Will Bliss

CHS 703

Final Project

Determining Risk Factors for Low Birth Weight

Introduction / Background

A large concern for newborn children is their weight at birth. An infant is diagnosed with low birth weight if it weights less than 2500 grams. This is a concern as low birth weight babies are at higher risk of infant mortality and birth defects. During pregnancy, a woman's traits (physically and behaviorally) can lead to an infant having normal birth weight. A study was done by University Baystate Medical Center Springfield, Massachusetts to collect data on 189 women who gave birth. Their medical information was recorded in attempt to recognize which common medical features lead to a woman giving birth to an infant with low or normal birth weight.

It is important to do this study as preterm birth and low birth weight is the number two cause of infant mortality in the US in 2018 (CDC Infant Mortality, 2021). The contribution of low birthweight to infant mortality in the United States is relatively greater now than it was in the past (Committee to Study the Prevention of Low Birthweight). While this may be due to more natural reasons other than an increased prevalence in low birth weight, it shows how low birth weight is still needing to be assessed. With this in mind, of course it is of great importance to reduce the frequency of infants born weighing less than 2500 grams.

In general, the primary determining factors of low birth weight infants are the mother's genetics, age, nutrition, prenatal care, and smoking status (Stanford University). Also, in a study conducted by the National Center for Health Statistics, it was found that black infants are twice as likely to weigh 2500 grams or less than those of white race. However, this is an old study, so we are unable to say if race is clinically significant. In our study, we will look at the primary factors that have been clinically proven, as well as some others in an effort discover more factors that lead to low birth weight. All variables observed in the study can be seen below:

Variable	**Description**	**Codes/Values**	**Name**
1	Identification Code	ID Number	ID
2	Low Birth Weight	1 = BWT<=2500g, 0 = BWT>2500	LOW
3	Age of Mother	Years	AGE
4	Weight of Mother at Last Menstrual Period	Pounds	LWT
5	Race	1 = White, 2 = Black, 3 = Other	RACE
6	Smoking Status During Pregnancy	0 = No, 1 = Yes	SMOKE
7	History of Premature Labor	0 = None, 1 = One, 2 = Two, etc.	PTL
8	History of Hypertension	0 = No, 1 = Yes	HT
9	Presence of Uterine irritability	0 = No, 1 = Yes	UI
10	Number of Physician Visits During the First Trimester	0 = None, 1 = One 2 = Two, etc.	FTV
11	Birth Weight	Grams	BWT

Methods

In our study, we will be applying logistic regression in an effort to determine whether the variables are risk factors for low birth weight. We will use all variables to start, excluding variable 1, the ID, and variable 11. Variable 11 is the numerical birth weight, which we do not need as we are using a binary variable (variable 2) to classify the newborn as having low birth weight or not. Variable 2 (LOW) is the response in this logistic regression. In order to deal with skewed outcomes, we also would like to turn the variables LWT and AGE into categorical variables with 3 outcomes, where each 3 outcomes have similar sample sizes. To do this, we separated the variables at the 33^{rd} and 66^{th} percentiles. In total, Table 1a shows the descriptive statistics for the categorical variables:

Table 1a: Summary of the Categorical Variables

Variable	**Value**	**N**	**Percent (%)**
LOW (Response)	0	130	68.78
	1	59	31.22
AGE (categorized)	1 (Age <= 20)	69	36.51
	2 (20 < Age <= 25)	66	34.92
	3 (25 < Age)	54	28.57
LWT (categorized)	1 (LWT <= 115)	67	35.45
	2 (115 < LWT <= 132)	60	31.75
	3 (132 < LWT)	62	32.80
RACE	1	96	50.78
	2	26	13.76
	3	67	35.45
SMOKE	0	115	60.85
	1	74	39.15
PTL	0	159	84.13
	1	30	15.87
HT	0	177	93.65
	1	12	6.35
UI	0	161	85.19
	1	28	14.81
FTV	0	100	52.91
	1	47	24.87
	2	42	22.22

Note the usage of PTL2, not PTL. We choose to use the PTL2 variable, which has been changed to a binary variable, where PTL2 = 0 means no history of premature labor, and PTL2 = 1 otherwise. We also reduced the values of FTV to 0, 1, and 2 to avoid a more drastic skew. The values are the same, except we combined all values after 2, so that it is “2 or more”. For continuous variables AGE and LWT, Table 1b provides a summary of all relevant information:

Table 1b: Summary of the Continuous Variables

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
AGE	23.24	5.30	14	45
LWT	129.81	30.58	80	250

For our calculations, we will need dummy variables for AGE, FTV, LWT, and RACE. This is because their numerical value has no representation of the variable’s meaning. For each dummy variable, we will need a reference value. We choose the reference values as follows (first bullet point is the template):

- VARIABLE – Reference: Reasoning
- AGE – 25 < Age: Age older than 25 is the reference because clinical studies show that older women give birth to infants with low birth weight more frequently
- FTV – 0: No physician visits is the reference since it is clinically proven that this leads to a higher risk of low birth weight, and it is the most prevalent response
- LWT – LWT <= 115: Weight 115 pounds and below is the reference since it is clinically proven than a woman’s weight is correlated with the birth weight, and it is the most prevalent response

- RACE – White: White is the reference because it is the most frequent value

Other variables have references where value = 1.

The goal of this project is to choose the best model that is made up of variables that are significant risk factors contributing to low birth weight. Since the response is binary, we will be applying logistic regression. We will use a stepwise process to determine the best model for the data. To choose significant variables, we will be using the likelihood ratio test (LRT). A LRT on a variable that results in a large test statistic will have a small p-value. A p-value less than 0.05 implies that the variable is significant. We will begin with a full model where every variable is used. Then, we will remove the variable with the least significance (smallest test statistic) and test the new model. We will repeat this step until all variables are significant. However, we will keep the clinically significant variables regardless of their significance in the likelihood ratio tests done on our data since these are already proven to be contributing factors. After reaching the best model, the effect of significant variables can be seen in Table 3. The actual models and the selection process can be seen in the supplementary table below:

Supplementary Table: Likelihood Ratio Tests for Low Birth Weight Factors

Model	Terms	-2Log-Lik	Chi-Square	df	p-value
1	age_cat1 + age_cat2 + lwt_cat2 + lwt_cat3 + race2 + race3 + smoke + ptl2 + ht + ui + ftv_cat1 + ftv_cat2	197.12	-	-	-
	- ftv_cat1 - ftv_cat2	197.34	.22	2	0.6380
	- race2 - race3	202.23	5.11	2	0.0238
	- pt12	205.35	8.23	1	0.0041
	- ht	203.03	5.91	1	0.0151
	- ui	199.11	1.99	1	0.1581
2	age_cat1 + age_cat2 + lwt_cat2 + lwt_cat3 + race2 + race3 + smoke + ptl2 + ht + ui	198.54	-	-	-
	- race2 - race3	204.59	6.05	2	0.0139
	- pt12	205.70	7.16	1	0.0074
	- ht	204.36	5.82	1	0.0159
	- ui	200.79	2.25	1	0.1336
3	age_cat1 + age_cat2 + lwt_cat2 + lwt_cat3 + race2 + race3 + smoke + ptl2 + ht	200.79	-	-	-
	- race2 - race3	200.61	5.82	2	0.0158
	- pt12	209.69	8.90	1	0.0029
	- ht	206.02	5.23	1	0.0222

Results

The clinically significant variables are AGE, LWT, and SMOKE (Stanford University). While studies show that prenatal care impacts the outcome of low birth weight, we will not include FTV as a clinically significant factor since FTV is only concerned with the first trimester. First, let's look the individual effect of each of each variable:

Table 2: Estimated Effect for Each Variable Based on Separate Univariate Logistic Regression Models

Variable	OR	SE	95% CI	p-value
age_cat1	1.58	0.64	(0.71, 3.51)	0.2644
age_cat 2	4.69	0.69	(.76, 3.77)	0.2021
lwt_cat2	0.60	0.22	(0.29, 1.25)	0.1690
lwt_cat3	0.37	0.15	(0.17, 0.81)	0.0125
race2	2.33	1.08	(0.94, 5.77)	0.0683
race3	1.89	0.66	(0.96, 3.74)	0.0674
smoke	2.02	0.65	(1.08, 3.78)	0.0276
ptl2	4.32	1.79	(1.92, 9.73)	0.0004
ht	3.37	2.05	(1.02, 11.09)	0.0461
ui	2.58	1.07	(1.14, 5.83)	0.0231
ftv_cat1	0.54	0.22	(0.25, 1.20)	0.1296
ftv_cat2	0.71	0.28	(0.32, 1.56)	0.3941

An odds ratio greater than one (OR > 1) implies that odds of low birth weight is higher for that value than the odds of low birth weight for the reference value. An OR < 1 implies that odds of low birth weight is lower for that value than the odds of low birth weight for the reference value. Furthermore, an OR = 1 implies that the value has no difference in odds of low birth weight than the reference value. Not every value can be looked at the same, though. Only odds ratios with a p-value < 0.05 should be more seriously studied.

