Identify risk factors associated with giving birth to a low birth weight baby

Hanchao Zhang

Meng Lan

Muming Tang

Professor Yizhe Zhao

Dec. 6th 2017

Abstract

The main purpose of the paper is to find the risk factors for low birth weight. The potential risk factors are mother's age, weight, race, the smoking status, whether the mother has premature labor before and the number of having premature labor, history of hypotension, presence of uterine irritability and the number of physician visits.

Linear logistic regression is performed to do univariate and multivariate analysis. Automatic variable selection procedure and general variable selection procedure are both implemented for comparison in the study. The former is to use Wald test and AIC for backward selection, and the latter is to use deviance for selection. In each model, the goodness-of-fit test is implemented. After checking the residuals, the goodness-of-fit test is implemented again.

The result suggests that the model selected by deviance is the best model. According to that model, the identified risk factors are history of premature labor, history of hypertension, presence uterine irritability, race and smoking status. Considering interaction, it is found that black mothers who smoke have a significantly higher risk compared to white mothers who do not smoke, when other factors held constant.

If the data with more observation can be accessed, the model will be more elaborate and if the especially when the data is not sparse in some categories. For exploratory analysis, a furthering study and analyze is also needed.

Key words: Logistic regression, Deviance, AIC, Wald Chi-square test

Introduction

The weight of a new born baby is a significant factor that affects the health status and development of a child. Those who were born with the weight under 2500 grams are called low birth weight children. Low birth weight may exert negative impacts on the intelligence and social adaptability of children. In addition, the mortality rate is higher in low birth weight babies than the normal weight group. (Su, 2013). "In both developed and developing countries, birth weight is probably the single most important factor that affects neonatal mortality, in addition to being a significant determinant of post-neonatal infant mortality and of infant and childhood morbidity". (Kramer,1987, pp663).

Therefore, studying the risk factors that could contribute to low birth weight becomes increasingly important. According to Kleinman and Kessel (1987), age, parity, marital status and education have different extents of effects on low birth rate among white race and black race. Dennis and Mollborn (2014) study how maternal age and ethnicity affect low birth weight and find that because of the difference in social and behavioral factors, low birth weight risk is not uniform in different ages or races. Meis, Ernest and Moore (1987) study how idiopathic premature labor affects low birth weight in both public and private patients. They find that idiopathic premature labor is more related to private low birth weight births (47.1%) than public births(24.8%). Kleinman and Madans (1985) find that, keeping other variables constant, women who have higher educational achievements are less likely to have babies with low birth weight. Those who receive education with less than 12 years have 2.38 times greater risk to have a low birth weight than those who receive more than 13 years of education. They also find that smoking is a crucial factor that increase the low birth rate. If all women could stop smoking during pregnancy, low birth rate will reduce by more than 10%. There is also another study showing that the low birth weight may cause hypertension in the adulthood in South Asian Region. (Ediriweera, Dilina, Perera,Flores &Samita,2017)

The object of this project is to find risk factors that have an association with low birth weight. Univariate and multivariate logistic regression models have been applied to identify the relationship between them. Then, the most suitable model is selected, followed by model checking.

Methodology

Variables

The criteria for low birth weight is 2500 grams. Those babies weighted under 2500 grams were categorized as low birth weight babies, and those above 2500 grams were categorized as normal birth weight. Race and Ethnicity were divided into three categories, white, black and others.

History of premature labor is included in our data. Premature labor is the primary reason for low birth weight *(J Nat Sci Biol Med. 2010 Jul-Dec; 1(1): 40–42.)*. In our data, instead of the premature labor, we use the history of premature labor instead. As we have already known that premature usually leads to not fully developed babies as well as low birth weight, gathering the premature labor in our data is not that meaningful as gathering the history of premature labor which could give us a sign of whether the past premature labor will influent the birth weight in the furthering delivery or not. In our data, the premature labor is very sparse, so it is important to regroup the data to new category that can fit the model better as well as make sense in real life. In our model, the premature labor is regrouped as unordinary categorical data, those who have had a premature labor experience before and those who have not had any premature labor experience.

History of Hypotension and presence of uterine irritability are two dichotomous variables in our data. We record the history of hypotension as yes or no. Hypotension of the mother during the pregnancy is associated with low birth weight *(BMJ. 2004 Dec 4; 329(7478): 1312.)*. We recorded the hypotension variable as yes whatever the mother of the baby had hypotension before the pregnancy or during the pregnancy. Meanwhile, the presence of uterine irritability can also lead to low birth weight since it could lead to premature labor which is, mentioned above, a primary reason for low birth weight.

Another dichotomous variable is smoking status. We only record the status as smoker or non-smoker. This means that whether a person stops smoking during the material period or not, the smoking status will be recorded as yes, a smoker, if they have previously smoked.

The times that the mother visits a physician is recorded in our data as "ftv". It is also a sparse data and needs to be reset to a new category. Since the time of visiting the physician six time is only 5 people in the data and the time of visit from 2 to 5 is also less than 30, meanwhile, it makes no sense if the data is regrouped to more than two levels in our data. The reason as follow, first the sparsity of the data needed to regroup, from the literature reviews, the group can be set to 0 as a visit, 1 as normal visit times, and 2 as not normal visit times, the clinical research shows that more than 6 times visit in the first three months of the labor is considered as too many visits. However, our dataset doesn't support such regroup method, because the number of 6 times visits mothers is too less, the compromise has to be made, the times of physician visits is, sadly, categorized by the visiting physician in first three month labor and not visit at all in the first three month labor.

Analysis plan

Demographic analysis

The first step is to do preliminary descriptive analysis and testing to characterize the variables and detect each variable's distribution difference in two groups where whether mothers delivering a low birth weight baby.

For continuous variables, mean and standard deviation are used as main statistics. As for testing, log transformation is needed first if the distributions of variables are not normal. Then a student's t test or non-parametric test like Wilcoxon signed-rank test is used depending on whether the distributions are normal or not.

For categorical variables, a contingency table is made for each variable to display the distribution, and a chi-square test is used to test the difference under the event and under the non-event.

Univariate analysis in logistic regression

In the second step, a one-variable logistic regression model is built for each factor to model the probability of delivering a low birthweight baby. The coefficient and the odds ratio from the result indicates the relationship between the covariate and dependent variable.

Multivariate analysis in logistic regression

After the univariate analysis, the mutual effect of covariates is taken into consideration by using multivariable logistic regression. Three models by using various model selection criteria are compared to find the potential risk factors.

Using the deviance as criterion, a general variable selection procedure is employed. First, compare the deviance of one-variable logistic model to that from the null model logit(p)=β_0 and find variables that significantly reduce the deviance. Set the threshold for the significant increase in deviance to be 2.07, $\chi^2(1, 0.15)$. Fit these variables, compute the increase in the deviance when each variable on its own is omitted from the set, and retain the variables whose omission induces significant increase. Then add back those left variables in the first step to see any of them can reduce the deviance significantly. After the final check to get the main factor model, interaction was considered. The interaction that brought lowest deviance will be retained in the last model.

