האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

Data Analysis - Birth Weights

Final Paper

GENERALIZED LINEAR MODELS (52542)

Eden Levy

206096588

Introduction

The data I will analyze is about babies' birth weights. The data includes 189 births collected at Baystate Medical Center, Springfield, Massachusetts, during 1986.
In my sub-data, there are 136 mothers with different behaviors and physical data during pregnancy.
Low birth weight affects infant mortality, and I would like to find out which risk factors associated with low birth weight, can alter the probability of it.
I will analyze the data and discover the main reasons and ask the question – which factors and behaviors of the mother can determine whether a baby will be born underweight.

The explained variable is low weight - an indicator (categorical variable).
1 if the weight < 2.5 kilograms and 0 if the weight > 2.5 kilograms.

The explanatory variables:

- Mother's age in years **(continuous variable)**
- Mother's weight in the last period in pounds **(continuous variable)**
- Mother's race ('white', 'black', 'other') **(categorical variable)** - I changed it to 3 dummy columns of {1,0}.
- Smoking status during pregnancy **(categorical variable)**
- History of hypertension **(categorical variable)**
- Presence of uterine irritability **(categorical variable)**
- Number of physician visits during the first trimester **(categorical variable)**
- Number of previous premature labors **(categorical variable)**
- Birth weight in grams **(continuous variable)** – I later removed this variable from the model because it fits perfectly with the explained variable.

In this data there are many categorical variables, and I would like to see their frequency against the explained variable - the Low weight. It can show us if when some categorical variable exists, it affects the babies' birth weight.

Frequency tables

Low weight	**Mothers' Race**			
	White	Black	Other	Row Total
0 (=no)	50	12	32	94
1 (=yes)	14	9	19	42
Column Total	64	21	51	136

From looking at the Mothers' Race frequency table, we can see that most low-birth-weight babies are from "other" race (by the numerical 19 babies are low weight in "other" race). But if we look at the number of black mothers that participate in the sample, we can see that there are only 21 black mothers, 51 "other" race mothers and 64 white mothers. So, if we look at the percentages, 43% among the black children were born underweight. Among the "other" race, only 37% of children were born underweight. Among the white children, only 21% of children were born underweight. I would like to prevent a multicollinearity so I will later remove part of the race variable.

Low weight	**Mothers' smoking status**		
	0 (=no)	1 (=yes)	Row Total
0 (=no)	64	30	94
1 (=yes)	21	21	42
Column Total	85	51	136

From looking at the above frequency table, we can see that for both cases, smoking and non-smoking mothers, there are 21 low birth weight babies. But there are 85 non-smoking mothers and 51 smoking mothers in total. So, if we look at the percentage, among the mothers that do not smoke, 24% gave birth to a low weight baby, and among the mothers that smoke, 41% gave birth to a low weight baby. From this data, we can assume that apparently there is an influence of the mother's smoking status on her baby's born weight.

Low weight	**History of hypertension**		
	0 (=no)	1 (=yes)	Row Total
0 (=no)	91	3	94
1 (=yes)	37	5	42
Column Total	128	8	136

From looking at the above frequency table, we can see that most of the mothers that have not a history of hypertension, gave birth to babies in normal weight (only 29% babies in low weight). But the number of participants in the sample, with a history of hypertension, is very low (only 8 mothers), and among the mothers that have a history of hypertension, 62.5% gave birth to a low weight baby. Maybe they should add observations of mothers with history of Hypertension, so the effect of this variable on birth weight will be clearer.

Low weight	**Uterine Irritability**		
	0 (=no)	1 (=yes)	Row Total
0 (=no)	81	13	94
1 (=yes)	34	8	42
Column Total	115	21	136

From looking at the above frequency table, we can see that only 21 mothers in the sample have uterine irritability, and 115 do not have uterine irritability. I look at the percentage, and we can see that among the mothers without uterine irritability, only 29% gave birth to a low weight baby, and among the mothers, with uterine irritability, 38% gave birth to a low weight baby.

Low weight	**Physician Visits during the first trimester**		
	0 (=no)	1 (=yes)	Row Total
0 (=no)	47	47	94
1 (=yes)	26	16	42
Column Total	73	63	136

From looking at the above frequency table, we can see that among the mothers that visited a physician during the first trimester of the pregnancy, only 16 out of 47 mothers gave birth to a low-birth-weight baby (34%), in comparison to mothers who did not visit a physician and gave birth to 26 out of 47 low weight babies (55%).

