US Births (2018) Data Analysis

STAT 448 Final Project

Ankit Datta - Group 7

Introduction:

The [dataset](https://www.kaggle.com/des137/us-births-2018) (https://www.kaggle.com/des137/us-births-2018) we investigated was the data containing information about births in the United States in 2018. US Births is a CDC maintained dataset populated with 55 variables and 3,801,534 observations. Columns in the dataset have information about the newborn's parents, place and time of birth, birth circumstances, biographical information, etc. The CDC user guide provides detailed information about what the variables mean and their coded values if they are a factor variable. For example, many variables such as RACE, EDUCATION, etc. have a coded value and a value of 9 represents missing or unknown information.

Based on this dataset, **the goal and question** I wanted to answer was whether we could **predict the odds of a cesarean birth based on any combination of a subset of these predictors**. In order to do so, the dataset must be cleaned and validated to remove any observations with NA or missing values for any predictor.

Looking at the initial dataset, we see variables such as DMAR, IMP_SEX, MAGE_IMPFLG, MAR_IMP, and more containing a very large number of NA's. We can also see that the max for all the variables is either 9, 99, 999, or 9999. In order to build a logistic regression model, we must remove all observations with an NA or a value which represents a missing variable since those observations will not help us build a model.

```
     ATTEND          BFACIL            BMI             CIG_0             DBWT         DLMP_MM          DLMP_YY          DMAR             DOB_MM           DOB_TT
 Min.   :1.000   Min.   :1.000   Min.   :13.00   Min.   : 0.000   Min.   : 227   Min.   : 1.00   Min.   :2016   Min.   :1.0      Min.   : 1.000   Min.   :   0
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:22.30   1st Qu.: 0.000   1st Qu.:2960   1st Qu.: 4.00   1st Qu.:2017   1st Qu.:1.0      1st Qu.: 4.000   1st Qu.: 759
 Median :1.000   Median :1.000   Median :25.80   Median : 0.000   Median :3300   Median : 7.00   Median :2017   Median :1.0      Median : 7.000   Median :1236
 Mean   :1.332   Mean   :1.036   Mean   :28.82   Mean   : 1.574   Mean   :3267   Mean   :10.93   Mean   :2368   Mean   :1.4      Mean   : 6.561   Mean   :1229
 3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:31.10   3rd Qu.: 0.000   3rd Qu.:3629   3rd Qu.:10.00   3rd Qu.:2018   3rd Qu.:2.0      3rd Qu.: 9.000   3rd Qu.:1733
 Max.   :9.000   Max.   :9.000   Max.   :99.90   Max.   :99.000   Max.   :9999   Max.   :99.00   Max.   :9999   Max.   :2.0      Max.   :12.000   Max.   :9999
                                                                                                                 NA's   :456083
     DOB_WK          DOB_YY          DWgt_R          FAGECOMB         FEDUC           FHISPX          FRACE15          FRACE31          FRACE6           ILLB_R
 Min.   :1.000   Min.   :2018   Min.   :100.0   Min.   :11.0   Min.   :1.000   Min.   :0.000   Min.   : 1.00   Min.   : 1.00   Min.   :1.000   Min.   :  3.0
 1st Qu.:3.000   1st Qu.:2018   1st Qu.:160.0   1st Qu.:28.0   1st Qu.:3.000   1st Qu.:0.000   1st Qu.: 1.00   1st Qu.: 1.00   1st Qu.:1.000   1st Qu.: 32.0
 Median :4.000   Median :2018   Median :182.0   Median :33.0   Median :4.000   Median :0.000   Median : 1.00   Median : 1.00   Median :1.000   Median : 84.0
 Mean   :4.059   Mean   :2018   Mean   :200.3   Mean   :39.6   Mean   :4.906   Mean   :1.647   Mean   :19.45   Mean   :19.22   Mean   :2.864   Mean   :397.1
 3rd Qu.:6.000   3rd Qu.:2018   3rd Qu.:212.0   3rd Qu.:38.0   3rd Qu.:6.000   3rd Qu.:1.000   3rd Qu.: 4.00   3rd Qu.: 4.00   3rd Qu.:4.000   3rd Qu.:888.0
 Max.   :7.000   Max.   :2018   Max.   :999.0   Max.   :99.0   Max.   :9.000   Max.   :9.000   Max.   :99.00   Max.   :99.00   Max.   :9.000   Max.   :999.0

     ILOP_R          ILP_R          IMP_SEX             IP_GON             LD_INDL              MAGER        MAGE_IMPFLG      MAR_IMP          MBSTATE_REC