The second model was selected by the Wald Chi-Square test. The backward selection was implemented. the variable whose p-value larger than the threshold (0.15) will be dropped in the model step by step, for each step, only the variable with not significant and largest p value will be dropped. The interaction term with the significant p value will be retained in the model for further study.

The third model was selected by Akaike information criterion (AIC). The backward selection was implemented and the variable with largest AIC will be dropped in the model.

Model checking

The final step is to do model checking including to evaluate the goodness-of-fit of the model and identify outliers. Deviance goodness of fit test or Hosmer-Lemeshow goodness of fit test is used depending on replicated data or unreplicated data for each pattern of covariate. The observation whose delta deviance larger than 4 is identified as outlier and should be deleted.

Model Analysis

Demographic Analysis

1. Age (age of mother in years)

In the study, the mean age of mothers who have a low birth weight baby is about 22.3 and its standard deviation is 4.5. For mothers who do not have a low birth weight, the mean age and standard deviation are slightly higher, which are approximately 23.7 and 5.6 respectively.

From the density curves and qq-plots, we can see the kurtosis of the distribution of age of mothers having a low birth weight baby is slightly smaller than normal distribution. The distribution of age of mothers not having a low birth weight baby is a little right-skewed. By using the Kolmogorov-Smirnov test, we did not find sufficient evidence to detect a difference between the first age group and normal distribution. For the second one, the p-value is smaller than significance level so its distribution is significantly different from normal distribution.

After doing the log transformation for age in two groups, the distribution of age of mothers having a low birth weight baby is still significantly different from normal distribution using Kolmogorov-Smirnov test. For data not normally distributed, we use Wilcoxon rank-sum test to test the location difference of distributions of age in two groups. The p-value is 0.247, larger than the significance level 0.15, so no sufficient evidence is found to suggests the locations of these two distributions are different.

2. Lwt (weight at the last menstrual period)

The mean weight of mothers having a low birth weight baby is about 122.1 and its standard deviation is around 26.6. Comparing to that, mothers not having a low birth weight baby tend to have higher weight mean and deviance which are 133.3 and 31.71 respectively.

From the density curves and qq-plots, the distributions of mothers' weight in both groups are right-skewed and have an obvious peak. By doing two Kolmogorov-Smirnov tests, with p-values smaller than the significance level (0.05), both distributions are significantly different from normal distribution.

After doing the log transformation for weight in two groups, the distribution of weight of mothers not having a low birth weight baby is still significantly different from normal distribution using Kolmogorov-Smirnov test. While for mothers having a low birth weight baby, we do not find sufficient evidence to detect a difference between this distribution and normal distribution. Wilcoxon rank-sum test is also used to test the location difference of distributions of in two groups. The p-value of Wilcoxon test is 0.013, which indicates these two distributions are different.

Table 1. demographic analysis

Effect	Low birth weight (1)	Normal birth weight (0)	P-value
Count	59	130	
Age (mean(sd))	23.66 (5.58)	22.31 (4.51)	0.247
Weight (mean(sd))	122.14 (26.56)	133.3 (31.72)	0.013
Race			0.08
White (1)	23	73	
Black (2)	11	15	
Other (3)	25	42	
Smoking Status			0.03
Smoker (1)	30	44	
Non-smoker (0)	29	86	
Premature labor			0.0002
Yes (1)	18	12	
No (0)	41	118	
Hypertension			0.036
Yes (1)	7	5	
No (0)	52	125	
Uterine irritability			0.02
Yes (1)	14	14	
No (0)	45	116	
Physician visits			0.133
Yes (1)	23	66	
No (0)	36	64	

3. Race

The frequency counts and percent of race in each group are shown in the 3 by 2 contingency table. For categorical variable with a large sample, we use Chi-Square test and the p-value is 0.08, which means that at the significance level of 0.15 race and whether to deliver a low birth weight baby are not independent.

4. Smk (smoking status)

The frequency counts and percent of smoking status in each group are shown in the contingency table. Using the Chi-Square test and the p-value is 0.03, which means that at the significance level of 0.15 smoking status and whether to deliver a low birth weight baby are associated.

5. Ptl (history of premature labor)

The frequency counts and percent of history of premature labor in each group are shown in the contingency table. The times mothers had premature labor is from 0 to 3. Because the original table is sparse so we regroup the times of premature labor into two levels: 0 and 1. 1 denotes that mothers had premature labor no matter how many times. 0 denotes that mothers never had premature labor.

Using the Chi-Square test and the p-value is smaller than 0.15, which means that at the significance level of 0.05 history of premature labor and whether to deliver a low birth weight baby are associated.

6. Ht (history of hypertension)

The frequency counts and percent of history of hypertension in each group are in the contingency table. Using the Chi-Square test and the p-value is 0.036, which means that at the significance level of 0.15 history of hypertension and whether to deliver a low birth weight baby are associated.

7. Ui (presence of uterine irritability)

The frequency counts and percent of presence of uterine irritability in each group are in the contingency table. Using the Chi-Square test and the p-value is 0.02, which means that at the significance level of 0.15 presence of uterine irritability and whether to deliver a low birth weight baby are associated.

8. Ftv (number of physician visits during the first trimester)

The frequency counts and percent of number of physician visits in each group are in the contingency table. The times mothers had physician visits is from 0 to 6. Considering the sparsity of the data, we regroup the number of physician visits into two levels: 0 and 1. 1 denotes mothers had physician visits no matter how many times. 0 denotes mothers never had physician visits.

Using the Chi-Square test and the p-value is smaller than 0.15, so whether mothers had physician visits and delivering a low birth weight baby is associated.

Univariate Analysis

Table 2. univariate analysis

Effect	Odds ratio	Confidence Interval	p value	Change in Deviance	AIC
age	0.95	0.893-1.011	0.1047	2.76	235.912
weight	0.986	0.974-0.998	0.0227	5.981	232.691
Race2 vs 1	2.327	0.938-5.772	0.0854	5.01	235.662
Race3 vs 1	1.889	0.955-3.735			
smk	2.022	1.081-3.783	0.0276	4.867	233.805
ptl	4.317	1.916-9.726	0.0004	12.774	225.898
ht	3.365	1.021-11.088	0.0461	4.022	234.65
UI	2.578	1.139-5.834	0.0231	5.076	233.596
ftv	0.62	0.331-1.159	0.1339	2.587	236.394

0.15 is assumed to be the significance level.

Age: From table 2, it can be seen that the p-value is 0.1047, less than the significance level (0.15), which means that age has an impact on low birth weight. The likelihood of delivering a low birth weight baby will decrease on average by 0.05 when age increases for one unit. We also checked the residuals of the regression and added the power of the age in the model to see whether there is a relationship between the higher order of the age. It turns out higher-order term is not significant so there is no need to consider it.