Low weight	**Previous Premature Labors**		
	0 (=no)	1 (=yes)	Row Total
0 (=no)	86	8	94
1 (=yes)	28	14	42
Column Total	114	22	136

From looking at the above frequency table, we can see that there are only 22 mothers with previous premature labors, and 114 without previous premature labors, so I will analyze the percentage. Among the mothers that have previous premature, 64% gave birth to a low weight baby, and among the mothers that did not have previous premature, only 24% gave birth low weight babies. So, we can assume that apparently this variable has an influence on the baby's birth weight and it probably explain it.

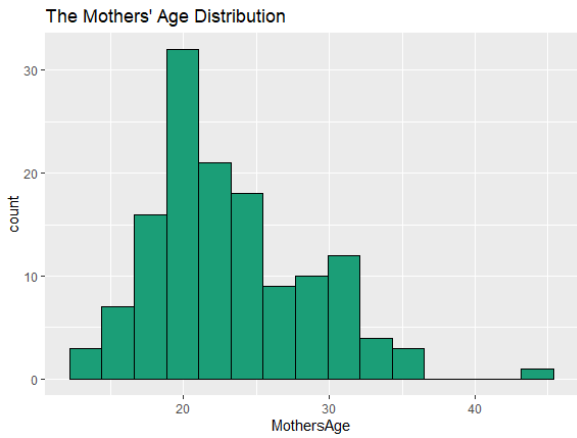
Now after I analyzed the frequencies, I have a hypothesis regarding which categorical variables probably explain the low weight.
We will look at the correlations between the continuous variables and try to conclude from them as well.

Correlation matrix of continuous variables

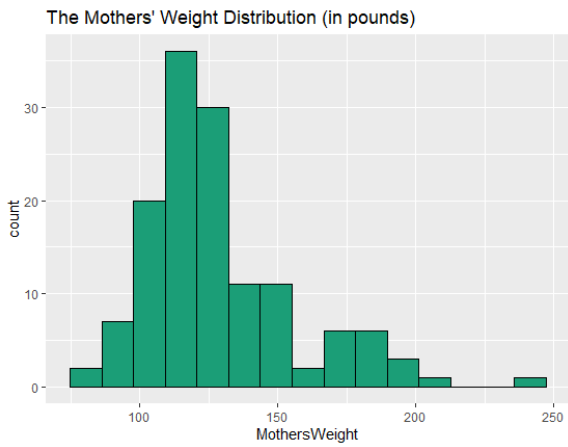
	MothersAge	MothersWeight	BirthWeight
MothersAge	1.000	0.185	0.028
MothersWeight	0.185	1.000	0.145
BirthWeight	0.028	0.145	1.000

We can see that there is a low correlation between all the continuous variables, but the lowest is the correlation between mothers' age and birth weight (0.028). It is a reasonable assumption because there is no reason that mother's age will affect the baby's birth weight. The correlation between mothers' weight and mothers' age is also low (0.18). Again, probably for the same reason, I do not expect to see a high correlation. The correlation between mothers' weight and babies' birth weight is only 0.145. I would expect that the correlation between the mothers' weight and babies' birth weight will be higher because usually, it is affected by the genes, but as we will see later, there are other explanations for low birth weight. Eventually, I can understand from this correlation matrix that the mothers' age and the mothers' weight probably do not explain the birth weight.

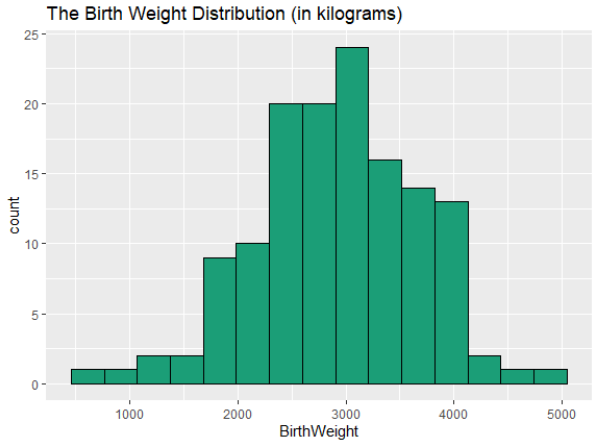
Marginal distributions of continuous variables



The Mothers' Age – we can see that most of the mothers are around 20 years old (the median is 22.5, mean is 23.25) but there are also older mothers. There is one mother that is 45 years old, which causes the chart to skew, and the histogram has a right tail because 45 is far from the average.



The Mothers' Weight – we can see that most of the mothers' weight is around 100 pounds (median is 122, mean is 128.9), but we still can see a 180-200 pounds. There is one mother that weighs well above the average (241 pounds), which causes the graph to skew with a right tail.



