PREDICTING HEAD CIRCUMFERENCE IN LOW BIRTHWEIGHT INFANTS : A MULTIPLE LINEAR REGRESSION APPROACH

ABSTRACT

This study examines the use of multiple linear regression to predict head circumference in low birth weight infants using the dataset "Low Birth Weight Infants." The dataset contains information for a sample of 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts. Predictor variables include birth weight, gestational age, and various demographic and clinical characteristics of the mother. The results indicate that the model has a moderate degree of accuracy in predicting head circumference using birth weight, gestational age and sex of the child.with an R-squared value of 0.76, the model is able to explain about 76% of variations in the data. Additionally, diagnostic checking will be performed to ensure the assumptions of the linear regression model are not violated.

Keywords: gestational age, birth weight , toxemia

PRELIMINARY ANALYSIS

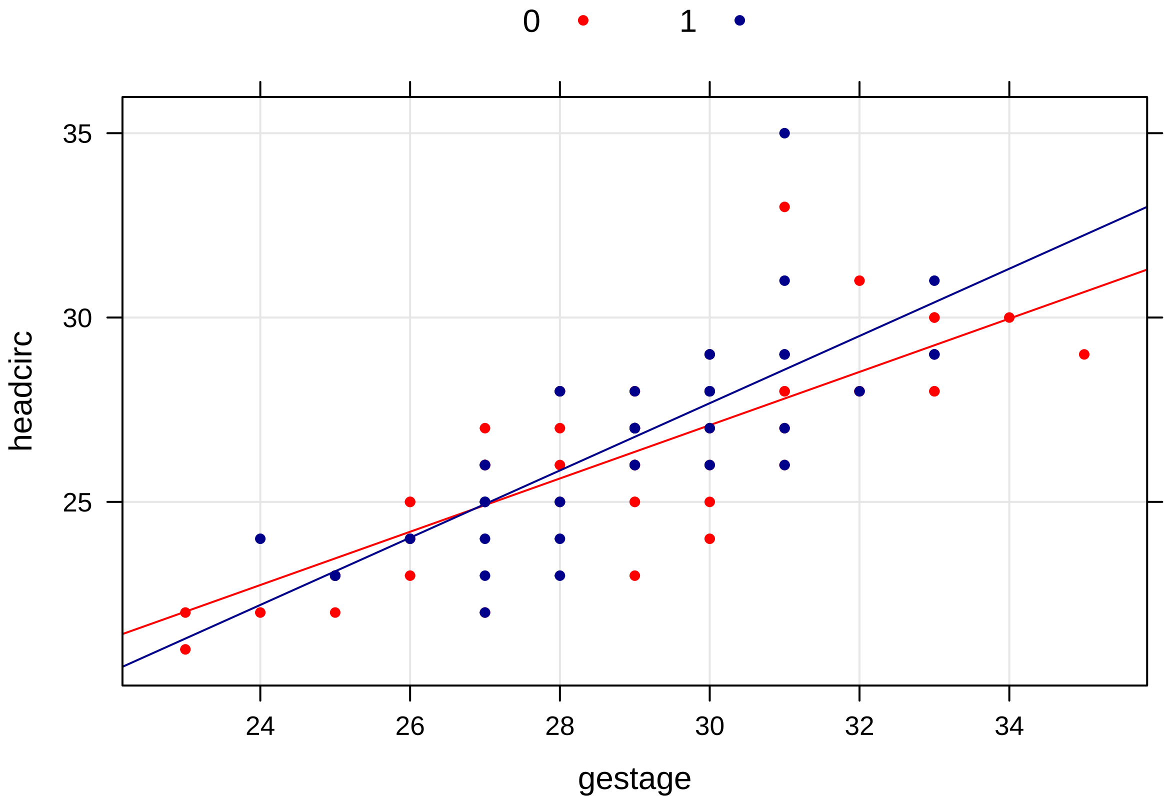
Descriptive Statistics

To begin, we describe the dataset using the five number summary.

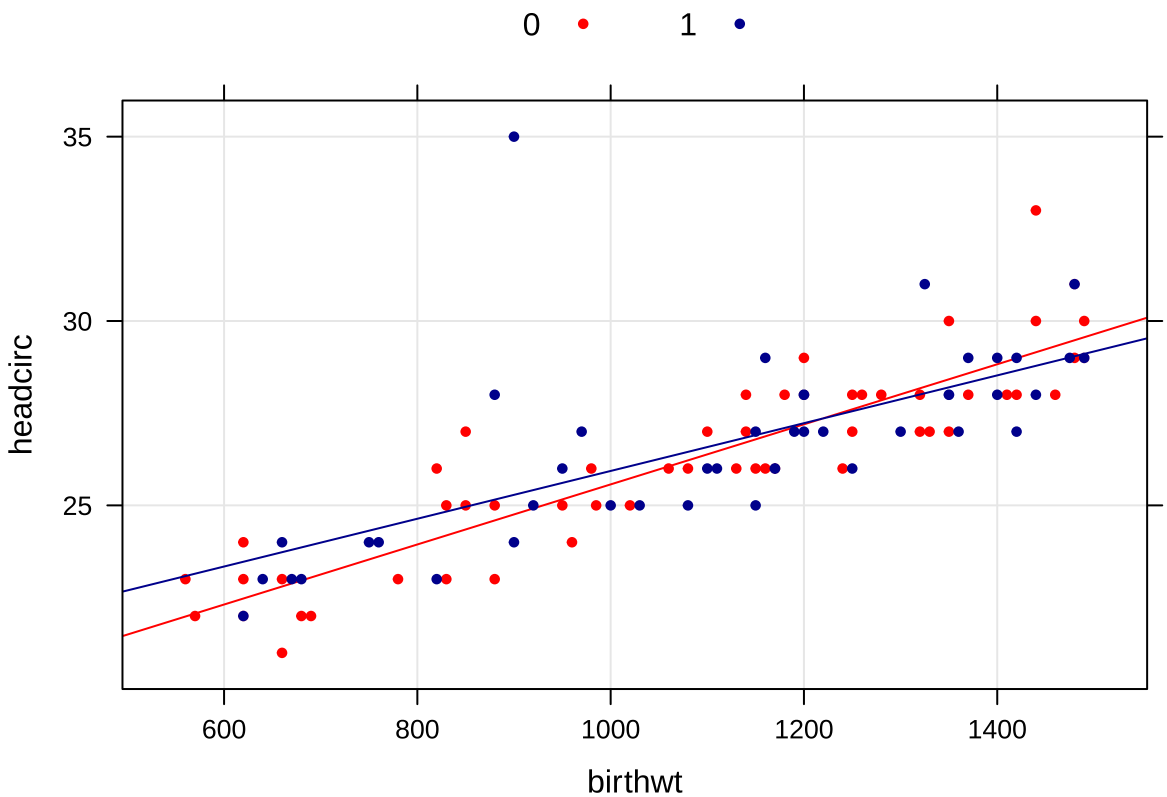
	mean	sd	IQR	0%	25%	50%	75%	100%	n
momage	27.73	5.9828965	9.00	14	23	28	32.00	41	100
gestage	28.89	2.5341904	4.00	23	27	29	31.00	35	100
toxemia	0.21	0.4093602	0.00	0	0	0	0.00	1	100
length	36.82	3.5714355	4.00	20	35	38	39.00	43	100
headcirc	26.45	2.5321169	3.00	21	25	27	28.00	35	100
birthwt	1098.85	269.9933174	446.25	560	880	1155	1326.25	1490	100
sbp	47.08	11.4032425	12.25	19	40	47	52.25	87	100
sex	0.44	0.4988877	1.00	0	0	0	1.00	1	100
grmhem	0.15	0.3588703	0.00	0	0	0	0.00	1	100
apgar5	6.25	2.4303427	3.00	0	5	7	8.00	9	100

From the summary statistics, we can see that the mean birth weight of the infants is 1099 grams, with a standard deviation of 269.99 grams. The average gestational age of the infants is 28.89 weeks, and the average age of the mothers is 27.73 years. The variable toxemia has a mean of 0.21, indicating that 21% of mothers had toxemia. The average head circumference of the infants is 26.45 centimeters. We can see that , the mean head circumference for male babies is slightly higher than females. On average, we can see that males babies have higher weights as compared to female babies.