 Min.   :  3.0   Min.   :  3.0   Mode:logical    Length:3801534     Length:3801534     Min.   :12.00   Mode:logical    Mode:logical     Min.   :1.000
 1st Qu.:888.0   1st Qu.: 28.0   TRUE:90         Class :character   Class :character   1st Qu.:25.00   TRUE:474        TRUE:1995        1st Qu.:1.000
 Median :888.0   Median : 81.0   NA's:3801444    Mode  :character   Mode  :character   Median :29.00   NA's:3801060    NA's:3799539     Median :1.000
 Mean   :751.1   Mean   :417.6                                                         Mean   :29.01                                    Mean   :1.233
 3rd Qu.:888.0   3rd Qu.:888.0                                                         3rd Qu.:33.00                                    3rd Qu.:1.000
 Max.   :999.0   Max.   :999.0                                                         Max.   :50.00                                    Max.   :3.000

     MEDUC           MHISPX            MM_AICU             MRACE15          MRACE31          MRACEIMP           MRAVE6           MTRAN              M_Ht_In
 Min.   :1.000   Min.   :0.0000   Length:3801534     Min.   : 1.000   Min.   : 1.000   Min.   :1          Min.   :1.000   Length:3801534     Min.   :30.00
 1st Qu.:3.000   1st Qu.:0.0000   Class :character   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:1          1st Qu.:1.000   Class :character   1st Qu.:62.00
 Median :4.000   Median :0.0000   Mode  :character   Median : 1.000   Median : 1.000   Median :1          Median :1.000   Mode  :character   Median :64.00
 Mean   :4.413   Mean   :0.6496                      Mean   : 1.942   Mean   : 1.634   Mean   :1          Mean   :1.523                      Mean   :64.31
 3rd Qu.:6.000   3rd Qu.:0.0000                      3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.:1          3rd Qu.:2.000                      3rd Qu.:66.00
 Max.   :9.000   Max.   :9.0000                      Max.   :15.000   Max.   :31.000   Max.   :1          Max.   :6.000                      Max.   :99.00
                                                                                       NA's   :3561344
    NO_INFEC          NO_MMORB          NO_RISKS           PAY             PAY_REC          PRECARE          PREVIS        PRIORDEAD          PRIORLIVE
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :1.000   Min.   : 0.00   Min.   : 0.0   Min.   : 0.0000   Min.   : 0.000
 1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.: 2.00   1st Qu.: 9.0   1st Qu.: 0.0000   1st Qu.: 0.000
 Median :1.0000   Median :1.0000   Median :1.0000   Median :2.000   Median :2.000   Median : 3.00   Median :12.0   Median : 0.0000   Median : 1.000
 Mean   :0.9905   Mean   :0.9936   Mean   :0.6933   Mean   :1.843   Mean   :1.742   Mean   : 5.32   Mean   :13.5   Mean   : 0.2819   Mean   : 1.338
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.: 3.00   3rd Qu.:14.0   3rd Qu.: 0.0000   3rd Qu.: 2.000
 Max.   :9.0000   Max.   :9.0000   Max.   :9.0000   Max.   :9.000   Max.   :9.000   Max.   :99.00   Max.   :99.0   Max.   :99.0000   Max.   :99.000

   PRIORTERM          PWgt_R       RDMETH_REC       RESTATUS        RF_CESAR            RF_CESARN             SEX                WTGAIN
 Min.   : 0.0000   Min.   : 75   Min.   :1.000   Min.   :1.000   Length:3801534     Length:3801534     Length:3801534     Min.   : 0.00
 1st Qu.: 0.0000   1st Qu.:130   1st Qu.:1.000   1st Qu.:1.000   Class :character   Class :character   Class :character   1st Qu.:20.00
 Median : 0.0000   Median :150   Median :1.000   Median :1.000   Mode  :character   Mode  :character   Mode  :character   Median :30.00
 Mean   : 0.7612   Mean   :176   Mean   :1.799   Mean   :1.332                                                            Mean   :31.62
 3rd Qu.: 1.0000   3rd Qu.:182   3rd Qu.:3.000   3rd Qu.:2.000                                                            3rd Qu.:40.00
 Max.   :99.0000   Max.   :999   Max.   :9.000   Max.   :4.000                                                            Max.   :99.00
```