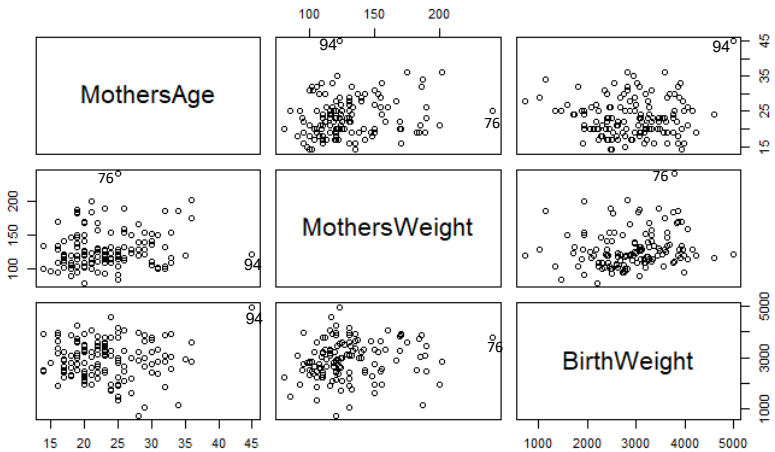
<u>The Birth Weight</u> – we can see that this distribution looks quite normal. Most of the weight is around 3000 grams (median is 2934 grams, mean is 2919 grams), but there are few below 1000 grams and above 4000 grams. Most of the birth weights are around the average but the birth weights are very diverse.

The observations that seem to be different from the rest:

	Low <dbl>	MothersAge <dbl>	MothersWeight <dbl>	Race <chr>
76	0	25	241	black
94	0	45	123	white

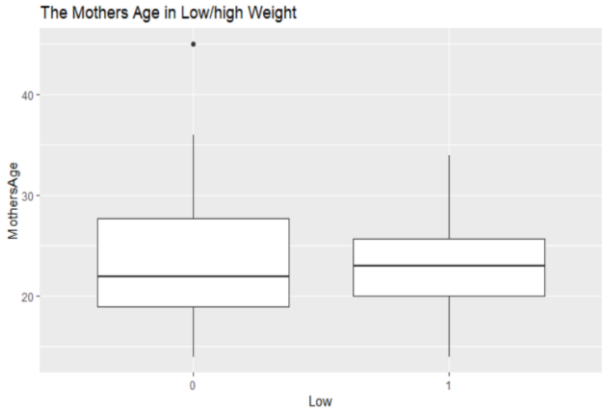
The mother's weight in observation number 76 is 241 pounds. The mother's age in observation 94 is 45 years old. We will see below whether the mothers' weight and mothers' age are included in the model, and whether those observations affect the model.

<u>The Scattering of the distributions</u>

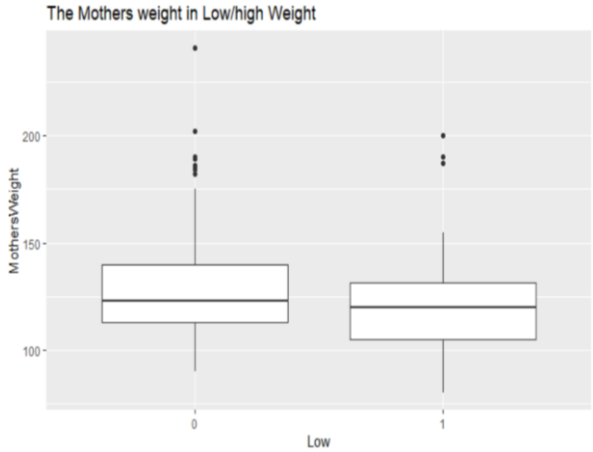


If we look at the distributions' scattering, we can see that generally, the scatter is normal between those three variables (besides the observations that I mentioned earlier). So, the sample of the mothers is quite diverse.
We can see in the scatter plots the two observations that I noticed before.

The relation between the low weight variable and the continuous variables



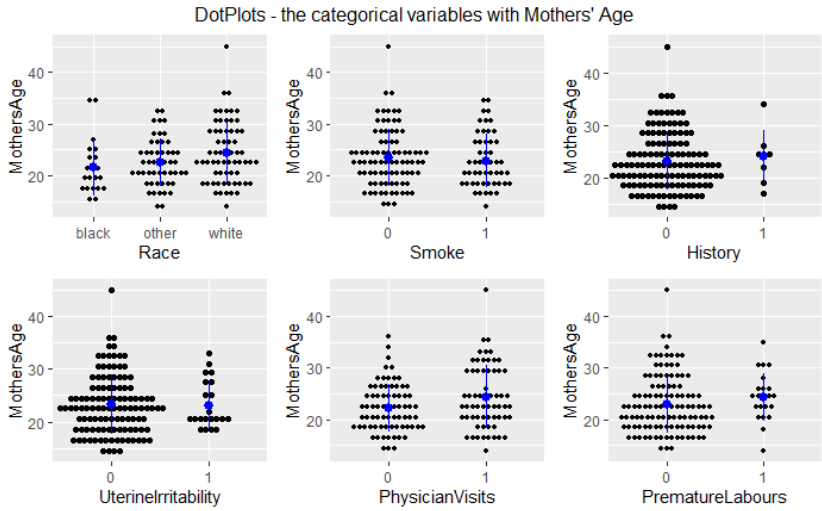
Mothers' Age –The median is around 22 (at low or normal birth weight). There is one observation that is over 40 (the 45 years old mother). We can see that the age probably does not explain the low/normal birth weight because they are quite the same, and that makes sense (from the reason I mentioned earlier).