When analyzing the model step by step, we remove the insignificant variables FTV and UI. Variables AGE, LWT, and SMOKE are not considered for removal since they are clinically significant. Table 3 shows the estimated effects for the variables used in our final model.

Table 3: Estimated Effect for *Significant* Variables Based on Separate Univariate Logistic Regression Models

Variable	OR	SE	95% CI	p-value
age_cat1	1.58	0.64	(0.71, 3.51)	0.2644
age_cat 2	4.69	0.69	(.76, 3.77)	0.2021
lwt_cat2	0.60	0.22	(0.29, 1.25)	0.1690
lwt_cat3	0.37	0.15	(0.17, 0.81)	0.0125
race2	2.33	1.08	(0.94, 5.77)	0.0683
race3	1.89	0.66	(0.96, 3.74)	0.0674
smoke	2.02	0.65	(1.08, 3.78)	0.0276
ptl2	4.32	1.79	(1.92, 9.73)	0.0004
ht	3.37	2.05	(1.02, 11.09)	0.0461

Discussions

Let's diagnose the individual effect of each variable used in our final model. In our final model, we have the following variables: age_cat1 (20 and younger), age_cat2 (21 – 25 years), lwt_cat2 (116 – 132 lbs), lwt_cat3 (133 lbs and above), race2 (black), race3 (other), smoke, ptl2, an ht. While we didn't consider removing AGE or LWT variables from our model because of their clinical significance, it should be noted that the age_cat1, age_cat2, and lwt_cat2 variables have a p-value well above 0.05. This implies that our data does not necessarily support the findings found in other research for age. This is a great example of why previous research is very important. Even though lwt_cat1 has a p-value > 0.05, lwt_cat2 has a p-value < 0.05. This shows that heavier women have lower odds of low birth weight than lighter women. The rest of the variables in the model have p-values around 0.05 or lower. Strictly speaking, however, we should only feel confident about p-values < 0.05.

There are three types of race in this study: white, black, and other. With white as the reference, we can see the effect that black and other races have on low birth weight compared to those that are white. Unfortunately, the p-values are greater than 0.05, so we cannot confidently infer on these odds ratios using 95% confidence. However, race is a clinically significant factor that contributes to low birth weight, so let's analyze the results from our data. According to our data, black women have odds of their child having low birth weight 2.33 times higher than the odds of white women. Yes, p-value > 0.05, but it should be noted that the confidence interval shows a skew well above 1. As for women of other race, the odds of giving birth to a child with low birth weight is 1.89 times higher than the odds for white women, with similar ideas for the p-value and confidence interval as black women.

As mentioned in the introduction, women who smoke are clinically proven to have an increased risk of their baby having low birth weight. Our data supports this claim, as the SMOKE variable has a p-value < 0.05, and OR > 1. Our study on data from Baystate Medical Center results in an odds ratio of 2.02, implying that women who smoke have an odds of a low birth weight child 2.02 times higher than the odds of a non-smoking woman having a low birth weight child.

A history of premature labor is another good contributor to our model. The p-value less than 0.05 proves that the variable is significant for our data. Our data shows that women who have a history of premature labor have odds of having a child with low birth weight 4.32 times higher than the odds of low birth weight for women with no history of premature labor. It should be noted that this is the highest odds ratio in our list of variables within our final model. Additionally, this is the most significant variable in our study with a p-value of 0.0004.

Our final contributing variable, HT, has a p-value less than 0.05 as seen in Table 3. Women with a history of hypertension have odds of giving birth to a low birth weight child 3.37 times higher than the odds of women with no history of hypertension. This makes sense as women with high blood pressure are considered to be of a worse health standing than those with a normal blood pressure (Mayo Foundation).

All in all, from our data and our prior research, a logistic regression model with response variable LOW and explanatory variables AGE, LWT, RACE, SMOKE, PTL, and HT is the best way to calculate the probability of a woman giving birth to a child weight less than 2500g. In order to arrive to this model, we decided to ignore variables UI and FTV. According to our data and clinical research, a presence of uterine irritability and the number of physician visits during the first trimester do not increase or decrease the odds of giving birth to a child that weighs less than 2500g.