Graphical Summaries

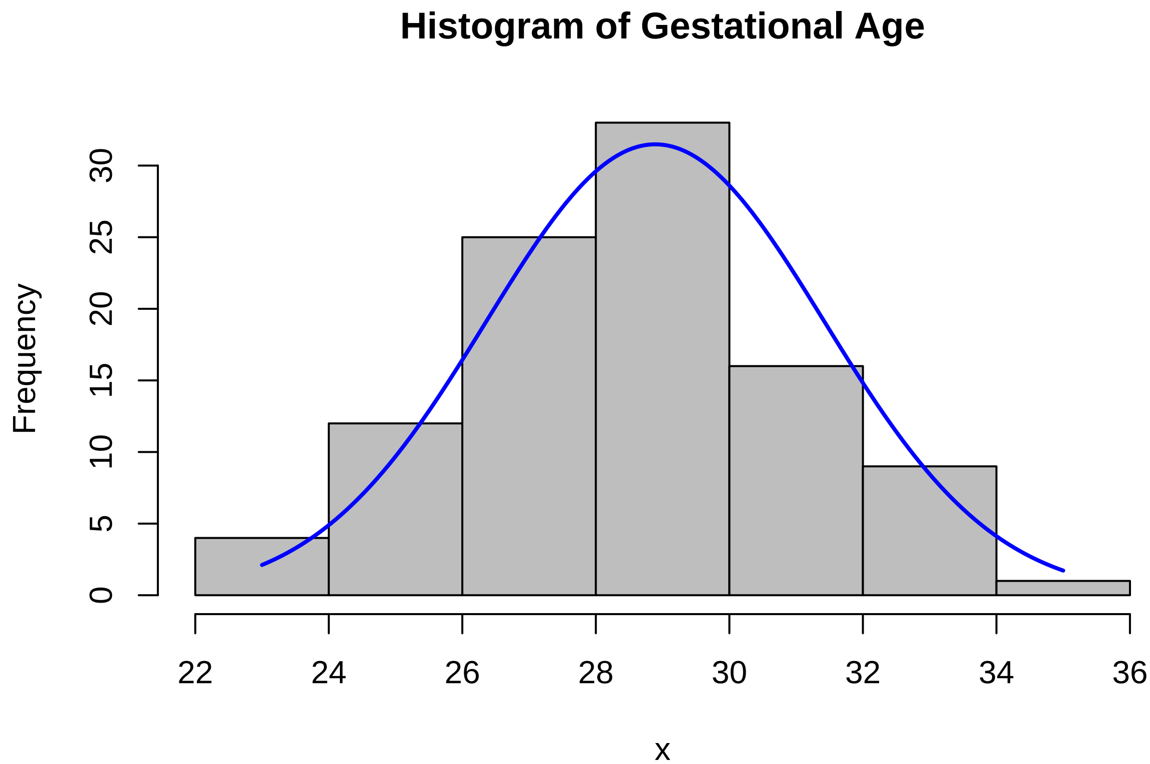


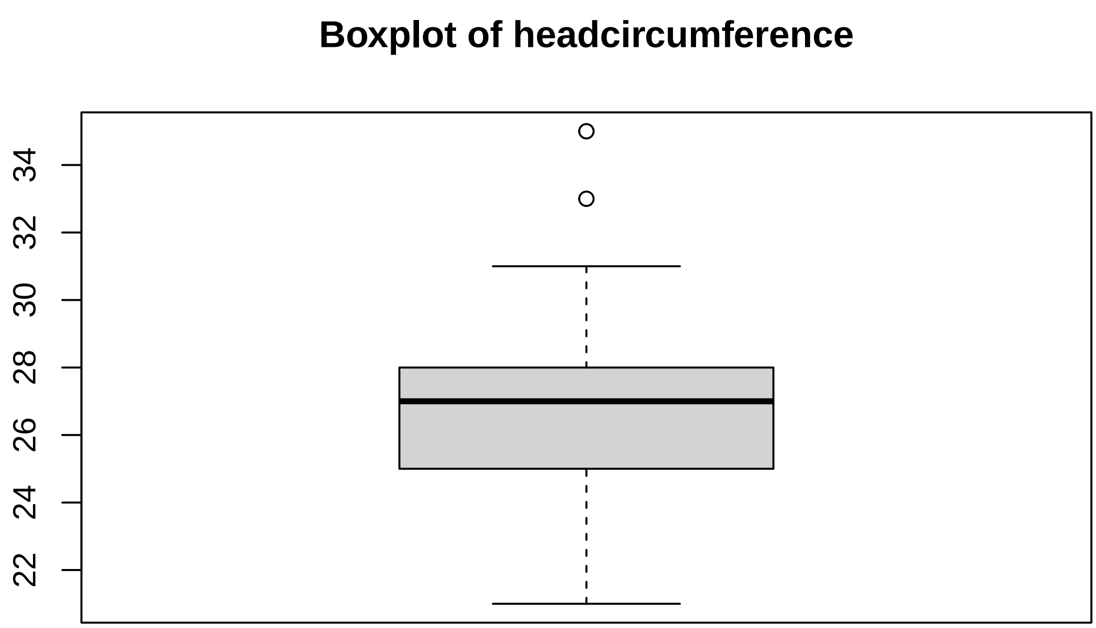
From the graph above we can see that the head circumference for both females and males increases as gestational age increases. However male head circumference tend to increase faster than female head circumference as gestational age increases.

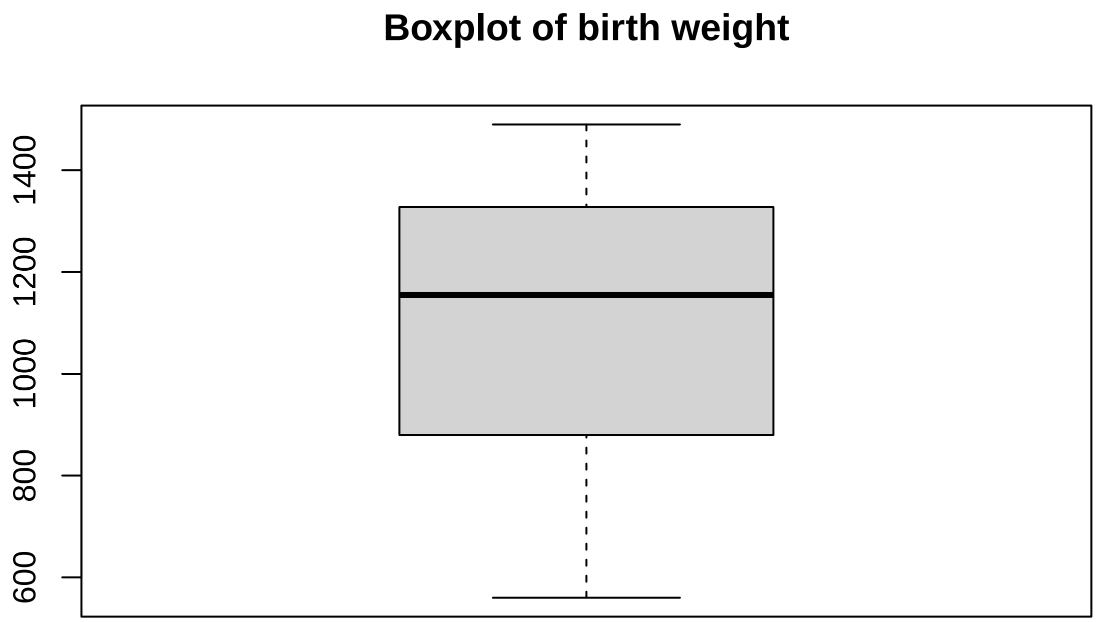


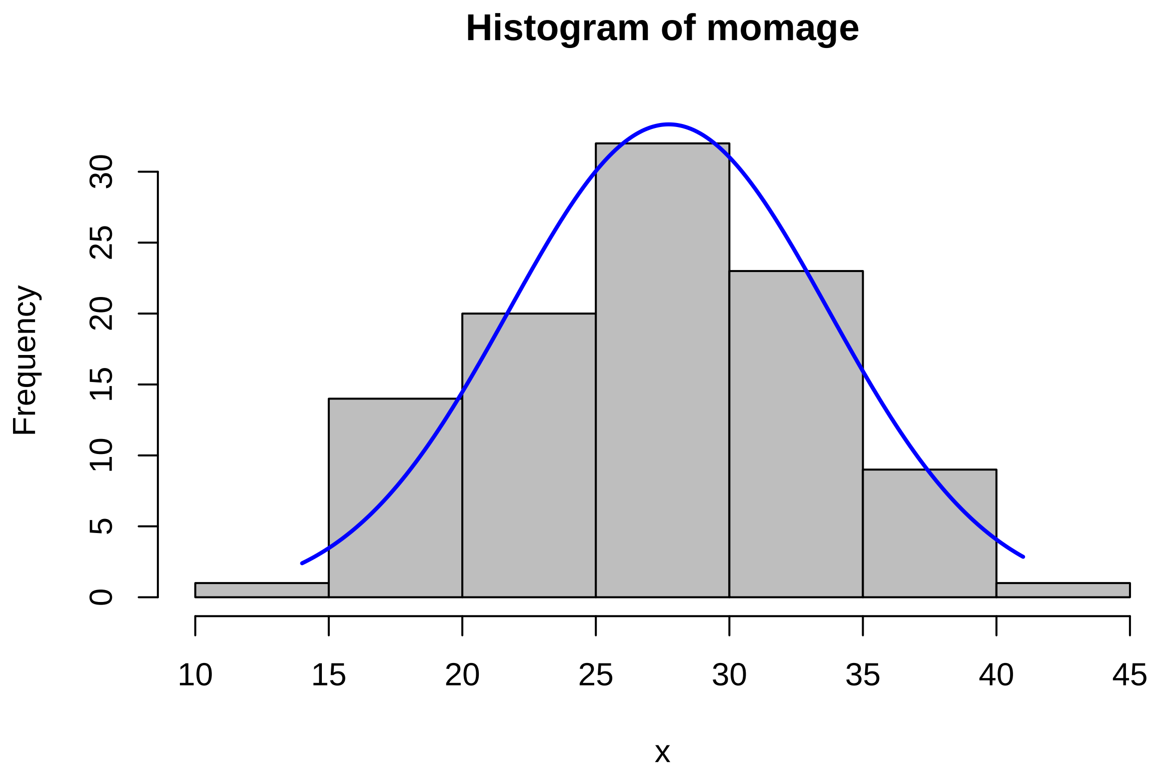
From the graph above we can see that the head circumference for both females and males increases as birth weight increases. However female head circumference tend to increase faster than male head circumference as birht weight increases.