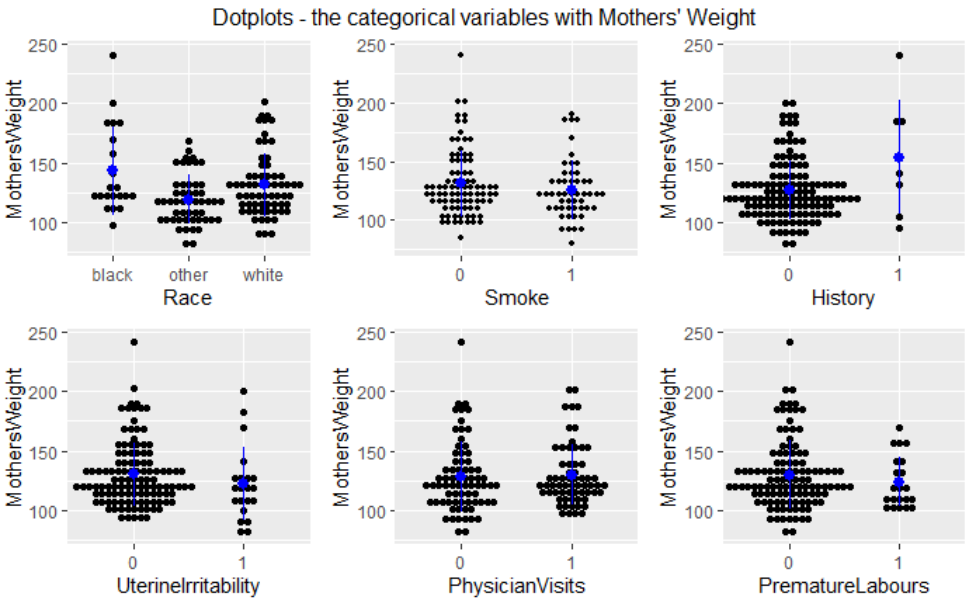
The Mothers' Weight -The average is around 125 pounds, and there are some observations above 0.75 of the samples (as we saw earlier, the mothers that weigh above 150 pounds). As we can see, there are mothers that weight above 150 pounds and gave birth to a baby with low/normal weight, so it also probably does not explain the low birth weight.

The Mothers' Age and Mothers' Weight with each categorical variable:

I chose to use Dot Plots. As I mentioned before, the number of the observations in each categorical variable is not the same (number of yes (1) ≠ number of no (0)) so, it is important to see the observation points more clearly and not to miss information.



From those Dot Plots of the categorical variables in front of the mothers' age, we can understand that the mother's age does not indicate other data about her. For example, the mothers' age does not affect whether the mother smokes or not, whether her race is black, white, or other, or whether the mother suffered from premature labor. The only thing that I can say is that the presence of uterine irritability appears at a young age (seeing from the dot plot that the median is around 22), and I do not have many observations of mothers with a history of hypertension, as I mentioned earlier in the frequency table. The median age of the black women in the sample is 20, and for the white women and other women, the median age is 25.



From the Dot Plots of the categorical variables in front of the mothers' weight, we can see that the there is no dramatic effect on the birth weight. In race, the white and other have a similar median but in the black it higher and skews. In addition, there are fewer observations of black mothers, contrary to white and other observations, as I mentioned before. In the white race, most of the observations are around the average and median, but there are a few exceptions that go above 0.75. If we look at the smoke variable, we can see that there is one exceptional observation in high weight (above 200 pounds) that does not smoke (maybe because of medical reasons). In addition, according to the observations I have, the mothers of high weight usually have not premature labors.