Weight at the last menstrual period: It can be seen that the p-value is 0.0227, less than the significance level, which means that weight at the last menstrual period has an impact on low birth weight. The likelihood of delivering a low birth weight baby will decrease on average by 1.4% when mother's weight at last menstrual increases for one pound.

Race: It can be seen that the p-value of the joined test of race is 0.0854, smaller than the significance level, which means that there is an association between race and low birth weight.

Smoking: It can be seen that the p-value is 0.0276, less than the significance level, which means that smoking is a factor that influences the low birth weight. The likelihood of smokers delivering a low birth weight baby over the likelihood of non-smokers delivering a low birth weight baby is 2.022. The 95% confident interval is [1.081, 3.783], which also indicates that smoking has a significant positive impact on low birth weight.

History of premature labor: It can be seen that the p-value is 0.0004, less than the significance level, which means that the history of premature labor is a factor that influences the low birth weight. The likelihood of mothers with history of premature labor delivering a low birth weight baby over the likelihood of mothers without history of premature labor delivering a low birth weight baby is 4.317. The 95% confident interval is [1.916, 9.726], which also suggests that history of premature labor has a significant positive impact on low birth weight.

History of hypertension: It can be seen that the p-value is 0.0461, less than the significance level, which means that history of hypertension is a factor that influences the low birth weight. The likelihood of mothers with history of hypertension delivering a low birth weight baby over the likelihood of mothers without history of hypertension delivering a low birth weight baby is 3.365. The 95% confident interval is [1.021, 11.088], which means that history of hypertension has a positive impact on low birth weight.

Presence of uterine irritability: It can be seen that the p-value is 0.0231 and it is less than the significance level, which means that presence of uterine irritability is a factor that influences the low birth weight. The likelihood of mothers with presence of uterine delivering a low birth weight baby over the likelihood of mothers without presence of uterine irritability delivering a low birth weight baby is 2.578. The 95% confident interval is [1.139, 5.834], which means that presence of uterine irritability has a positive impact on low birth weight.

Physician visits during the first trimester: It can be seen that the p-value is 0.1339 and it is less than the significance level, which means that visiting physicians is a factor that influences the low birth weight. The likelihood of mothers with physician visits during the first trimester delivering a low birth weight baby over the likelihood of mothers without physician visits delivering a low birth weight baby is 0.62, which indicates that more physician visits lower the risk of having a low birth weight baby.

Multivariate Analysis

Deviance Method

From the univariate analysis, set 0.15 as the significance level, the result in Table 2 suggests that all the eight variables are considered to significantly reduce the deviance so they are all fitted together for later analysis. After computing the increase in the deviance when each variable on its own is omitted from the set (Table 3 in appendix), those variables whose omission induces significant increase in the deviance are shown in Table 4. Adding back any other variables does not lower the deviance so it is the main factor model. The result suggests that finally the variables retained in the model are race, smoking status, history of premature labor, history of hypertension and presence of uterine irritability, which means these variables are considered to have a significant association with whether mothers will deliver a low birth weight baby.

Table 4. main factor model using deviance selection

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	joint test Chi-Square	Pr > ChiSq
Intercept	0.6688	0.3954	2.8612	0.0907		
race 2	0.3656	0.3076	1.4131	0.2345	7.2994	0.026
race 3	0.3292	0.2631	1.566	0.2108		
smk	0.4653	0.1959	5.644	0.0175	5.644	0.0175
ptl	0.5952	0.221	7.257	0.0071	7.257	0.0071
ht	0.685	0.3214	4.5428	0.0331	4.5428	0.0331
UI	0.4197	0.227	3.4173	0.0645	3.4173	0.0645

Considering interaction, it is found that the interaction between race and smoking status lowers the deviance the most after it is added into the main factor model (Table 5 in appendix). To better interpret the effect of interaction, 5 dummy variables are created to represent each level of the interaction item.

The p-value of the joint test of the dummy variables are less than the significance level, which means that the effect of smoking status to delivering a low birth weight baby significantly differs in race.

The adjusted odds ratio for history of premature labor shows that, holding other covariates constant, the likelihood of mothers who had premature labor delivering a low birth weight baby over the likelihood of mothers who never had premature labor delivering a low birth weight baby is 3.556. Mothers who had premature labor tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

Table 6. multivariate analysis using deviance selection

Effect	Coefficient	Odds Ratio	Confidence Interval	P-value
Ptl	0.6344	3.556	1.448-8.736	0.0057
Ht	0.7098	4.135	1.091-15.68	0.0368
UI	0.5276	2.872	1.122-7.354	0.0278
Race 1 & Smk 1	0.2721	17.849	2.189-145.546	0.4521
Race 2 & Smk 0	0.0697	14.578	1.472-144.356	0.8955
Race 2 & Smk 1	1.5194	62.128	5.573-692.658	0.0132
Race 3 & Smk 0	0.4317	20.938	2.581-169.882	0.2272
Race 3 & Smk 1	0.3169	18.665	1.744-199.778	0.5913

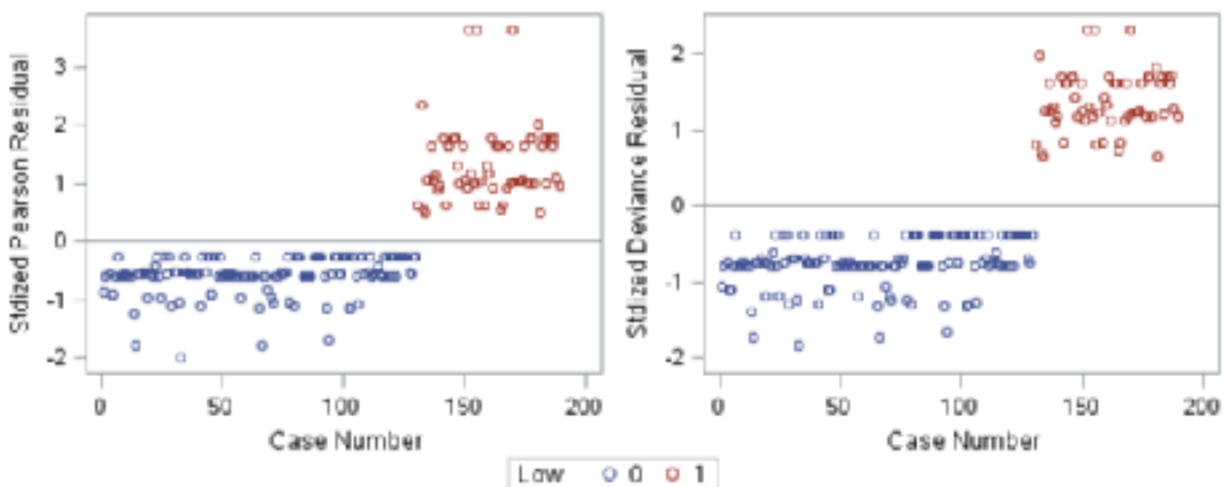
The adjusted odds ratio for history of hypertension shows that, holding other covariates constant, the likelihood of mothers who had hypertension delivering a low birth weight baby over the likelihood of mothers who never had hypertension delivering a low birth weight baby is 4.135. Mothers who had hypertension tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

The adjusted odds ratio for presence of uterine irritability shows that, holding other covariates constant, the likelihood of mothers with uterine irritability delivering a low birth weight baby over the likelihood of mothers without uterine irritability delivering a low birth weight baby is 2.872. Mothers with uterine irritability tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

For black mother who smoke, the odds of having a low birth weight baby is approximately 62 times the odds of white mothers who do not smoke, when other factors hold constant.