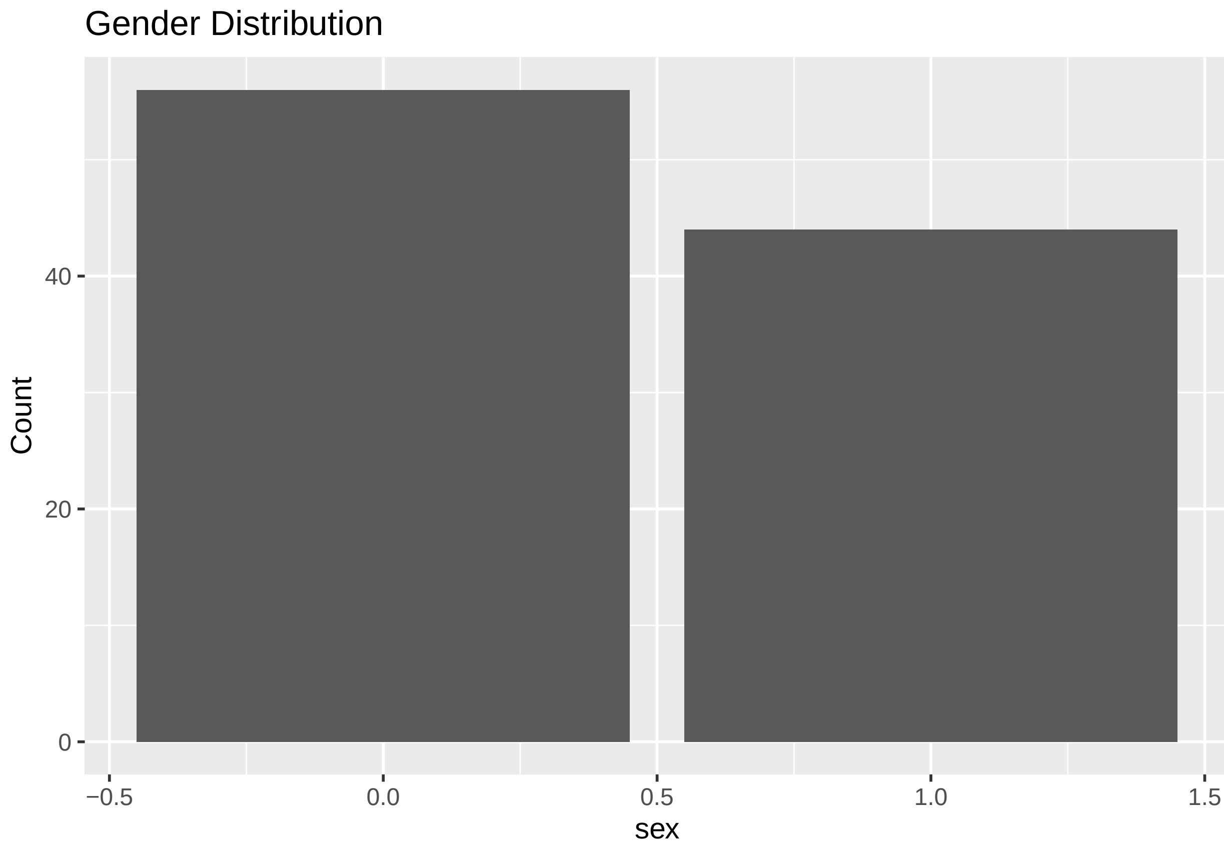
Checking the Distributions of the data






Boxplot of birth weight
600
800
1000
1200
1400


Histogram of momage
Frequency
0
5
10
15
20
25
30
10
15
20
25
30
35
40
45
x



The graphs reveal that gestational age and momage approximately follow a normal distribution. However, the distribution of head circumference appears slightly skewed to the right, indicating a potential deviation from normality. Moreover, the presence of outliers in the data is evident. In contrast, the distribution of birth weight is slightly skewed to the left. Furthermore, the histogram illustrates a greater frequency of female babies (coded as 0) compared to male babies (coded as 1).

COrrelation Matrix

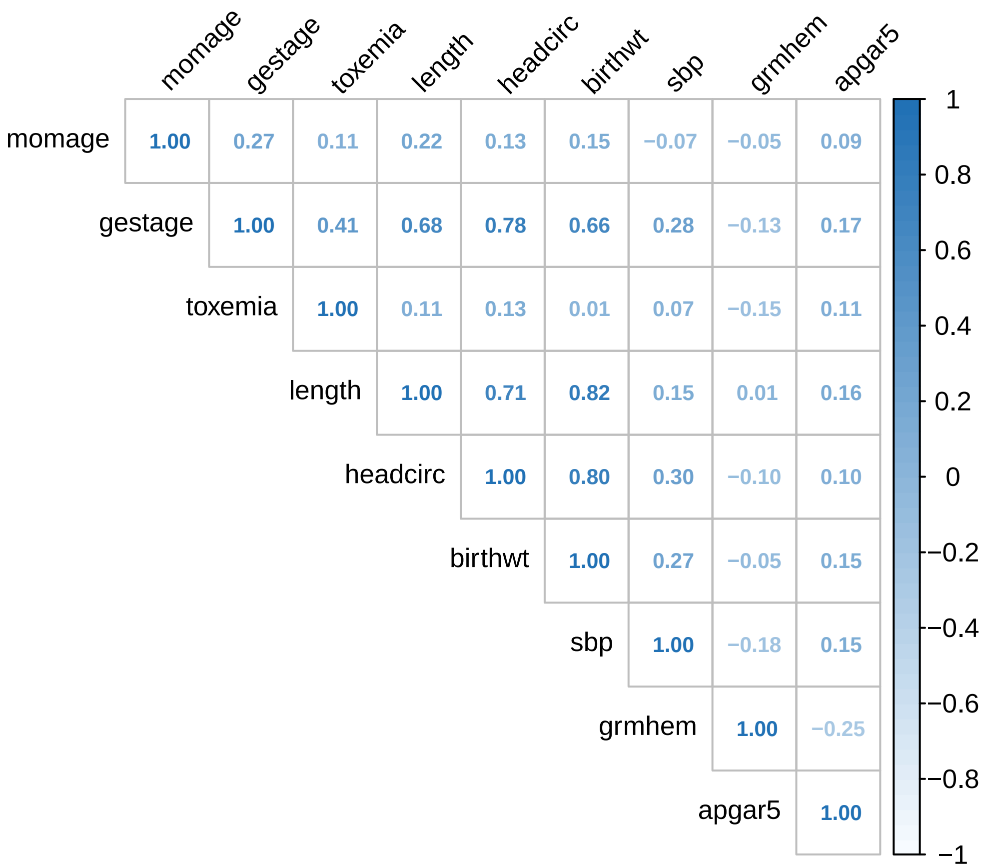
	momage	gestage	toxemia	length	headcirc	birthwt	sbp	grmhem	apgar5
momage	1.00	0.10	0.02	0.06	-0.04	-0.02	-0.38	-0.21	-0.05
gestage	0.10	1.00	0.29	0.83	0.91	0.82	0.25	-0.54	0.00
toxemia	0.02	0.29	1.00	-0.12	-0.04	-0.21	-0.04	-0.43	0.05
length	0.06	0.83	-0.12	1.00	0.91	0.96	0.09	-0.29	-0.03
headcirc	-0.04	0.91	-0.04	0.91	1.00	0.95	0.32	-0.40	-0.05
birthwt	-0.02	0.82	-0.21	0.96	0.95	1.00	0.26	-0.32	-0.02
sbp	-0.38	0.25	-0.04	0.09	0.32	0.26	1.00	-0.50	0.13
grmhem	-0.21	-0.54	-0.43	-0.29	-0.40	-0.32	-0.50	1.00	-0.60
apgar5	-0.05	0.00	0.05	-0.03	-0.05	-0.02	0.13	-0.60	1.00

n= 9

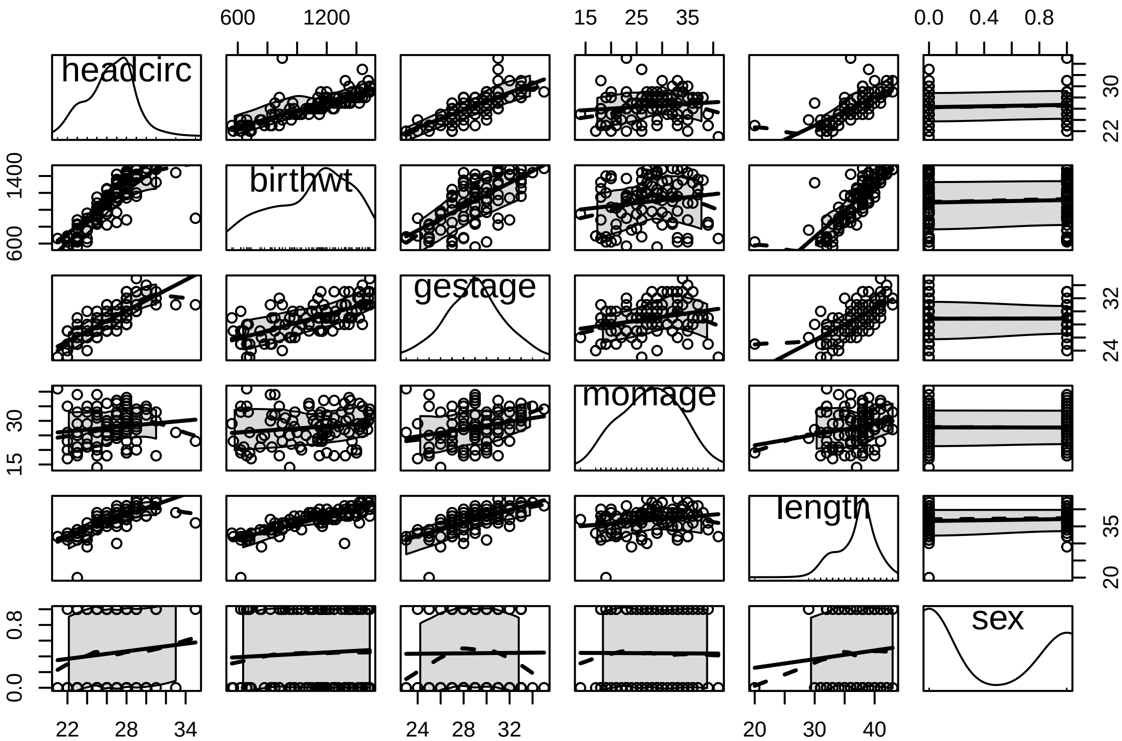
P

	momage	gestage	toxemia	length	headcirc	birthwt	sbp	grmhem	apgar5
momage		0.8059	0.9536	0.8705	0.9253	0.9603	0.3153	0.5842	0.9084

gestage	0.8059		0.4557	0.0060	0.0006	0.0073	0.5186	0.1351	0.9947
toxemia	0.9536	0.4557		0.7487	0.9163	0.5901	0.9088	0.2541	0.9054
length	0.8705	0.0060	0.7487		0.0008	0.0000	0.8094	0.4548	0.9420
headcirc	0.9253	0.0006	0.9163	0.0008		0.0000	0.4071	0.2873	0.8937
birthwt	0.9603	0.0073	0.5901	0.0000	0.0000		0.5030	0.3935	0.9659
sbp	0.3153	0.5186	0.9088	0.8094	0.4071	0.5030		0.1714	0.7305
grmhem	0.5842	0.1351	0.2541	0.4548	0.2873	0.3935	0.1714		0.0844
apgar5	0.9084	0.9947	0.9054	0.9420	0.8937	0.9659	0.7305	0.0844	



From the correlation plot above , head circumference is highly correlated with birhtweight with a coefficient of 0.80 and gestational age with a coefficient of 0.78. We can also see that head circumference and lenght are highly correlated with a coefficient of 0.71.Length has a moderate positive correlation with gestage, headcirc, and birthwt, and a weak positive correlation with sbp and apgar5.Birthwt has a strong positive correlation with length and headcirc, and a weak positive correlation with momage, gestage, and apgar5. We can see from the p-values table that , all variables pair with p-value less than 0.05 have some positive correlation.This suggests that there is strong evidence to reject the null hypothesis that there is no correlation between the variables, and to conclude that the observed correlation coefficient is not likely to have arisen by chance.