Linear Regression Model

First, I checked the linear model, with all the variables and later I did a Generalized linear model.

```
Call:
lm(formula = df_birth$Low ~ MothersAge + MothersWeight + Race +
    Smoke + History + UterineIrritability + PhysicianVisits +
    PrematureLabours, data = df_birth)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7539 -0.2774 -0.1547  0.3988  0.8955

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.533775   0.270199   1.975  0.05040 .
MothersAge            -0.001148   0.007300  -0.157  0.87530
MothersWeight         -0.001638   0.001526  -1.073  0.28510
Raceother             -0.051631   0.121851  -0.424  0.67249
Racewhite             -0.208105   0.116183  -1.791  0.07567 .
Smoke1                 0.182169   0.090038   2.023  0.04516 *
History1               0.305813   0.169300   1.806  0.07325 .
UterineIrritability1   0.057292   0.103942   0.551  0.58248
PhysicianVisits1      -0.032269   0.080828  -0.399  0.69040
PrematureLabours1      0.308590   0.105423   2.927  0.00406 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4319 on 126 degrees of freedom
Multiple R-squared:  0.1905,    Adjusted R-squared:  0.1327
F-statistic: 3.295 on 9 and 126 DF,  p-value: 0.001224
```

I saw that four variables are significant:
Mothers' smoking status during pregnancy, mothers' race (the white mothers), history of hypertension and premature labors.
The significance can show us that, apparently, these variables have the most impact on the low birth weight.

GLM - Generalized Linear Model

I decided to use the GLM because now I consider regression with a binary response, where the only possible values of the response Y are 0 or 1 (Low weight – yes (1) or not (0)).
I used the distribution function of the logit distribution, given by: $g(u) = \frac{e^u}{1+e^u}$

First, the low weight categorical variable and the birth weight continuous variable are giving the same information. I decided to remove the variable "Birth Weight" from the model because it is just an indicator of the birth weight variable. (with the birth weight continuous variable, the model will fit perfectly, and it creates multicollinearity because this variable gives us the exact same information).

We will look at the model with all the variables except "birth weight":

Full Model

(except continuous variable "birth weight")

```
Call:
glm(formula = Low ~ MothersAge + MothersWeight + Race + Smoke +
    History + UterineIrritability + PhysicianVisits + PrematureLabours,
    family = binomial(link = "logit"), data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7261  -0.7791  -0.5260   0.9782   2.1567

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.687016   1.538533   0.447  0.65521
MothersAge          -0.018399   0.044347  -0.415  0.67822
MothersWeight       -0.009912   0.008887  -1.115  0.26472
Raceother           -0.235179   0.635964  -0.370  0.71153
Racewhite           -1.250884   0.641191  -1.951  0.05107 .
Smoke1               1.108868   0.522130   2.124  0.03369 *
History1             1.525710   0.886446   1.721  0.08522 .
UterineIrritability1 0.326618   0.547950   0.596  0.55113
PhysicianVisits1    -0.163409   0.450374  -0.363  0.71673
PrematureLabours1    1.445181   0.537025   2.691  0.00712 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 168.14  on 135  degrees of freedom
Residual deviance: 140.93  on 126  degrees of freedom
AIC: 160.93

Number of Fisher Scoring iterations: 4
```

In the full model (without the continuous variable birth weight), many variables are not significant, and the AIC in the model is 160.93.
When we perform a binary regression analysis, we must check for redundancy in the explanatory variables, to prevent multicollinearity.
Therefore, we must find out which variables are unnecessary.
From looking at the significance we can see that probably the mothers' weight and age are unnecessary in the model, and only the white race is significant. Also, the presence of uterine irritability and physician visits are probably not explaining the low birth weight.

To test which variables do not explain the low weight, and whether we remove them, this model will be better, I used the back-ward method, and removed one variable any time (the variable with the lowest deviance). This method prevents overfitting (from removing too many variables) because it stops when the AIC is lowest before it raises again. The backward method omits the unnecessary variables:
First, it omits the physician visits variable (AIC became 159.06), second, the "black" + "other" race (AIC became 157.16). Then the Mothers' age (AIC became 155.43), the uterine irritability (AIC became 153.78) and the Mothers' weight (AIC became 153.26).
The reason that we keep only the white variable (or in the same way we can keep the black and other variables and omit the white because they complement). It makes sense because when we keep only the white variable, it prevents multicollinearity.

The model after removing those variables:

The new Model

```
Call:
glm(formula = Low ~ White + Smoke + History + PrematureLabours,
    family = binomial(link = "logit"), data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8514  -0.7501  -0.5859   1.0569   2.2134

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -1.1345     0.3119  -3.638 0.000275 ***
White1             -1.2248     0.4962  -2.468 0.013576 *
Smoke1              1.2349     0.4900   2.520 0.011724 *
History1            1.2003     0.8010   1.499 0.134002
PrematureLabours1   1.4147     0.5189   2.726 0.006404 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 168.14  on 135  degrees of freedom
Residual deviance: 143.26  on 131  degrees of freedom
AIC: 153.26

Number of Fisher Scoring iterations: 4
```