The last step is model checking. In this model, the data is replicated so the Deviance goodness of fit test is used. The p-value of is 0.54, larger than the significance level, which suggests that the current model dose not lack goodness-of-fit. The deviance standardized residuals are shown in Figure 1, and the observations whose value lager than 2 or less than -2 are deleted.

Figure 1. deviance standardized residuals and standardized Pearson residuals of cases for model selected by deviance



Wald Chi-Square test method

The second model is selected using backward selection by Wald Chi-Square test. In the first step, only the p-value of coefficient of age is 0.17, which is larger than the significance level, so it is dropped first. After the backward selection procedure, mother's age, race, smoking status, history of premature labor, history of hypertension and presence of uterine irritability are kept in the main effect model.

Table 7. multivariate analysis using Wald test selection

Effect	Coefficient	Odds Ratio	Confidence Interval	P-value
lwt	-0.0189	0.981	0.967-0.996	0.0112
smk	0.4672	2.546	1.130-5.737	0.0242
ptl	0.6039	3.346	1.367-8.191	0.0082
ht	0.9954	7.321	1.775-30.191	0.0059
race=1 & ui=1	-1.0616	1.137	0.287-4.058	0.0860
race=2 & ui=0	-0.1707	2.773	0.922-8.346	0.7213
race=2 & ui=1	2.4970	39.931	2.668-597.604	0.0282
race=3 & ui=0	-0.4531	2.090	0.797-5.481	0.2649
race=3 & ui=1	0.3781	4.798	1.161-19.837	0.5255

Besides, the interaction was also added by the Wald test. Since the interaction of race and presence of uterine irritability is the most significant comparing with other interactions, this interaction was added in the model. To specify the effect of interaction, 5 dummy variables were created representing different combination of race and presence of uterine irritability and using white mothers who had no uterine irritability as reference level for comparison.

The adjusted odds ratio for weight shows that, holding other covariates constant, the likelihood of mothers delivering a low birth weight baby will decrease on average by 1.9% when weight increases for one pound. Mothers with lower weight tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

The adjusted odds ratio for smoking status shows that, holding other covariates constant, the likelihood of mothers who smoke delivering a low birth weight baby over the likelihood of mothers who never smoke delivering a low birth weight baby is 2.546. Mothers who smoke tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

The adjusted odds ratio for history of premature labor shows that, holding other covariates constant, the likelihood of mothers who had premature labor delivering a low birth weight baby over the likelihood of mothers who never had premature labor delivering a low birth weight baby is 3.346. Mothers who had premature labor tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

The adjusted odds ratio for history of hypertension shows that, holding other covariates constant, the likelihood of mothers who had hypertension delivering a low birth weight baby over the likelihood of mothers who never had hypertension delivering a low birth weight baby is 7.321. Mothers who had hypertension tend to have a higher risk of delivering a low birth weight baby when other conditions stay the same.

Comparing to white mothers without uterine irritability, white mothers who have uterine irritability increase the odds by 1.137 when other conditions are the same. For black mothers who has uterine irritability, the odds of having a low birth weight baby increases by 40 comparing to white mothers who has no uterine irritability when other factors held constant.

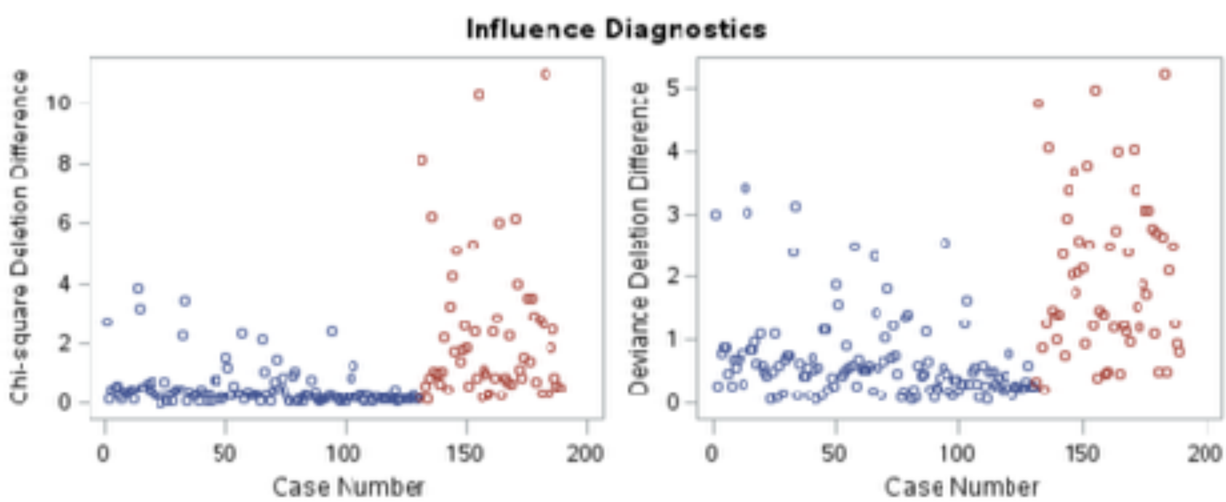


Figure 2. delta deviance of cases for model selected by Wald test

As for model checking, the data is replicated in this model so the Hosmer-Lemeshow goodness of fit test is used. The p-value is 0.6107, larger than the significance level, which suggests that the current model dose not lack goodness-of-fit. The delta deviance of cases is shown in Figure 2, and the observations whose value lager than 4 are deleted.

AIC

Table 8. multivariate analysis using AIC selection

Effect	Coefficient	Odds Ratio	Confidence Interval	P-value
lwt	-0.0147	0.985	0.972-0.999	0.0336
ptl	0.7141	4.172	1.710-10.179	0.0017
ui	0.3423	1.983	0.812-4.844	0.1331

In the third model selected using backward selection by Akaike information criterion (AIC), whether to deliver a low birth weight baby is significantly associated with mother's weight at the last menstrual period, history of premature labor and uterine irritability.