REGRESSION MODELS

The next step in this project involves selecting the best model for the data. I will use AIC , LASSO , Cp and d R-Squared model selection criteria.

Full regression model

We will first run the regression model with all the independent variables and interactions to determine the significant variables.

```
Call:
lm(formula = headcirc ~ ., data = low_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8131 -0.6869 -0.0500  0.3018  7.8667

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.9545996  2.1908027   3.174  0.00206 **
momage      -0.0269360  0.0226181  -1.191  0.23682
gestage      0.5203399  0.0861789   6.038 3.44e-08 ***
```

```
toxemia     -0.5308288  0.3744738  -1.418  0.15978
length       0.0191553  0.0676638   0.283  0.77775
birthwt      0.0041124  0.0009096   4.521 1.87e-05 ***
sbp          0.0085803  0.0123204   0.696  0.48795
sex          0.2605805  0.2625683   0.992  0.32365
grmhem      -0.2016148  0.3832020  -0.526  0.60009
apgar5      -0.0626223  0.0556996  -1.124  0.26388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.276 on 90 degrees of freedom
Multiple R-squared:  0.7691,    Adjusted R-squared:  0.746
F-statistic: 33.31 on 9 and 90 DF,  p-value: < 2.2e-16
```

We can see that gestational age and birth weight are the significant (p-value < 05) variable in the model

Reduced regression model

```
Call:
lm(formula = headcirc ~ gestage + birthwt, data = low_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0350 -0.7271 -0.0765  0.3472  8.5402

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.3080154  1.5789429   5.262 8.54e-07 ***
gestage     0.4487328  0.0672460   6.673 1.56e-09 ***
birthwt     0.0047123  0.0006312   7.466 3.60e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.274 on 97 degrees of freedom
Multiple R-squared:  0.752, Adjusted R-squared:  0.7469
F-statistic: 147.1 on 2 and 97 DF,  p-value: < 2.2e-16
```

In this model , we remove all insignificant variables from the initial model.

Reduced regression with Interactions

```
Call:
lm(formula = headcirc ~ gestage + birthwt + gestage * (as.factor(sex)) +
    birthwt * (as.factor(sex)), data = low_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0799 -0.6226 -0.0940  0.4105  7.6358

Coefficients:
```

```
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            11.1101375  1.8958315   5.860 6.80e-08 ***
gestage                 0.3021018  0.0848625   3.560 0.000584 ***
birthwt                 0.0059261  0.0008635   6.863 7.13e-10 ***
as.factor(sex)1        -7.8162725  3.3002668  -2.368 0.019914 *
gestage:as.factor(sex)1  0.3754163  0.1371646   2.737 0.007415 **
birthwt:as.factor(sex)1 -0.0025180  0.0012294  -2.048 0.043341 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.238 on 94 degrees of freedom
Multiple R-squared:  0.7731,    Adjusted R-squared:  0.7611
F-statistic: 64.07 on 5 and 94 DF,  p-value: < 2.2e-16
```

Now, in this model we included interaction terms. We can see from initial plots that there was interaction between gestational age and sex and birth weight and sex. The adjusted R-squared increased which means that the interactions effect is significant in predicting the model.

AIC Approach

```
Start:  AIC=53.32
headcirc ~ gestage + birthwt + gestage * (as.factor(sex)) + birthwt *
    (as.factor(sex)) + momage + length + sbp

                         Df Sum of Sq    RSS    AIC
- length                  1    0.0273 142.38 51.335
- sbp                     1    0.3284 142.68 51.546
- momage                  1    1.1466 143.50 52.118
<none>                                142.36 53.316
- birthwt:as.factor(sex)  1    5.4978 147.85 55.105
- gestage:as.factor(sex)  1    9.5914 151.95 57.836

Step:  AIC=51.34
headcirc ~ gestage + birthwt + as.factor(sex) + momage + sbp +
    gestage:as.factor(sex) + birthwt:as.factor(sex)

                         Df Sum of Sq    RSS    AIC
- sbp                     1    0.3050 142.69 49.549
- momage                  1    1.1268 143.51 50.123
<none>                                142.38 51.335
- birthwt:as.factor(sex)  1    5.4894 147.87 53.118
+ length                  1    0.0273 142.36 53.316
- gestage:as.factor(sex)  1    9.6135 152.00 55.869

Step:  AIC=49.55
headcirc ~ gestage + birthwt + as.factor(sex) + momage + gestage:as.factor(sex) +
    birthwt:as.factor(sex)

                         Df Sum of Sq    RSS    AIC
- momage                  1    1.3164 144.00 48.467
<none>                                142.69 49.549
+ sbp                     1    0.3050 142.38 51.335
- birthwt:as.factor(sex)  1    5.7003 148.39 51.466
```

```
+ length                  1    0.0040 142.68 51.546
- gestage:as.factor(sex)  1    9.9515 152.64 54.291

Step:  AIC=48.47
headcirc ~ gestage + birthwt + as.factor(sex) + gestage:as.factor(sex) +
    birthwt:as.factor(sex)

                         Df Sum of Sq    RSS    AIC
<none>                               144.00 48.467
+ momage                  1    1.3164 142.69 49.549
+ sbp                     1    0.4947 143.51 50.123
+ length                  1    0.0025 144.00 50.466
- birthwt:as.factor(sex)  1    6.4260 150.43 50.833
- gestage:as.factor(sex)  1   11.4760 155.48 54.135


Call:
lm(formula = headcirc ~ gestage + birthwt + as.factor(sex) +
    gestage:as.factor(sex) + birthwt:as.factor(sex), data = low_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0799 -0.6226 -0.0940  0.4105  7.6358

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             11.1101375  1.8958315   5.860 6.80e-08 ***
gestage                  0.3021018  0.0848625   3.560 0.000584 ***
birthwt                  0.0059261  0.0008635   6.863 7.13e-10 ***
as.factor(sex)1         -7.8162725  3.3002668  -2.368 0.019914 *
gestage:as.factor(sex)1  0.3754163  0.1371646   2.737 0.007415 **
birthwt:as.factor(sex)1 -0.0025180  0.0012294  -2.048 0.043341 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.238 on 94 degrees of freedom
Multiple R-squared:  0.7731,    Adjusted R-squared:  0.7611
F-statistic: 64.07 on 5 and 94 DF,  p-value: < 2.2e-16
```

In this model, the AIC starts at 53.32, the model then drop one variable each time until it has the lowest AIC 48.47 Therefore we have gestage, birthweight and the interactions of gestage , birthweight and sex as the significant variables. The adjusted R-Squared increased in this model to 0.7611

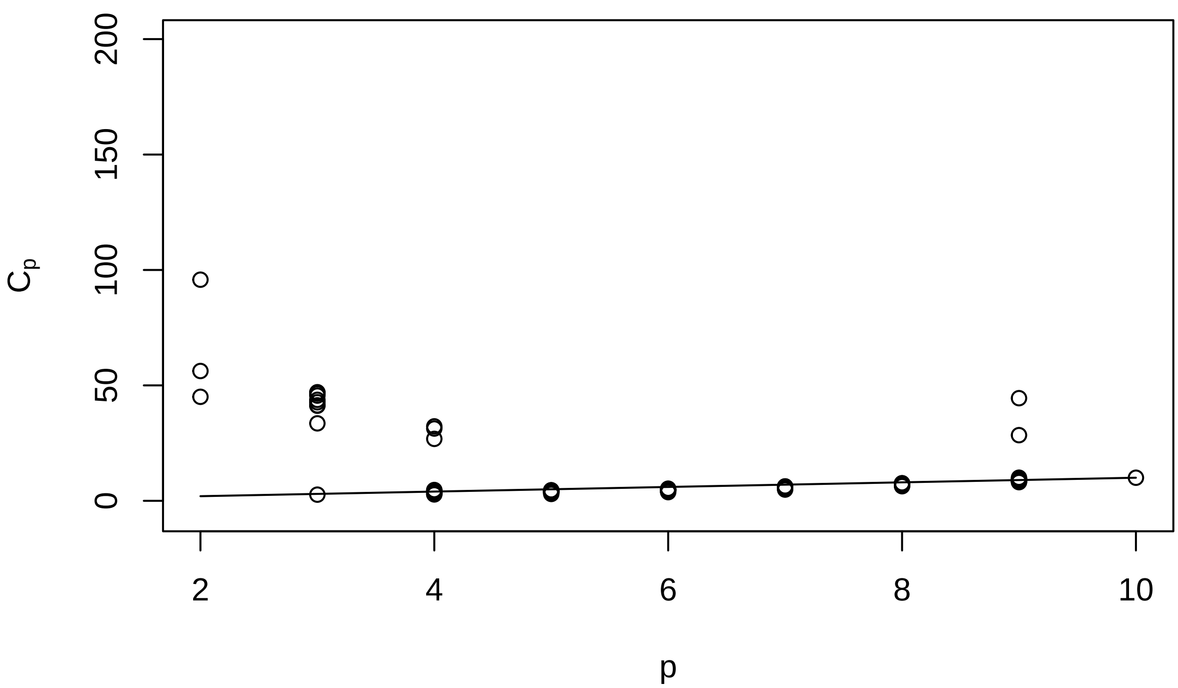
LASSO Model

```
10 x 1 sparse Matrix of class "dgCMatrix"
                     s1
(Intercept) 12.976461800
momage       .
gestage      0.329674713
toxemia      .
length       .
birthwt      0.003593972
```

```
sbp          .
sex          .
grmhem       .
apgar5       .
```

In the LASSO model, We can also see that the insignificant variables are been dropped off. Hence we have birth weight and gest age as the significant variables.