References

Abramowicz, M., & Kass, E. H. (1966). Pathogenesis and prognosis of prematurity. *The New England journal of medicine*, *275*(16), . https://doi.org/10.1056/NEJM196610202751605

Centers for Disease Control and Prevention. (2021, September 8). *Infant mortality*. Centers for Disease Control and Prevention. Retrieved May 4, 2022, from https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm

Committee to Study the Prevention of Low Birthweight; Division of Health Promotion and Disease Prevention; Institute of Medicine. Preventing Low Birthweight. Washington (DC): National Academies Press (US); 1985 Jan 1. 1, The Significance of Low Birthweight. Available from: https://www.ncbi.nlm.nih.gov/books/NBK214473/

Lumley J. (1980). Better perinatal health. Australia. *Lancet (London, England)*, *1*(8159), 79–81. https://doi.org/10.1016/s0140-6736(80)90503-6

Mayo Foundation for Medical Education and Research. (2022, January 14). *How high blood pressure can affect the body*. Mayo Clinic. Retrieved May 8, 2022, from https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20045868

McCormick, M. C., Shapiro, S., & Starfield, B. H. (1985). The regionalization of perinatal services. Summary of the evaluation of a national demonstration program. *JAMA*, *253*(6), 799–804.

National Center for Health Statistics. Health, United States, 1982. Washington, D.C: U.S. Government Printing Office; Dec, 1982. (DHHS No. (PHS) 83-1232. Public Health Service).

Shapiro S. (1954). Influence of birth weight, sex, and plurality on neonatal loss in United States. *American journal of public health and the nation's health*, *44*(9), 1142–1153. https://doi.org/10.2105/ajph.44.9.1142

Stanford University. (n.d.). *Causes and Implications of Low Birth Weight Infants*. Primary factors of low birth weight infants. Retrieved May 4, 2022, from https://web.stanford.edu/group/virus/herpes/2000/primaryf.htm

Wilcox, A. J., & Russell, I. T. (1983). Birthweight and perinatal mortality: I. On the frequency distribution of birthweight. *International journal of epidemiology*, *12*(3), 314–318. https://doi.org/10.1093/ije/12.3.314