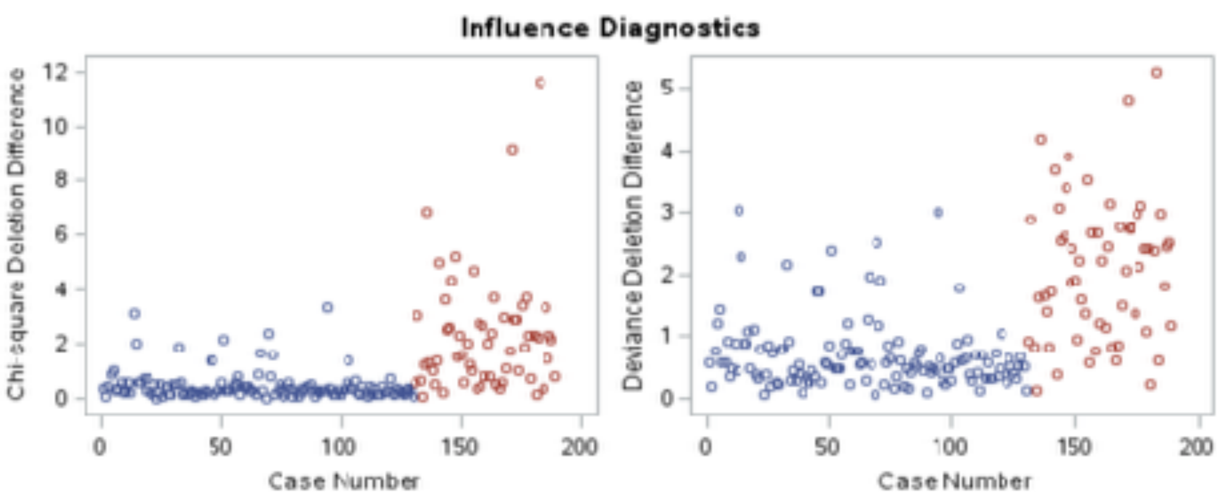
The adjusted odds ratio for weight says that, holding other covariates constant, the likelihood of delivering a low birth weight baby will decrease on average by 1.5% when weight increases for one unit. Mothers with lower weight results in a higher risk in delivering a low birth weight baby when other factors hold constant.

The adjusted odds ratio for premature labor shows that, holding other covariates constant, the likelihood of mothers who had premature labor delivering a low birth weight baby over the likelihood of mothers who never had premature labor delivering a low birth weight baby is 4.172. Mothers who had premature labor tend to have a higher risk of delivering a low birth weight baby when other factors hold constant.

The adjusted odds ratio for uterine irritability shows that, holding other covariates constant, the likelihood of mothers who had uterine irritability delivering a low birth weight baby over the likelihood of mothers who never had uterine irritability delivering a low birth weight baby is 1.983. Mothers who had premature labor tend to have a higher risk of delivering a low birth weight baby when other factors hold constant.

As for model checking, the data is replicated in this model so the Hosmer-Lemeshow goodness of fit test is used. The p-value is 0.46, larger than the significance level, which suggests that the current model dose not lack goodness-of-fit. The delta deviance of cases is shown in Figure 3, and the observations whose value lager than 4 are deleted.

Figure 3. delta deviance of cases for model selected by AIC



Conclusion

This study discusses how to find the risk factors for delivering a low birth weight baby by using three different criteria for selecting a proper model. Even though the models are quite different, as shown following, the interpretations of the effects are similar. In each model, the history of premature labor, the history of hypotension, the presence of uterine irritability are risk factors that increase the odds of delivering a low birth weight baby. For each model, since the variables was selected by different criteria, they have unique variables. Although more than one model can fit well, we choose the model selected by Deviance method, because this method covers fewest variables and all of them are significant with a smallest deviance. Moreover, when comparing the Deviance of the model selected by AIC and Wald Chi-square test, the Wald Chi-square model had the less Deviance. When comparing the AIC of the model selected by Wald Chi-square method and Deviance, they had same AIC value, which means, the model selected by the Deviance had the lowest Deviance and lowest AIC at the same time, better than the Model selected by AIC and Wald Chi-square.

Table 9. the last models using three selection criteria

variable	level	estimate (Deviance)	P > Chi Sq	estimate (Wald)	P > ChiSq	estimate (AIC)	P > ChiSq
Intercept		0.412	0.3593	3.549	0.0025	1.6321	0.4367
ptl	1	3.556	0.0057	0.6039	0.0082	0.7141	0.0017
ht	1	4.135	0.0368	0.9954	0.0059	-1.4773	0.4427
UI	1	2.872	0.0278			0.3423	0.1331
smk	1			0.4672	0.0242		
age						0.0424	0.6195
lwt				-0.0189	0.0112	-0.0147	0.0336
dummy	race = 1 and smk = 1	17.849	0.4521				
dummy	race = 2 and smk = 0	14.578	0.8955				

dummy	race = 2 and smk = 1	62.128	0.0132				
dummy	race = 3 and smk = 0	20.938	0.2272				
dummy	race = 3 and smk = 1	18.665	0.5913				
dummy	race = 1 and ui = 1			-1.0616	0.0860		
dummy	race = 2 and ui = 0			-0.1701	0.7213		
dummy	race = 2 and ui = 1			2.4970	0.0282		
dummy	race = 3 and ui = 0			-0.4531	0.2649		
dummy	race = 3 and ui = 1			0.3781	0.5255		
age*ht						0.1078	0.2042

Moreover, The deviance model which containing the less variable might cost research less money to gather the data and also less energy to process the data.

Discussion

After selection by deviance, it is concluded that low birth weight has an association with history of premature labor, history of hypertension and presence of uterine irritability, race and smoking status. This conclusion is consistent with many other studies. For example, Dennis and Mollborn (2014) find that low birth weight differs in different races. Ediriweera, Dilina, Perera,Flores &Samita(2017) find that there is an association between low birth weight and hypertension. However, there are some insufficiency in the model, since we do not have multiple stratifications of some variables. We just classify them by two categories: yes or no, which is too general and unclear. For instance, smoking status may have different categories: very heavy smokers, heavy smokers, light smokers and non-smokers. History of hypertension is also not clear when just using yes or no, because the hypertension before pregnancy or during pregnancy may be different. In addition, other causes apart from the eight risk factors may exert impacts on low birth weight such as socioeconomic disadvantages and different educational levels of women. This means that we need to read more related articles to deeply understand the relationship between low birth weight and different factors. Furthermore, other statistical analysis models should be applied and compared so that we can choose the most accurate model both theoretically and technically.