Model Selection by CP

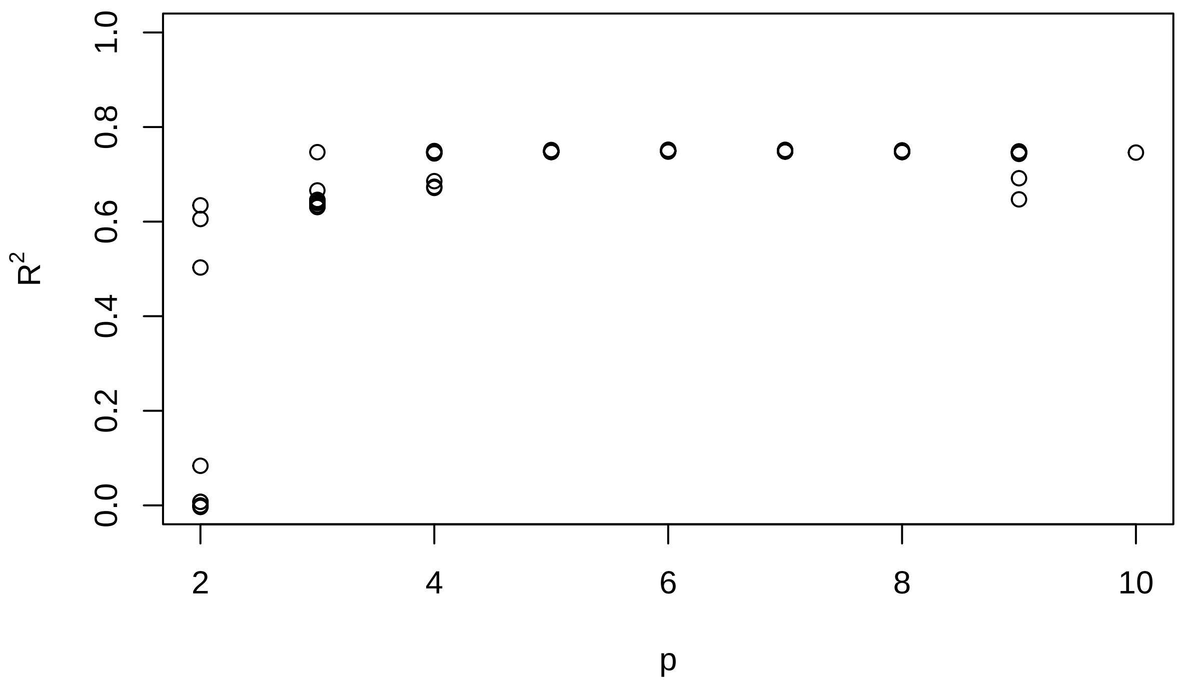



```
Call:
lm(formula = y ~ foo)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.0350 -0.7271 -0.0765  0.3472  8.5402 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 8.3080154  1.5789429   5.262 8.54e-07 ***
foogestage  0.4487328  0.0672460   6.673 1.56e-09 ***
foobirthwt  0.0047123  0.0006312   7.466 3.60e-11 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.274 on 97 degrees of freedom
Multiple R-squared:  0.752, Adjusted R-squared:  0.7469
F-statistic: 147.1 on 2 and 97 DF,  p-value: < 2.2e-16
```

In the Cp model, We can also see that the insignificant variables are been dropped off. Hence we have birthweight and gestage as the significant variables.

Model Selection by R-Squared



MODEL SELECTION

COMPARING REGRESSION MODELS PERFORMANCE

Metrics for model 1 (model with all variables)

```
# A tibble: 1 x 5
  adj.r.squared sigma   AIC   BIC  p.value
          <dbl> <dbl> <dbl> <dbl>    <dbl>
1         0.746  1.28  344.  373. 5.73e-25
```

Metrics for model2 (model with all significant variables)

```
# A tibble: 1 x 5
  adj.r.squared sigma   AIC   BIC  p.value
          <dbl> <dbl> <dbl> <dbl>    <dbl>
1         0.747  1.27  337.  348. 4.28e-30
```

Metrics for model 3(model with interactions)

```
# A tibble: 1 x 5
  adj.r.squared sigma   AIC   BIC  p.value
          <dbl> <dbl> <dbl> <dbl>    <dbl>
1         0.761  1.24  334.  352. 9.11e-29
```

Metrics for model 4(model with interactions)

```
# A tibble: 1 x 5
  adj.r.squared sigma   AIC   BIC  p.value
          <dbl> <dbl> <dbl> <dbl>    <dbl>
1         0.761  1.24  334.  352. 9.11e-29
```

Metrics for model 5(model with interactions)

```
# A tibble: 1 x 5
  adj.r.squared sigma   AIC   BIC  p.value
          <dbl> <dbl> <dbl> <dbl>    <dbl>
1         0.747  1.27  337.  348. 4.28e-30
```

From the above computations we can see that model 3 and 4 (model with interactions effect of sex) has the lowest AIC, BIC, P-Value and biggest adjusted R-Squared.For the rest of the analysis we will use model 3.

Model interpretation

```r
summary(low.lm3)
```

```

Call:
lm(formula = headcirc ~ gestage + birthwt + gestage * (as.factor(sex)) +
    birthwt * (as.factor(sex)), data = low_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0799 -0.6226 -0.0940  0.4105  7.6358

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             11.1101375  1.8958315   5.860 6.80e-08 ***
gestage                  0.3021018  0.0848625   3.560 0.000584 ***
birthwt                  0.0059261  0.0008635   6.863 7.13e-10 ***
```

```
as.factor(sex)1         -7.8162725  3.3002668  -2.368 0.019914 *
gestage:as.factor(sex)1  0.3754163  0.1371646   2.737 0.007415 **
birthwt:as.factor(sex)1 -0.0025180  0.0012294  -2.048 0.043341 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.238 on 94 degrees of freedom
Multiple R-squared:  0.7731,    Adjusted R-squared:  0.7611
F-statistic: 64.07 on 5 and 94 DF,  p-value: < 2.2e-16
```

This is our final model. From the coefficients, we can form our model as follows:

$HeadCircumference = 11.1101375 + 0.3021018(Gestage) + 0.0059261(birthweight) - 7.8162725(sex1) + 0.3754163(sex1 * gestage) - 0.0025180(sex1 * birthweight)$

Coefficients

The intercept is 11.1101375cm. This means that, on average head circumference will be 11.1101375cm when all other variables are zero.

The coefficient of gestage shows that for any additional week in gestational age , the average head circumference will increases by 0.3021018cm holding birthweight and sex constant.

The coefficient of birth weight shows that, for any additional gram increase in birth weight, hea circumference will increase by 0.0059261cm holding gestational age and sex constant.

The coefficient of sex shows that, the average head circumference for male infants is 7.8162725cm lesser than for female when gestational age and birthweight is zero.

This coefficient value cannot be relied upon since it not meaningful.

The coefficient of sex1*gestage shows that, for any additional week in gestational age, the average increase in head circumference for male infants is 0.3754163 higher for males than for female holding all other variables in the model constant.

The coefficient of sex1*birthweight shows that, for any additional gram in birth weight , the average increase in head circumference for male infants is 0.0025180 lower for males than for female holding all other variables in the model constant.

Residual standard error

The smaller the residual standard error, the better the fit. Hence, the result shows that our selected model predicts the PIQ with an average error of 1.238 on 94 degrees of freedom.

Overall F-statistics

The null hypothesis states that there is no relationship between the dependent variable and the independent variable(s) and the alternative hypothesis is that there exists at least one $\beta_i > 0$. In this study, the *F-statistics* is 64.07 with degrees of freedom on 5 and 94. The p-value < 2.2e-16 which is less than $\alpha = 0.05$. This means that there is at least one significant relationship between the independent variables and dependent variables in the model.

Multiple R-squared

The R-Squared measure the variability in the data that can be explained by the model.WIth an R-Squared of 0.7611 Our model explains around 76.11% the data variability.

Iindividual T-Test for the coefficients (β_i)

Gesage (β_1)

The null hypothesis states that,the coefficient for the gestage in the regression model is not significantly different from zero and the alternative hypothesis is that the coefficient for the gestage in the regression model is significantly different from zero. In this study, the T-statistics is 3.560.The p-value is 0.000584 which is less than $\alpha = 0.05$. We reject the null hypothesis and conclude that the coefficient for the gestate variable is significantly different from zero, indicating that, gestational age is a significant and good predictor variable for the head circumference even with birth weight ,sex, interactions of sex with birth weight and gestational age in the data.

birthweight (β_2)

The null hypothesis states that , the coefficient for the birthweight in the regression model is not significantly different from zero and the alternative hypothesis is that the coefficient for the birthweight in the regression model is significantly different from zero. In this study, the T-statistics is 6.863.The p-value is 7.13e-10 which is less than $\alpha = 0.05$. We reject the null hypothesis and conclude that the coefficient for the birthweight variable is significantly different from zero, indicating that, birth weight is a significant and good predictor variable for the head circumference even with gestage,sex,interactions of sex with birth weight and gestational age in the data.

sex (β_3)

The null hypothesis states that , the coefficient for sex in the regression model is not significantly different from zero and the alternative hypothesis is that the coefficient for the sex in the regression model issignificantly different from zero. In this study, the T-statistics is -2.368 .The p-value is 0.019914 which is less than $\alpha = 0.05$. We reject the null hypothesis and conclude that the coefficient for the sex variable is significantly different from zero, indicating that, sex is a significant and good predictor variable for the head circumference even with gestage ,gestational age,interactions of sex with birth weight and gestational age in the data.

Interaction of sex and gestage (β_4)

The null hypothesis states that , the coefficient for the (β_4) is not significantly different from zero and the alternative hypothesis is that the coefficient for the (β_4) in the regression model is significantly different from zero. In this study, the T-statistics is 2.737.The p-value is 0.007415 which is less than $\alpha = 0.05 = 0.05$. We reject the null hypothesis and conclude that the coefficient for the interaction of sex and gestage variable is significantly different from zero, indicating that, the interaction of sex and gestage is a significant and good predictor variable for the head circumference even with gestage,gestational age and interactions of sex with birth weight.