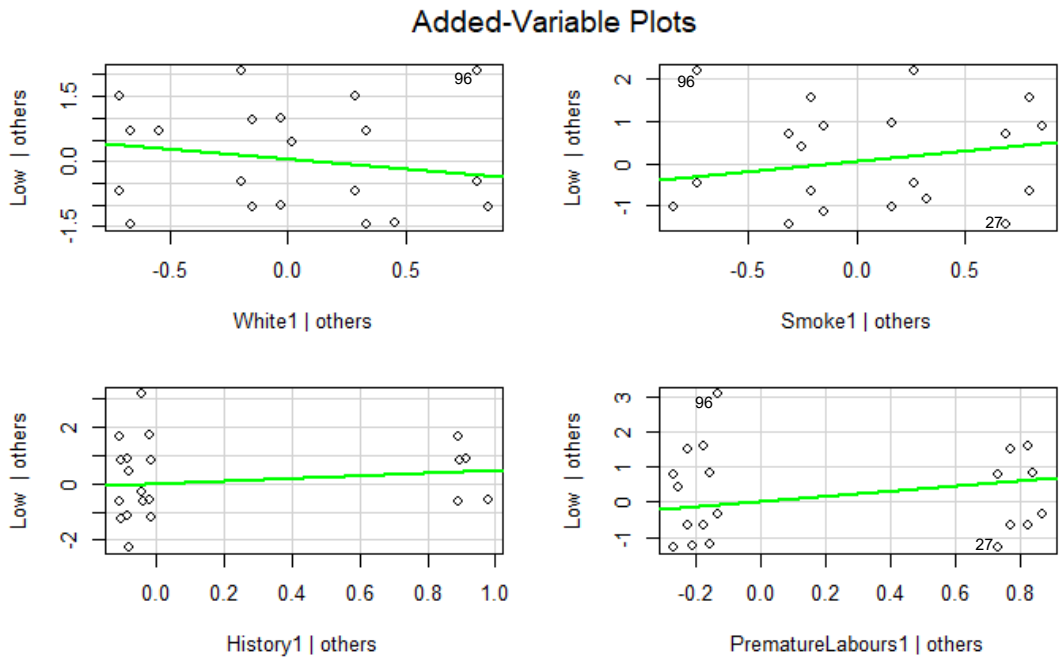
The old Model

(includes all the variables)

```
Call:
glm(formula = Low ~ MothersAge + MothersWeight + Race + Smoke +
    History + UterineIrritability + PhysicianVisits + PrematureLabours,
    family = binomial(link = "logit"), data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7261  -0.7791  -0.5260   0.9782   2.1567

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           0.687016   1.538533   0.447  0.65521
MothersAge           -0.018399   0.044347  -0.415  0.67822
MothersWeight        -0.009912   0.008887  -1.115  0.26472
Raceother            -0.235179   0.635964  -0.370  0.71153
Racewhite            -1.250884   0.641191  -1.951  0.05107 .
Smoke1                1.108868   0.522130   2.124  0.03369 *
History1              1.525710   0.886446   1.721  0.08522 .
UterineIrritability1  0.326618   0.547950   0.596  0.55113
PhysicianVisits1     -0.163409   0.450374  -0.363  0.71673
PrematureLabours1     1.445181   0.537025   2.691  0.00712 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 168.14  on 135  degrees of freedom
Residual deviance: 140.93  on 126  degrees of freedom
AIC: 160.93

Number of Fisher Scoring iterations: 4
```

As we can see, the new model explains the low weight better, most of the variables are significant, and we can see that the AIC is now 153.26. The white race variable seems to reduce the probability that a low-birth-weight baby will be born. In the new model - smoking, history, and premature labors increase the probability of a low birth weight.

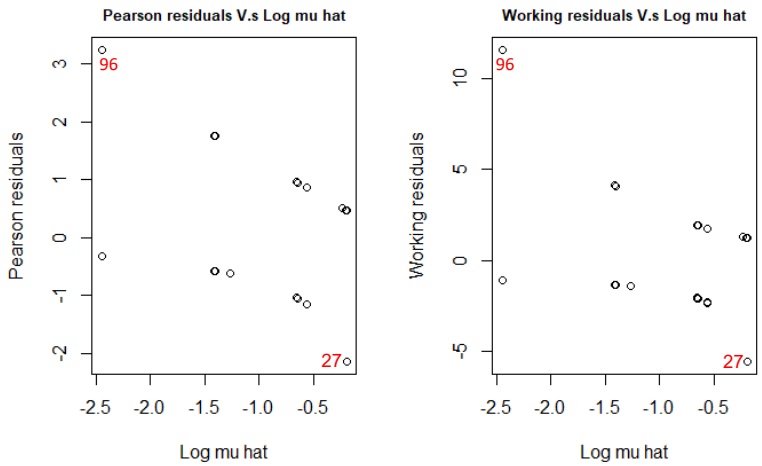
Now, we will look at the scattering of the explained variable in front of each explanatory variable. This plot identifies the points with the largest residuals and the points with the largest partial leverage. we would like to see which observation would change the slope.