Reference

Dennis.J &Mollborn.S (2014) Young maternal age and low birth weight risk: An exploration
Of racial/ethnic disparities in the birth outcomes of mothers in the United States. *NIH Public Access Author Manuscript 50(4)*. 625-634

Ediriweera.DS &Dilina.N & Perera.U& Flores.F &Samita.S (2017) Risk of low birth weight on adulthood hypertension-evidence from a tertiary care hospital in a South Asian country, SriLanka: a retrospective cohort study BMC Public Health 17. 358

Kleinman.JC&Kessel.SS. (1987) Racial differences in low birth weight.Trends and risk factors.*The New England Journal of Medicine. 317(12)*.749-753

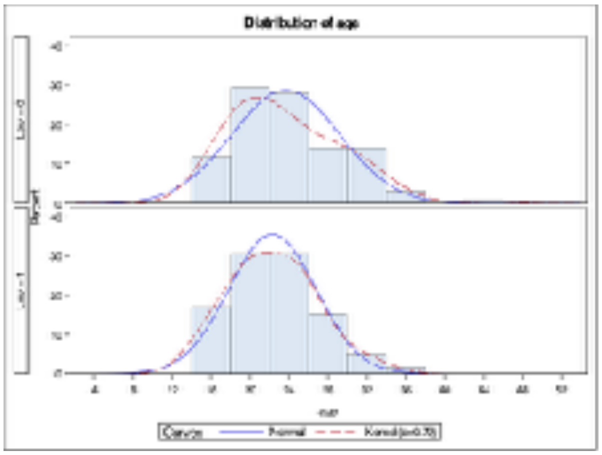
Kleinman JC & Madans JH(1985) The effects of maternal smoking, physical stature, and educational attainment on the incidence of low birth weight. *American Journal of Epidemiology 121(6)* 843-855

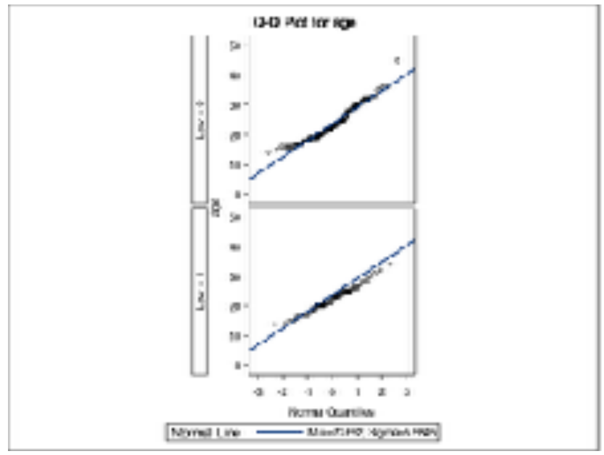
Kramer (1987). Reviews Analyses. *Bulletin of the World Health Organization.65(5)*.663-737

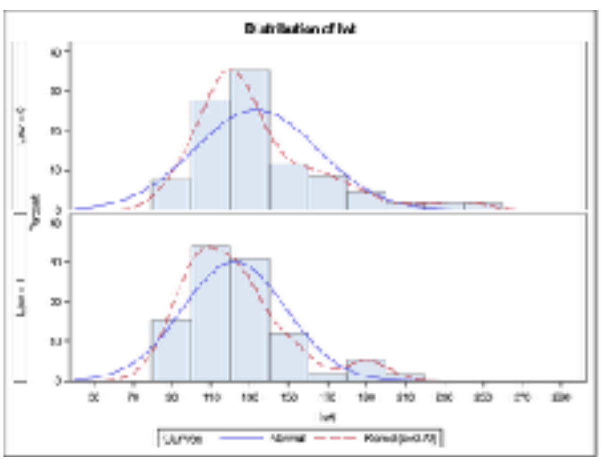
Meis PJ,Ernest JM& Moore ML(1987) Causes of low birth weight births in public and private patients. *American Journal of obstetrics and gynecology 156(5)* 1165-1168

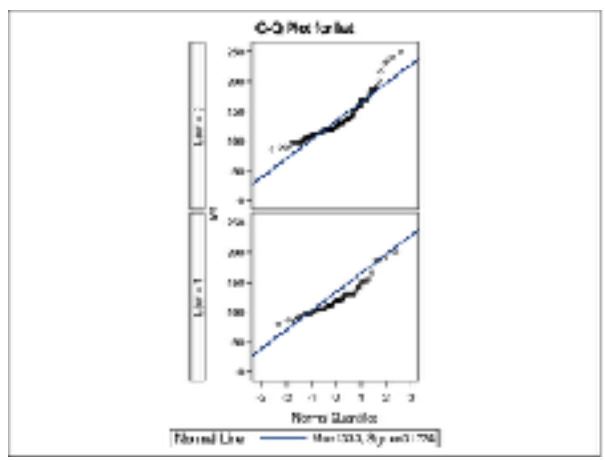
Su.D(2013) The study of risk factors on low birth weight. *China Medical Herald.28.60-62*

Appendix









Kolmogorov-Smirnov Test for Normal Distribution		
Group	**Statistic**	**p value**
Low = 1	0.088	>0.15
Low = 0	0.109	<0.01
Low = 1 (for log age)	0.090	>0.15
Low = 0 (for log age)	0.086	0.02

Kolmogorov-Smirnov Test for Normal Distribution		
Group	**Statistic**	**p Value**
Low = 1	0.146	<0.010
Low = 0	0.171	<0.01
Low = 1 (for log lwt)	0.104	0.114
Low = 0 (for log lwt)	0.124	<0.010

Table 3. deviance in variable omission

Main effect of the Deviance selection	**Deviance**	**Change of the Deviance**	**Pr > ChiSq**
full	196.7291		0.1044
age	178.8628	17.8663	0.1463
lwt	159.0722	37.6569	0.048
race	202.7575	-6.0284	0.0669
smk	200.5453	-3.8162	0.082
ptl	203.9375	-7.2084	0.0538
ht	203.7303	-7.0012	0.061
UI	198.9738	-2.2447	0.0857
ftv	194.0611	2.668	0.1303

Table 5. deviance of models including interaction

Interaction for Deviance selected model	**Deviance**	**DF**	**Value/DF**	**Pr > ChiSq**
main effect	19.7291	20	0.9865	0.475
race*smk	16.5984	18	0.9221	0.5509

race*ptl	19.2151	18	1.0675	0.3787
race*ht	19.4217	18	1.079	0.3663
race*UI	18.6123	18	1.034	0.4161
smk*ptl	19.232	19	1.0122	0.442
smk*ht	19.7017	19	1.0369	0.4127
smk*UI	19.3049	19	1.016	0.4374
smk*ptl	19.232	19	1.0122	0.442
ptl*ht	18.725	19	0.9855	0.4746
ptl*UI	17.9739	19	0.946	0.5242
ht*UI	19.7291	20	0.9865	0.475