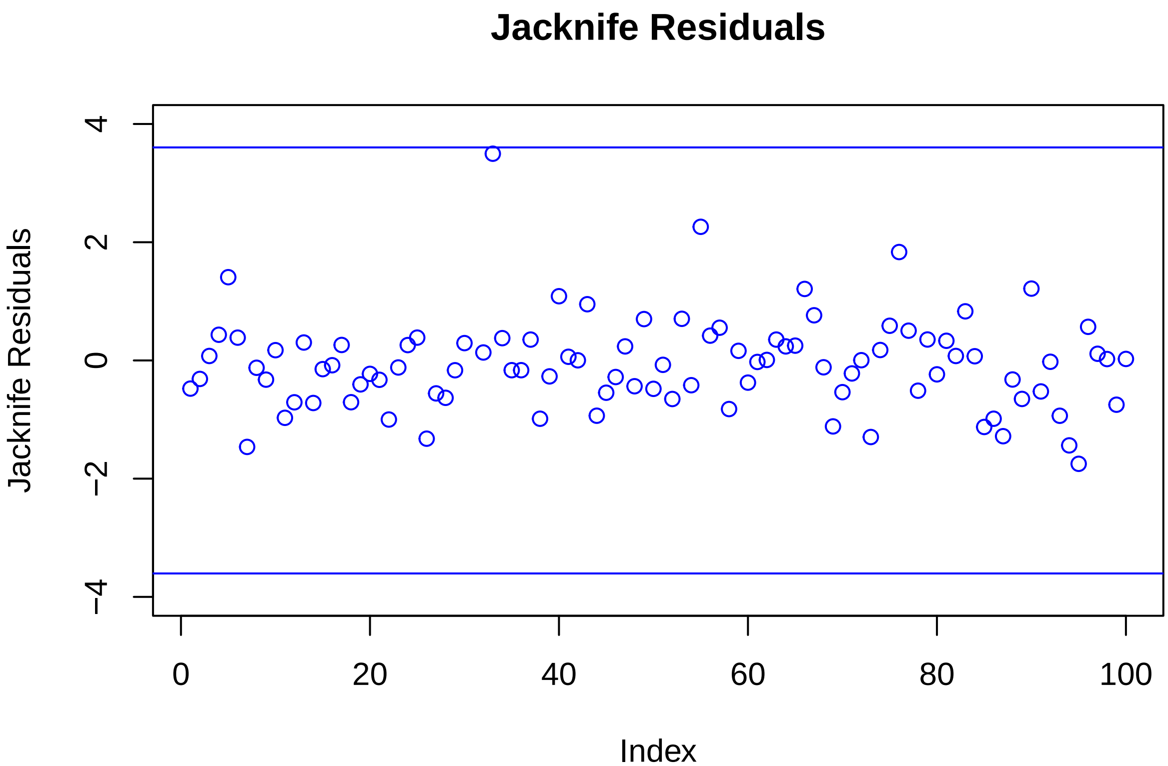
Interaction of sex and birth weight(β_5)

The null hypothesis states that,the coefficient for the (β_5)in the regression model is not significantly different from zero and the alternative hypothesis is that the coefficient for the (β_5) in the regression model is significantly different from zero. In this study, the T-statistics is -2.048. The p-value is 0.043341 which is less than $\alpha = 0.05 = 0.05$. We reject the null hypothesis and conclude that the coefficient for the gestage variable is significantly different from zero, indicating that, the interaction of sex and birth weight is a significant and good predictor variable for the head circumference even with gestational age, birth weight ,sex, and interactions of sex with gestage in the data.

MODEL DIAGNOSTICS

In this part we check to see whether all residual assumptions of linear regression models have been achieved. We will check for normality, linearity , equal variance , independence , and colinearity of residuals.

External Residual



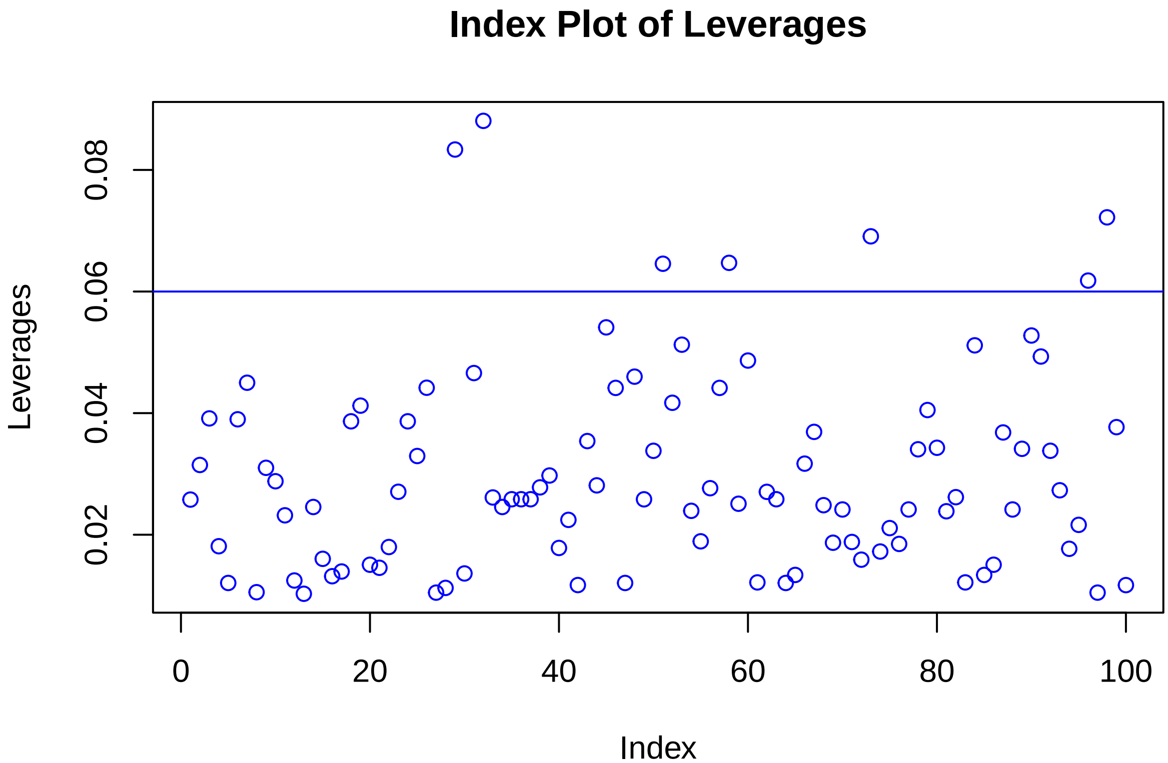
We can see that there are no critical outliers in the data.