We can see that the exceptional observations that modify the trend of the regression line are 96 and 27.
Next, we will check the residuals graphs.

Residuals

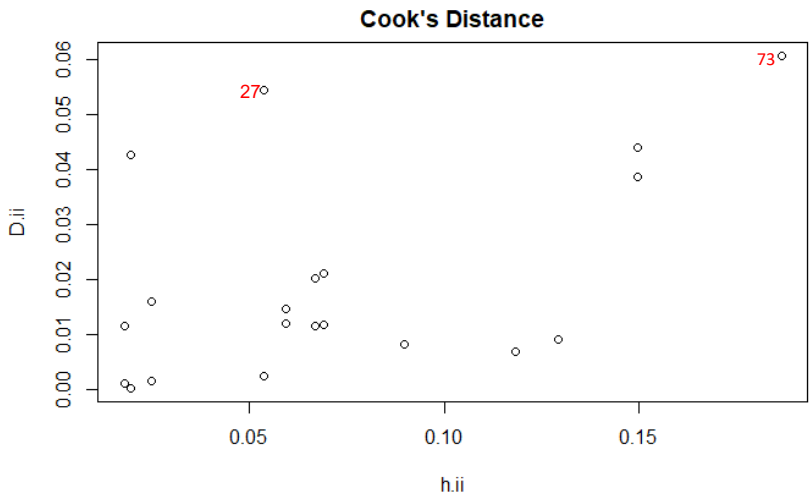
I checked the residuals because the residuals may show us if there are exceptional values requiring checking if there are outliers that disturb the model.
I compared between the Pearson residuals versus the Log mu hat, and the Working residuals versus the Log mu hat. I used a log mu hat because I wanted to spread out the points.



We can see that observation number 96 repeats as an outlier in those graphs. It is not certain that 27 is an outlier, and we will check it in the cook's distance.

Cook's distance

The Cook distance finds influential outliers in a set of predictor variables. It is a way to identify points that negatively affect the model.
We would like to see which observations in the model can Interfere with the model. I marked the observations above 0.05 as outliers.



In the cook's distance, the observations that are outliers are 73 and 27.

Observations suspected to be outliers from the tests I did:

96:

Low <dbl>	MothersAge <dbl>	MothersWeight <dbl>	Race <chr>	Smoke <chr>	History <chr>	PrematureLabours <chr>	BirthWeight <dbl>
1	29	130	white	0	0	0	1021

This mother is a white non-smoking mother. She had no history of hypertension or premature labors. However, despite that, she gave birth to a low weight baby. All the signs according to my model, were predicting that she will give birth to a baby at a normal weight. Probably, that is the reason it is an outlier.

73:

Low <dbl>	MothersAge <dbl>	MothersWeight <dbl>	Race <chr>	Smoke <chr>	History <chr>	PrematureLabours <chr>	BirthWeight <dbl>
0	19	184	white	1	1	0	3756

This mother is a white smoking mother. She had a history of hypertension, and despite that, she gave birth to a normal weight baby. She does not have premature labor (a variable that I saw that affects a lot on the predictive).

27:

Low <dbl>	MothersAge <dbl>	MothersWeight <dbl>	Race <chr>	Smoke <chr>	History <chr>	PrematureLabours <chr>	BirthWeight <dbl>
0	35	121	black	1	0	1	2948

This mother is a black smoking mother. She had premature labor. Although a high probability of a low weight baby, she gave birth normal weight baby.
She does not have a history of hypertension.
We saw earlier that this variable is not significant in the model.

I tried to remove each observation separately and see how it affects the model. I found out that omitting the observation number 96, improves the model the most.

GLM model without the outlier observation (96)

The new Model

(without the observation 96)

```
Call:
glm(formula = Low[no.96] ~ White[no.96] + Smoke[no.96] + History[no.96] +
    PrematureLabours[no.96], family = binomial(link = "logit"),
    data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9229  -0.7312  -0.7276   1.0659   1.7079

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.1826     0.3170  -3.730 0.000191 ***
White[no.96]1            -1.4207     0.5223  -2.720 0.006526 **
Smoke[no.96]1             1.4095     0.5125   2.750 0.005951 **
History[no.96]1           1.2288     0.8058   1.525 0.127248
PrematureLabours[no.96]1  1.4506     0.5245   2.766 0.005678 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.77  on 134  degrees of freedom
Residual deviance: 138.14  on 130  degrees of freedom
AIC: 148.14

Number of Fisher Scoring iterations: 4
```