SAS codes

Dup checking

```
libname sasfp
	"/folders/myshortcuts/SASUniversityEdition/study_design/final project";
run;

run;
data fp;
	set sasfp.data1;
	if ftv = 0 then ftv2 = 0;
	else ftv =1;
	if ptl = 0 then ptl = 0;
	else ptl = 1;
run;
```

```
proc sort data = fp
    nodupkey
    dupout = dups;
    by _all_;
run;
proc print
    data = dups;
run;


libname sasfp
        "/folders/myshortcuts/SASUniversityEdition/study_design/final project";
run;

run;


data fp;
        set sasfp.data1;
        if ftv = 0 then ftv2 = 0;
        else ftv =1;
        if ptl = 0 then ptl = 0;
        else ptl = 1;
run;


* full;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age lwt race smk ptl ht UI ftv / aggregate scale=none lackfit;
        output out=predict;
run;


*age;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=lwt race smk ptl ht UI ftv / aggregate scale=none lackfit;
        output out=predict;
run;
```

```
*lwt;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")=age race smk ptl ht UI ftv / aggregate scale=none lackfit;
	output out=predict;
run;


*race;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")=age lwt smk ptl ht UI ftv / aggregate scale=none lackfit;
	output out=predict;
run;

*smk;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")=age lwt race ptl ht UI ftv / aggregate scale=none lackfit;
	output out=predict;
run;

*ptl;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")=age lwt race smk ht UI ftv / aggregate scale=none lackfit;
	output out=predict;
run;

*ht;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")=age lwt race smk ptl UI ftv / aggregate scale=none lackfit;
	output out=predict;
run;

*UI;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
```

```
        model low(even = "1")=age lwt race smk ptl ht ftv / aggregate scale=none lackfit;
        output out=predict;
run;

*ftv;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age lwt race smk ptl ht UI / aggregate scale=none lackfit;
        output out=predict;
run;

* first round lwt out;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age race smk ptl ht UI ftv / aggregate scale=none lackfit;
run;

* age;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI ftv / aggregate scale=none lackfit;
run;

*race;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age smk ptl ht UI ftv / aggregate scale=none lackfit;
run;

*smk;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age race ptl ht UI ftv / aggregate scale=none lackfit;
run;

*ptl;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
```

```
        model low(even = "1")=age race smk  ht UI ftv / aggregate scale=none lackfit;
run;

*ht;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age race smk ptl UI ftv / aggregate scale=none lackfit;
run;

*UI;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age race smk ptl ht ftv / aggregate scale=none lackfit;
run;

*ftv;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")=age race smk ptl ht UI  / aggregate scale=none lackfit;
run;

*second round age out;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI ftv / aggregate scale=none lackfit;
run;

*race;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= smk ptl ht UI ftv / aggregate scale=none lackfit;
run;

*smk;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race ptl ht UI ftv / aggregate scale=none lackfit;
run;
```

```
*ptl;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ht UI ftv / aggregate scale=none lackfit;
run;

*ht;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl UI ftv / aggregate scale=none lackfit;
run;

*UI;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht  ftv / aggregate scale=none lackfit;
run;

*ftv;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI / aggregate scale=none lackfit;
run;

*third round ftv out;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI / aggregate scale=none lackfit;
run;

*race;
proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= smk ptl ht UI / aggregate scale=none lackfit;
run;

*smk;
proc logistic data=fp;
```

```
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race ptl ht UI / aggregate scale=none lackfit;
run;


*ptl;
proc logistic data=fp;
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race smk ht UI / aggregate scale=none lackfit;
run;

*ht;
proc logistic data=fp;
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race smk ptl UI / aggregate scale=none lackfit;
run;


*UI;
proc logistic data=fp;
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race smk ptl ht / aggregate scale=none lackfit;
run;


*Best model;

proc logistic data=fp;
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race smk ptl ht UI / aggregate scale=none lackfit;
run;


* interaction ;

proc logistic data=fp;
		class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
		model low(even = "1")= race smk ptl ht UI / aggregate scale=none lackfit;
run;
```

```
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI race*smk/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI race*ptl/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI race*ht/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI race*UI/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI smk*ptl/ aggregate scale=none lackfit;
run;


proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI smk*ht/ aggregate scale=none lackfit;
run;


proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(even = "1")= race smk ptl ht UI smk*UI/ aggregate scale=none lackfit;
run;


proc logistic data=fp;
```

```
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI smk*ptl/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI ptl*ht/ aggregate scale=none lackfit;
run;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI ptl*UI/ aggregate scale=none lackfit;
run;


*BEST;

proc logistic data=fp;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI race*smk/ aggregate scale=none lackfit influence;
        output out = deviance  difdev=difdev;
run;

data outliers;
set deviance;
if difdev < 4 and difdev > -4;
run;

proc logistic data=outliers order=data  plots = all;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
        model low(even = "1")= race smk ptl ht UI race*smk/ aggregate scale=none lackfit ;

run;


data outliers;
set deviance;
if difdev < 4 and difdev > -4;
if race = 1 and smk = 0 then dummy = "race = 1 and smk = 0";
if race = 1 and smk = 1 then dummy = "race = 1 and smk = 1";
```

```
if race = 2 and smk = 0 then dummy = "race = 2 and smk = 0";
if race = 2 and smk = 1 then dummy = "race = 2 and smk = 1";
if race = 3 and smk = 0 then dummy = "race = 3 and smk = 0";
if race = 3 and smk = 1 then dummy = "race = 3 and smk = 1";
run;

proc logistic data=outliers plots = all;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0") dummy(ref="race = 1 and smk = 0");
	model low(even = "1")= ptl ht UI dummy/ aggregate scale=none lackfit;

run;



libname sasfp
	"/folders/myshortcuts/SASUniversityEdition/study_design/final project";
run;

run;
data fp;
	set sasfp.data1;
	if ftv = 0 then ftv2 = 0;
	else ftv =1;
	if ptl = 0 then ptl = 0;
	else ptl = 1;
run;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")=age lwt race smk ptl ht UI ftv / HIER=single lackfit scale=none
		selection=forward  sls=0.15 sle=0.15 aggregate;
run;

* BEST main model;
proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
	output out = wald stdreschi = stdperason;
	title "best model";
run;
```

```
data adjusted;
set wald;
if stdperason < 2 and stdperason > -2;
run;


proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit ;
	title "interaction";
run;

*race interaction;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*lwt   / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*smk   / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*ptl   / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*ht   / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*UI   / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate
```

```
lackfit;
run;


*smk interaction;
proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI smk*ptl  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI smk*ht  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI smk*UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI smk*lwt  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