Output the largest residual

```
      31
8.810603
```

The largest residual value is 8.810603

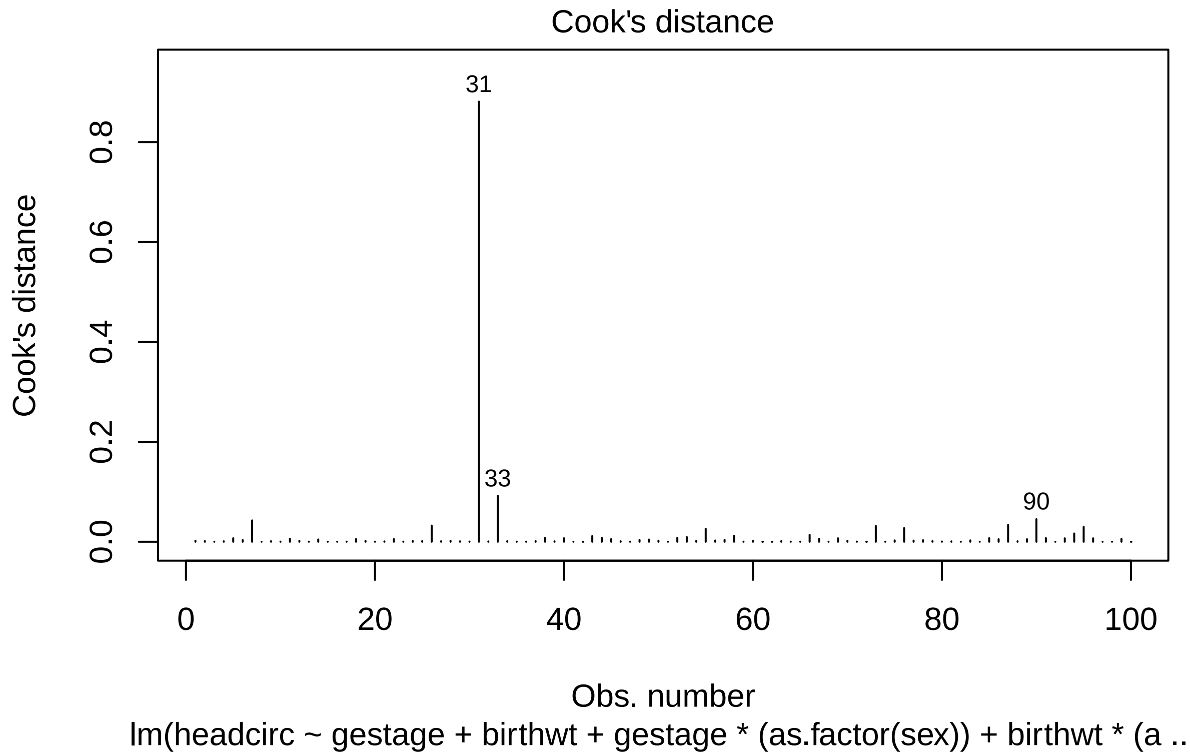
Leverage

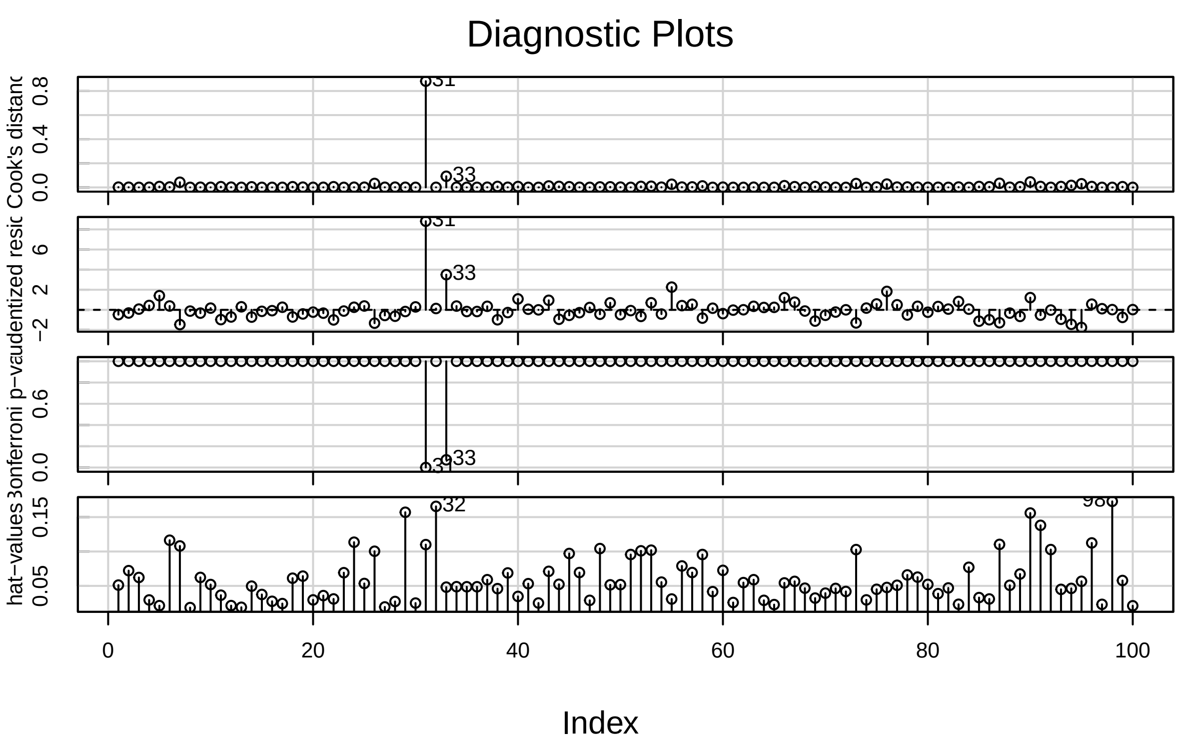



```
integer(0)
```

All the points above the 0.15 line indicates high leverage points which have the tendency of influencing the data.

Influential





```
Potentially influential observations of
     lm(formula = headcirc ~ gestage + birthwt + gestage * (as.factor(sex)) +      birthwt * (as.factor

   dfb.1_ dfb.gstg dfb.brth dfb.a.() dfb.g:.( dfb.b:.( dffit   cov.r   cook.d
6   0.00   0.00     0.00     0.08    -0.05    -0.03     0.14    1.20_*  0.00
24  0.00   0.00     0.00    -0.06     0.05     0.01     0.09    1.20_*  0.00
29  0.03  -0.05     0.07    -0.02     0.03    -0.05    -0.07    1.26_*  0.00
31  0.00   0.00     0.00    -1.57_*   1.96_*  -1.69_*   3.10_*  0.03_*  0.88
32 -0.03   0.05    -0.06     0.02    -0.03     0.04     0.06    1.28_*  0.00
33 -0.06  -0.12     0.51     0.03     0.07    -0.36     0.79_*  0.53_*  0.09
55  0.14  -0.02    -0.16    -0.08     0.01     0.12     0.40    0.80_*  0.03
91  0.00   0.00     0.00     0.15    -0.15     0.07    -0.21    1.22_*  0.01
98  0.00   0.00     0.00    -0.01     0.01    -0.01     0.01    1.29_*  0.00
   hat
6   0.12
24  0.11
29  0.16
31  0.11
32  0.17
33  0.05
55  0.03
91  0.14
98  0.17
```

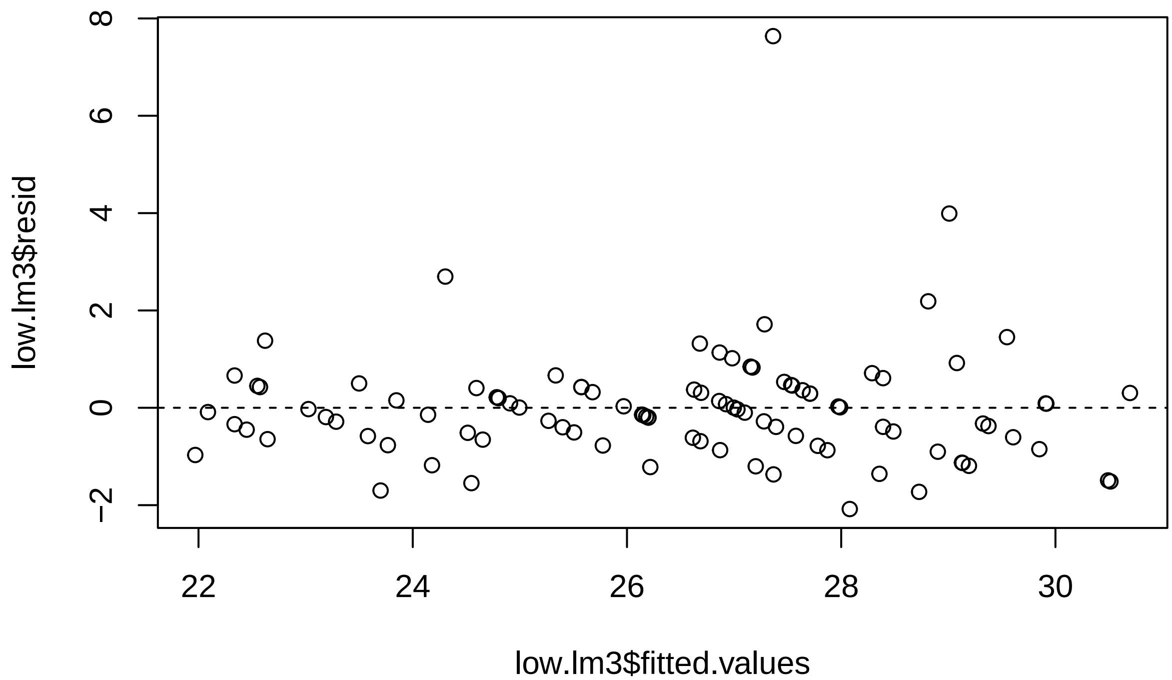
The table gives the points that are influential in the dataset

Bonferroni p-values for testing outliner

```
   rstudent unadjusted p-value Bonferroni p
31 8.810603         6.7066e-14   6.7066e-12
```

With a very small Bonferroni p- value which is less than 0.05, we can conclude that the effect of outliers in the data is significant.

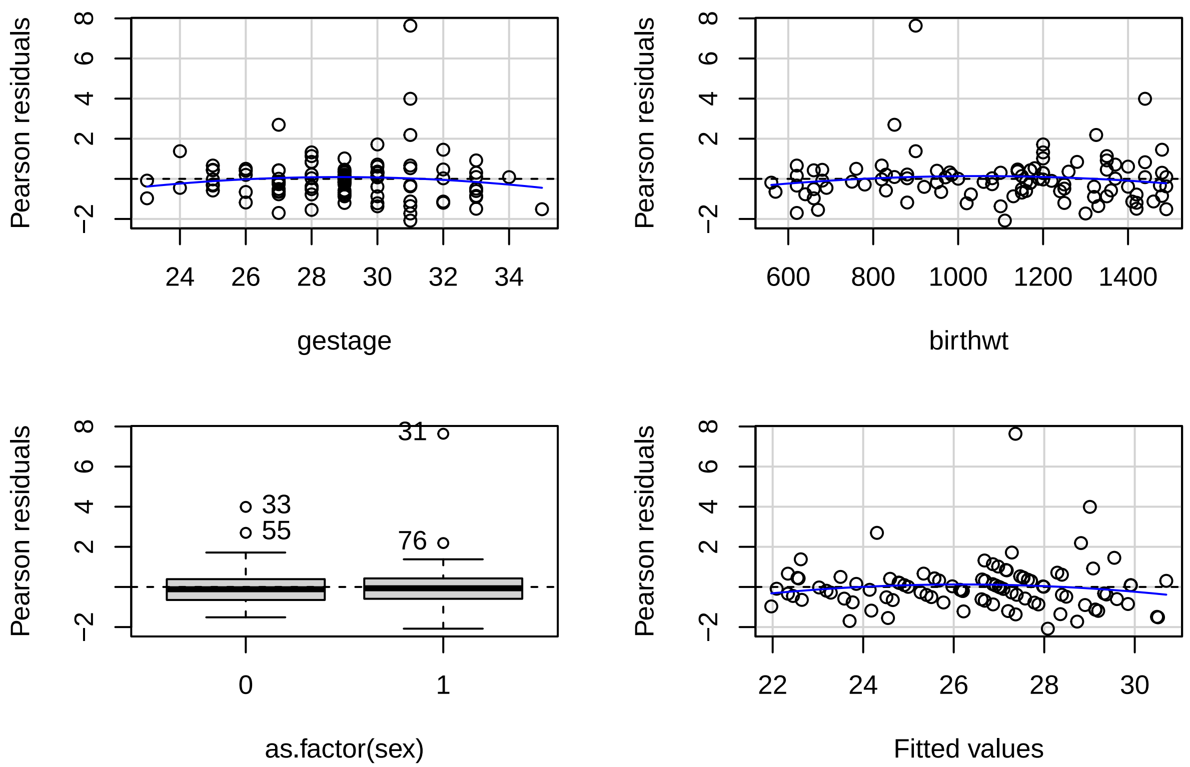
Homoscedasticity (Equal Variance Assumption)




```
    studentized Breusch-Pagan test

data:  low.lm3
BP = 10.676, df = 5, p-value = 0.05819
```

With a large Breusch Pagan p- value which is greater than 0.05, we can conclude that the variances are constant , hence Homoscedasticity is satisfied