SAS CODE

```
libname final 'Z:\CHS703\Final';

proc contents data=final.lowbwt2;
run;

proc freq data=final.lowbwt2;
tables low race smoke ht ui ftv ptl2;
run;

proc means data = final.lowbwt2;
var age lwt;
run;

proc univariate data = final.lowbwt2;
var lwt;
output out = percentiles_lwt pctlpts = 33, 67 pctlpre = P_;
run;

proc print data = percentiles_lwt;
run;

** cut off is 115 and 132 pounds (from table);
/*
min = 80
max = 250
1st group= 80-115
2nd group = 116-132
3rd group = 133 - 250

*/

proc univariate data = final.lowbwt2;
var age;
output out = percentiles_age pctlpts = 33, 67 pctlpre = P_;
run;

proc print data = percentiles_age;
run;


** cut off is 20 and 25 years (from table);
/*
min = 14
max = 45
1st group= 14-20
2nd group = 21-25
3rd group = 26 - 40

*/

data lowbwt;
set final.lowbwt2;
if ftv=0 then ftv_cat = 0;
else if ftv = 1 then ftv_cat = 1;
else if ftv in (2,3,4,6) then ftv_cat = 2;
```

```
if lwt <= 115 the lwt_cat = 1;
else if 115 < lwt <= 132 then lwt_cat = 2;
else if 132 < lwt then lwt_cat = 3;

if age <= 20 then age_cat = 1;
else if 20 < age <= 25 then age_cat = 2;
else if 25 < age then age_cat = 3;
run;

proc freq data = lowbwt;
table lwt_cat age_cat nftv;
run;


** create dummy variables;

data lowbwt2;
set lowbwt;
ftv_cat0 = (ftv_cat=0);
ftv_cat1 = (ftv_cat=1);
ftv_cat2 = (ftv_cat=2);

age_cat1 = (age_cat=1);
age_cat2 = (age_cat=2);
age_cat3 = (age_cat=3);

lwt_cat1 = (lwt_cat=1);
lwt_cat2 = (lwt_cat=2);
lwt_cat3 = (lwt_cat=3);
run;

***********************************;
** Note: Race dummy variable already in dataset;

proc freq data = lowbwt2;
tables ftv_cat ftv_cat0-ftv_cat2 age_cat age_cat1-age_cat3 lwt_cat lwt_cat1-
lwt_cat3 race race1-race3;
run;


** Likelihood ratio tests (for interest only, not for paper);

proc logistic data=lowbwt2 descending;
      model LOW = age_cat1 /risklimits;
      **model LOW = age_cat2 /risklimits;
      **model LOW = age_cat3 /risklimits;
      **model LOW = lwt_cat1 /risklimits;
      **model LOW = lwt_cat2 /risklimits;
      **model LOW = lwt_cat3 /risklimits;
      **model LOW = race1 /risklimits;
      **model LOW = race2 /risklimits;
      **model LOW = race3 /risklimits;
      **model LOW = smoke /risklimits;
      **model LOW = ptl /risklimits;
      **model LOW = ht /risklimits;
      **model LOW = ui /risklimits;
```

```
        **model LOW = ftv_cat0 /risklimits;
        **model LOW = ftv_cat1 /risklimits;
        **model LOW = ftv_cat2 /risklimits;
        output out = results predicted = p lower = lcl upper = ucl;
run;

/* Table 2 */
** AGE;
proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2/dist = binomial link=logit;
estimate '1 vs 3' age_cat1 1 age_cat2 0/exp;
estimate '2 vs 3' age_cat1 0 age_cat2 1/exp;
run;

** LWT;
proc genmod data = lowbwt2 descending;
model low = lwt_cat2 lwt_cat3/dist = binomial link=logit;
estimate '2 vs 1' lwt_cat2 1 lwt_cat3 0/exp;
estimate '3 vs 1' lwt_cat2 0 lwt_cat3 1/exp;
run;

** RACE;
proc genmod data = lowbwt2 descending;
model low = race2 race3/dist = binomial link=logit;
estimate 'Black vs White' race2 1 race3 0/exp;
estimate 'Other vs White' race2 0 race3 1/exp;
run;

** SMOKE;
proc genmod data = lowbwt2 descending;
model low = smoke /dist = binomial link=logit;
estimate 'Smoker' smoke 1/exp;
run;

** PTL;
proc genmod data = lowbwt2 descending;
model low = ptl2 /dist = binomial link=logit;
estimate 'PTL' ptl2 1/exp;
run;

** HT;
proc genmod data = lowbwt2 descending;
model low = ht /dist = binomial link=logit;
estimate 'HT' ht 1/exp;
run;

** UI;
proc genmod data = lowbwt2 descending;
model low = ui /dist = binomial link=logit;
estimate 'UI' ui 1/exp;
run;

** FTV;
proc genmod data = lowbwt2 descending;
model low = ftv_cat1 ftv_cat2 /dist = binomial link=logit;
estimate '1 vs 0' ftv_cat1 1 ftv_cat2 0/exp;
estimate '2 vs 0' ftv_cat1 0 ftv_cat2 1/exp;
```

```
run;

** Likelihood ratio tests;
** do not try models without age, lwt, or smoke since clinically sig. ;
** Model 1;
proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht ui
ftv_cat1 ftv_cat2 /dist = binomial;

contrast "-ftv" ftv_cat1 1 ftv_cat2 1;
contrast "-race" race2 1 race3 1;
contrast "-ptl2" ptl2 1;
contrast "-ht" ht 1;
contrast "-ui" ui 1;
run;

proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht ui
ftv_cat1 ftv_cat2 / link = logit dist = binomial;
run;

** Model 2: Model 1 w/o FTV;
proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht ui
/dist = binomial;

contrast "-race" race2 1 race3 1;
contrast "-ptl2" ptl2 1;
contrast "-ht" ht 1;
contrast "-ui" ui 1;
run;

proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht ui
/ link = logit dist = binomial;
run;

** Model 3: Model 1 w/o FTV UI;
proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht
/dist = binomial;

contrast "-race" race2 1 race3 1;
contrast "-ptl2" ptl2 1;
contrast "-ht" ht 1;
run;

proc genmod data = lowbwt2 descending;
model low = age_cat1 age_cat2 lwt_cat2 lwt_cat3 race2 race3 smoke ptl2 ht /
link = logit dist = binomial;
run;
```