* ptl interaction;
proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ptl*lwt  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ptl*ht  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

proc logistic data=adjusted;
```

```
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ptl*UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;

run;




* ht interaction;
proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ht*ui  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;



proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ht*lwt  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;

run;



* ui interaction;
proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI ui*lwt  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;

run;


* Best model with interaction;


proc logistic data=fp plots = all;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*ui / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate
lackfit;
	output out = wald difdev = stdperason;
run;



*outliers;
data adjusted;
```

```
set wald;
if stdperason < 4 and stdperason > -4;

run;

proc logistic data=adjusted plots = all;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*ui / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;

run;




* Stratified with race;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
	output out = wald stdreschi = stdperason;
run;


data adjusted;
set wald;
if race = 1;
if stdperason < 2 and stdperason > -2;

run;


proc logistic data=adjusted;
	class smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0")  ;
	model low(event = "1")= lwt  smk ptl ht UI / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;



data adjusted;
set wald;
```

```
if stdperason < 2 and stdperason > -2;
if race = 2;
run;


proc logistic data=adjusted;
        class smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0") ;
        model low(event = "1")= lwt  smk ptl ht UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;

data adjusted;
set wald;
if stdperason < 2 and stdperason > -2;
if race = 3;
run;


proc logistic data=adjusted;
        class smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0")  ;
        model low(event = "1")= lwt  smk ptl ht UI  / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;



*stratified by ui;


data adjusted;
set wald;
if ui = 0;
if stdperason < 2 and stdperason > -2;

run;


proc logistic data=adjusted;
        class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0")  ftv(ref="0")  ;
        model low(event = "1")= race lwt  smk ptl ht / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;
```

```
data adjusted;
set wald;
if ui = 1;
if stdperason < 2 and stdperason > -2;

run;



proc logistic data=adjusted;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0")  ftv(ref="0")  ;
	model low(event = "1")= race lwt  smk ptl ht / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;



* stratified by age_group;



data adjusted;
set wald;
if stdperason < 2 and stdperason > -2;
if age < 20 then age = 1;
else if age < 34 then age = 2;
else age = 3;

if age = 1;
run;

proc logistic data=adjusted;
	class a race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0")  ftv(ref="0")  ;
	model low(event = "1")= race lwt  smk ptl ht / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
run;



*DUMMY;

proc logistic data=fp;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0");
	model low(event = "1")= lwt race smk ptl ht UI race*ui / HIER=single lackfit scale=none sls=0.15 sle=0.15 aggregate lackfit;
	output out = wald difdev = stdperason;
```

```
run;


data stratified;
set wald;
if race = 1 and ui = 0 then dummy = "race = 1 and ui = 0";
if race = 1 and ui = 1 then dummy = "race = 1 and ui = 1";
if race = 2 and ui = 0 then dummy = "race = 2 and ui = 0";
if race = 2 and ui = 1 then dummy = "race = 2 and ui = 1";
if race = 3 and ui = 0 then dummy = "race = 3 and ui = 0";
if race = 3 and ui = 1 then dummy = "race = 3 and ui = 1";
run;


proc logistic data=stratified;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0") dummy(ref="race = 1 and ui = 0");
	model low(even = "1")=lwt  smk ptl ht  dummy  /  aggregate scale=none lackfit;
run;


data stratified;
set wald;
if stdperason < 4 and stdperason > -4;
if race = 1 and ui = 0 then dummy = "race = 1 and ui = 0";
if race = 1 and ui = 1 then dummy = "race = 1 and ui = 1";
if race = 2 and ui = 0 then dummy = "race = 2 and ui = 0";
if race = 2 and ui = 1 then dummy = "race = 2 and ui = 1";
if race = 3 and ui = 0 then dummy = "race = 3 and ui = 0";
if race = 3 and ui = 1 then dummy = "race = 3 and ui = 1";
run;


proc logistic data=stratified;
	class race(ref="1") smk(ref="0") ptl(ref="0") ht(ref="0") UI(ref="0") ftv(ref="0") dummy(ref="race = 1 and ui = 0");
	model low(even = "1")=lwt  smk ptl ht  dummy  /  aggregate scale=none lackfit;
run;
```

```
libname sasfp
        "/folders/myshortcuts/SASUniversityEdition/study_design/final project";
run;

data fp;
        set sasfp.data1;
        logage = log(age);
        loglwt = log(lwt);
        if ftv = 0 then ftv =0;
        else ftv = 1;
        if ptl = 0 then ptl = 0;
        else ptl = 1;
run;

proc means data=fp;
        class low;
        var age;
run;


Proc univariate data=fp nonprint;
   class low;
   histogram age/ normal kernel;
   qqplot age / normal(mu=est sigma=est color=red l=2)
           square;
run;

Proc univariate data=fp nonprint;
   class low;
   histogram logage/ normal kernel;
   qqplot logage / normal(mu=est sigma=est color=red l=2)
           square;
run;

proc ttest data = fp;
class low;
var age;
run;
```

```
proc ttest data = fp;
class low;
var logage;
run;


proc means data=fp;
        class low;
        var lwt;
run;


Proc univariate data=fp nonprint;
   class low;
   histogram lwt/ normal kernel;
   qqplot lwt / normal(mu=est sigma=est color=red l=2)
          square;
run;

Proc univariate data=fp nonprint;
   class low;
   histogram loglwt/ normal kernel;
   qqplot loglwt / normal(mu=est sigma=est color=red l=2)
          square;
run;

proc ttest data = fp;
class low;
var lwt;
run;

proc ttest data = fp;
class low;
var loglwt;
run;



proc freq data=fp;
```

```
	table race * low / nocol norow chisq;
run;


proc logistic data= fp;
	class race(ref = "1");
	model low = race;
	run;

proc freq data=fp;
	table smk * low / nocol norow chisq;
run;

proc freq data=fp;
	table ptl * low / nocol norow chisq;
run;

proc freq data=fp;
	table ht * low / nocol norow chisq;
run;

proc freq data=fp;
	table UI * low / nocol norow chisq;
run;

proc freq data=fp;
	table ftv * low / nocol norow chisq;
run;

libname sasfp
	"/folders/myshortcuts/SASUniversityEdition/study_design/final project";

data FP;
	set sasfp.data1;
	if ftv > 0 then ftv = 1;
	if ptl > 0 then ptl = 1;
run;
```

```
* logistic 1 -- age;
title 'Age of the mothers';

proc logistic data=fp plots = all;
	model low(event = "1")=age / aggregate=(age) scale=none lackfit influence;
RUN;


* logistic 1 -- lwt;
title 'weight at last menstrul period';

proc logistic data=fp;
	model low(event = "1")=lwt / aggregate=(lwt) scale=none lackfit;
RUN;

* logistic 1 -- race;
title 'Race';

proc logistic data=fp;
	class race(ref = "1");
	model low(event = "1")=race / aggregate=(race) scale=none lackfit;
RUN;

* logistic 1 -- smk;
title 'Indicator for smoking status';

proc logistic data=fp;
	class smk(ref = "0");
	model low(event = "1")=smk/ aggregate=(smk) scale=none lackfit;
RUN;

* logistic 1 -- ptl;
title 'history of premature labor';

proc logistic data=fp;
	class ptl(ref = "0");
	model low(event = "1")=ptl / aggregate=(ptl) scale=none lackfit;
RUN;

* logistic 1 -- ht;
```

```
title 'history of hypertension';

proc logistic data=fp;
	class ht (ref = "0");
	model low(event = "1")=ht / aggregate=(ht) scale=none lackfit;
RUN;

* logistic 1 -- UI;
title 'present of uterine irritability';

proc logistic data=fp;
	class UI(ref = "0");
	model low(event = "1")=UI / aggregate=(UI) scale=none lackfit;
RUN;

* logistic 1 -- ftv;
title 'number of pysician visits during the first trimester';

proc logistic data=fp;
	class ftv(ref = "0");
	model low(event = "1")=ftv / aggregate=(ftv) scale=none lackfit;
RUN;
```