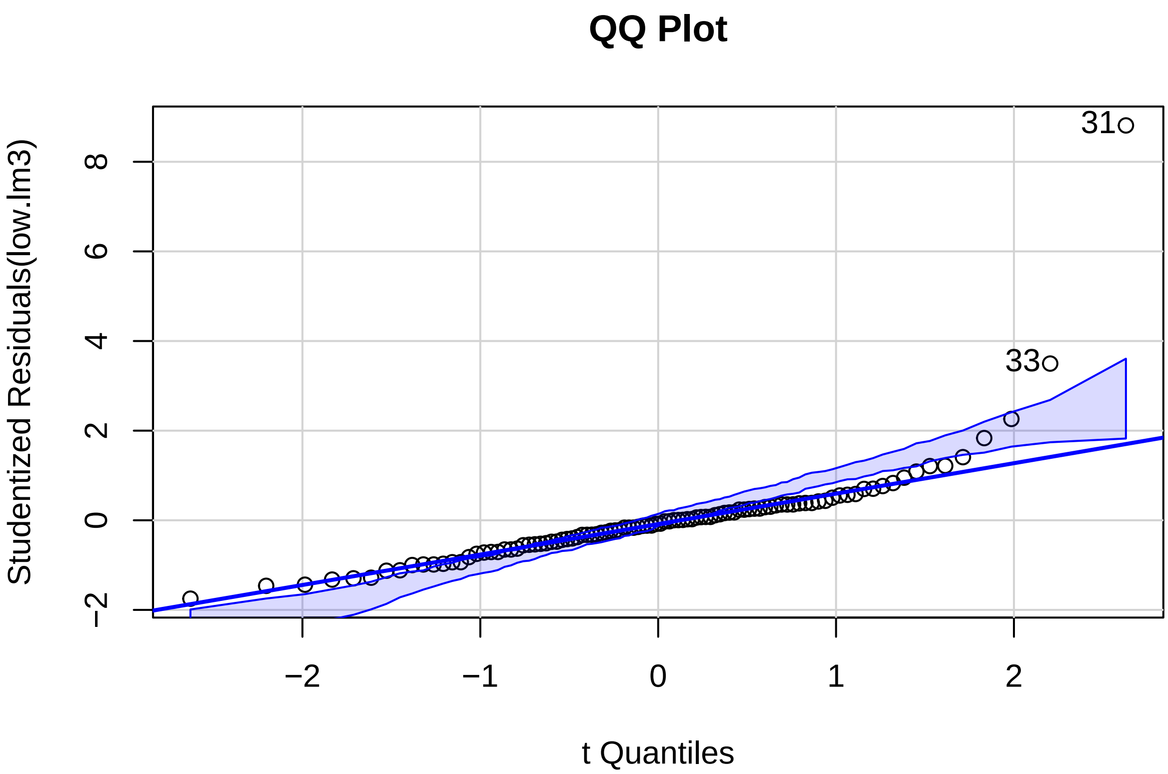
Linearity



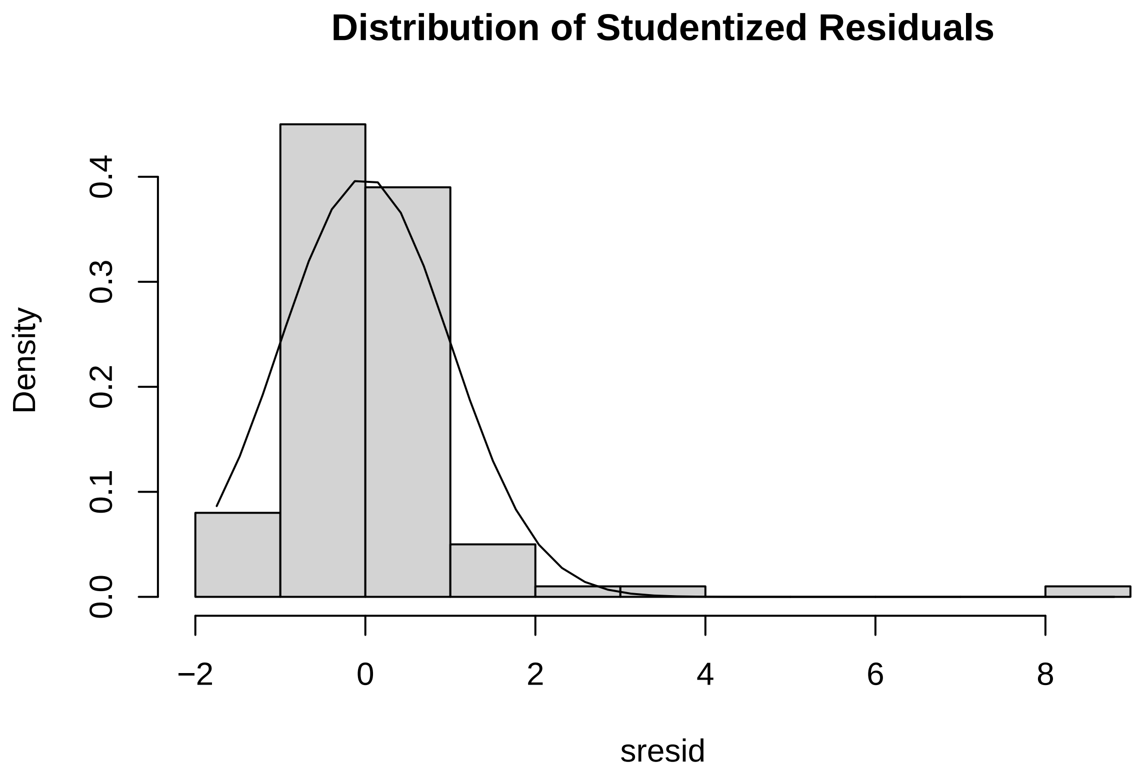
	Test stat	Pr(>|Test stat|)
gestage	-0.9498	0.3447
birthwt	-1.0410	0.3006
as.factor(sex)		
Tukey test	-1.1015	0.2707

From the plots and the test , we can conclude that the residuals are somewhat linear.

Normality Test




```
[1] 31 33
```




```
    Shapiro-Wilk normality test

data:  residuals(low.lm3)
W = 0.7808, p-value = 6.639e-11
```

From the normal plots and the shapiro normality test , we can say the residuals are not normally distributed.

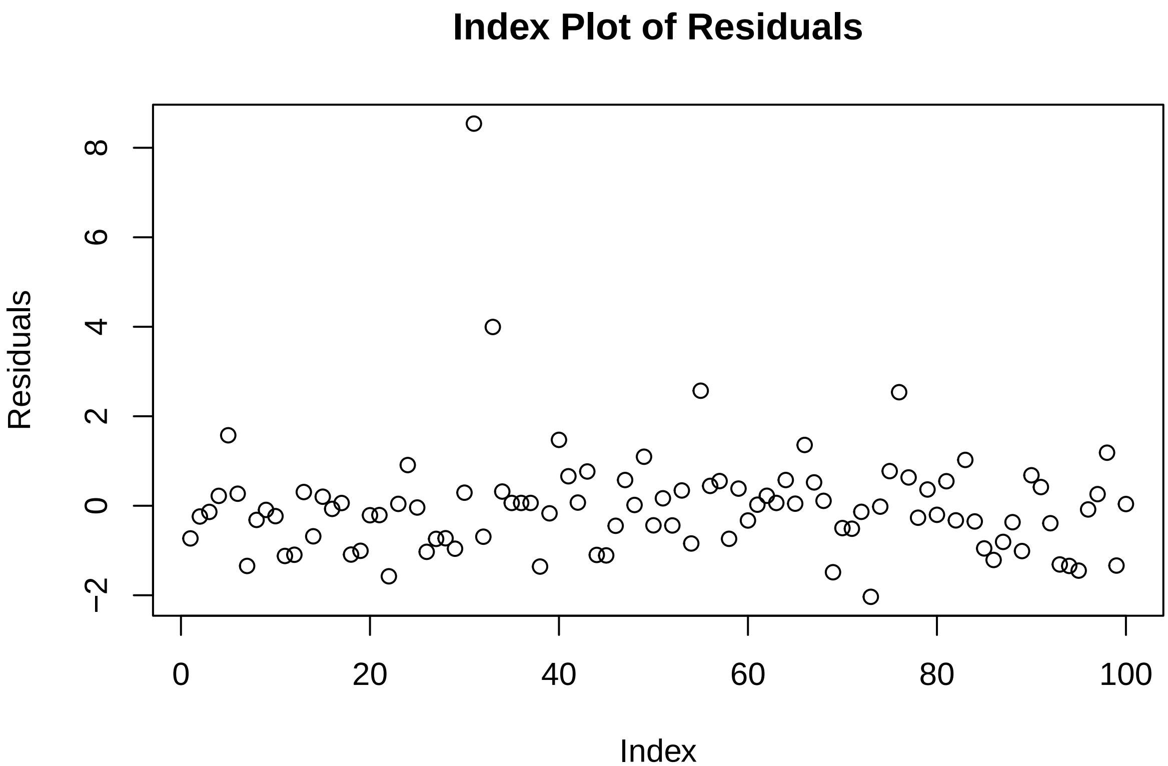
variance inflation factor (Multicolinearity)

```
there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif

              gestage                 birthwt         as.factor(sex)
             2.988805                3.512528             175.182450
gestage:as.factor(sex) birthwt:as.factor(sex)
           255.251407               33.116183

there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif

              gestage                 birthwt         as.factor(sex)
                FALSE                   FALSE                   TRUE
gestage:as.factor(sex) birthwt:as.factor(sex)
                 TRUE                    TRUE
```

with large variance inflation factor, we can conclude that there is colinearity in the data.

residual plot vs. PIQ (independence)

```
 lag Autocorrelation D-W Statistic p-value
   1       0.1264567      1.744769   0.164
 Alternative hypothesis: rho != 0
```



The residuals are spread out around zero, Hence the residuals are independent.The p-value for durbinWatsonTest confirms this.

#CONCLUSIONS From the results of the study, we can conclude that there is a significant positive relationship between head circumference and gestational age, birth weight, and sex of low birthweight infants. The model also reveals that there are interactions between birth weight and gestational age, as well as sex. This means that the relationship between head circumference and these factors is not constant, but varies depending on the sex type . This suggests that these factors can be used to predict the head circumference of low birth weight infants.

The R-squared value of 0.7611 indicates that the model explains 76.11% of the variation in the data. Thissuggests that the model is a good fit for the data, but there may be other factors that are not accounted for in the model that could also influence head circumference.

It is important to note that the model does not meet all the assumptions of multiple linear regression.This indicates that the results may not be entirely reliable and should be interpreted with caution. Further research is needed to validate these findings and to identify other factors that may be influencing head circumference in low birth weight infants.