The old Model

(includes the observation 96)

```
Call:
glm(formula = Low ~ White + Smoke + History + PrematureLabours,
    family = binomial(link = "logit"), data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8514  -0.7501  -0.5859   1.0569   2.2134

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.1345     0.3119  -3.638 0.000275 ***
White1            -1.2248     0.4962  -2.468 0.013576 *
Smoke1             1.2349     0.4900   2.520 0.011724 *
History1           1.2003     0.8010   1.499 0.134002
PrematureLabours1  1.4147     0.5189   2.726 0.006404 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 168.14  on 135  degrees of freedom
Residual deviance: 143.26  on 131  degrees of freedom
AIC: 153.26

Number of Fisher Scoring iterations: 4
```

The model now looks better. The significant variables became more significant and the AIC became lower (148.14).

We can also see that the History variable is not significant, but removing this variable, get a model that all the variables are significant perfectly (multicollinearity) and I want to prevent it.

I saw that if I omit the observation number 27 it also gives me an improvement, and I wanted to see what happens if I remove it from the model as well.

GLM model without the outlier observations (96 and 27)

The new Model

(without the observations 96 and 27)

```
Call:
glm(formula = Low[no.96_27] ~ White[no.96_27] + Smoke[no.96_27] +
    History[no.96_27] + PrematureLabours[no.96_27], family = binomial(link = "logit"),
    data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0795  -0.7299  -0.5361   0.9915   1.7102

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                -1.1867     0.3193  -3.717 0.000202 ***
White[no.96_27]1           -1.5975     0.5488  -2.911 0.003605 **
Smoke[no.96_27]1            1.5853     0.5373   2.950 0.003173 **
History[no.96_27]1          1.1853     0.8149   1.454 0.145828
PrematureLabours[no.96_27]1  1.6411     0.5438   3.018 0.002546 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.04  on 133  degrees of freedom
Residual deviance: 134.15  on 129  degrees of freedom
AIC: 144.15

Number of Fisher Scoring iterations: 4
```

The old Model

(includes the observations 96 and 27)

```
Call:
glm(formula = Low[no.96] ~ White[no.96] + Smoke[no.96] + History[no.96] +
    PrematureLabours[no.96], family = binomial(link = "logit"),
    data = df_birth)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9229  -0.7312  -0.7276   1.0659   1.7079

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.1826     0.3170  -3.730 0.000191 ***
White[no.96]1            -1.4207     0.5223  -2.720 0.006526 **
Smoke[no.96]1             1.4095     0.5125   2.750 0.005951 **
History[no.96]1           1.2288     0.8058   1.525 0.127248
PrematureLabours[no.96]1  1.4506     0.5245   2.766 0.005678 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.77  on 134  degrees of freedom
Residual deviance: 138.14  on 130  degrees of freedom
AIC: 148.14

Number of Fisher Scoring iterations: 4
```

The model without the two observations is even better, the AIC get improved, and the variables are more significant.
Because the sub-data have 136 observations, It possible to omit two observations if it improves my model.
I decided that the last model without the two observations is the best-predicting model I can get with my observations.

Conclusion

In this analysis, I tried to find out which variables influence the most on the babies' birth weight. I tried to fit the model that will include the explanatory variables and the observations that predict well the explained variable. I looked at the significance and the AIC and checked the outlier observations that perhaps if I remove them, the model will be a better predicting. I found out that observation number 96 is an outlier in the residual graphs, and 27 is an outlier in the Cook distance graph. I decided to remove them from the model.
I wanted to see what would happen if I removed the observations from the beginning, maybe the variables were different, and the model has been changed. After I checked it, I found out that even if I omit those observations from the beginning, I get the same model and the same explanatory variables.
Now I would like to compare the coefficients of the first full model and the fit model:

The Coefficients of estimates in the full model:

Intercept	Mothers' Age	Mothers' Weight	White	Black	Smoke	History	Uterine irritability	Physician visits	Premature labors
0.451	-0.018	-0.009	-1.015	0.235	1.108	1.525	0.326	-0.163	1.445

The Coefficients of estimates in the fit model:

Intercept	White	Smoke	History	Premature Labors
-1.186	-1.597	1.585	1.185	1.641

We can see that the model improved the prediction in the variables that were chosen for the fit model. All those variables that have a positive prediction, increased (Smoke, History, and premature labors) and negative predicted, decreased (White) – which shows that the model improved.
In Addition, the AIC decreased by 16.78 from the first model (160.93) to the last model (144.15) and I took the AIC as a measure of the good fitting.
Finally, the improvement in the significance and prediction of low birth weight shows that a suitable model has been found.