In order to answer this question, we need to determine which subset of predictors are of interest and then determine how cesarean births are represented in our dataset. Due to the fact that there are over 3 million observations, we cannot construct plots due to lack of computational power so we must look at the relevant statistics in order to make conclusions.

Doing some preliminary analysis, we can see that we have Boolean, character, and numeric variables that we all represent coding according to the CDC user guide which will need to be translated to categorical variables to set up for our analysis later.

We would also like to focus our analysis on the most relevant predictors such as MRACE6+FRACE6 (races of parents), FEDUC+MEDUC (education level of parents), DOB_MM, DOB_TT, DOB_WK (month, time, and day of the week of birth), NO_RISKS (risk factors during pregnancy), PRECARE/PREVIS (prenatal care and visits), PRIORDEAD/PRIORLIVE/PRIORTERM (past history of births and terminated pregnancies of mother), and finally RDMETH_REC (method of delivery).

Methods:

I use R initially to subset, validate, verify, and make initial observations about our dataset.

We must first determine which births in our dataset are cesarean births. In order to do so, we must look at the RDMETH_REC variable.

RDMETH_REC	**Delivery Method Recode**	1	Vaginal (excludes vaginal after previous C-section)
		2	Vaginal after previous c-section
		3	Primary C-section
		4	Repeat C-section
		5	Vaginal (unknown if previous c-section)
		6	C-section (unknown if previous c-section)
		9	Not stated

As is shown, the values 3, 4, and 6 are coded for C-sections. Thus, we first create a new column in our dataset CSECTION which will be a binary variable representing whether a birth was a cesarean birth. If the observation has value 3, 4, or 6 then CSECTION = 1 else CSECTION = 0. By doing this, we are setting up a Bernoulli distribution representing having a C-section or not that can be used in our logistic regression to predict the odds. We can also not include the RDMETH_REC variable in our subset as we have already extracted the relevant information.

```
> summary(us_births$RDMETH_REC)
      1       2       3       4       5       6       9
2507707   79017  695090  515780    1810     502    1628
> sum(us_births$csection == TRUE)
[1] 1211372
>
```

Clearly, 695090+515780+502 = 1,211,372 so CSECTION is valid. Initially, this shows us that around 1,211,372/3,801,534 = 31.9% of births in 2018 were C-section births.

Next, we choose our initial subset of relevant variables that can be used in our logistic regression model. By looking at the variables and what they represent, we decide to keep the following 25 variables and CSECTION:

- Child Info: DBWT (birth weight), MONTH (month born), TIME (time of day), DAY (day of week), and SEX
- Parental Info: MAR (marital status), FAGE (father age), FEDUC (father education), FRACE (father race), MAGE, MEDUC, MRACE, ICU (admitted to ICU), M_HT_IN (mother height), PREWEIGHT (pre-pregnancy weight), WTGAIN (weight gain), BMI, CIG_0 (daily cigarettes before birth)
- Birth Info: NO_RISKS (no risk factors), PRECARE (time when prenatal care started), PREVIS (number of pretanal visits), PRIORDEAD (prior births deceased), PRIORLIVE (prior births living), PRIORTERM (prior pregnancy terminations), CESAR_PREV (number of previous C-sections)

We only include the 26/55 variables since other variables in our original dataset contained irrelevant information or redundant information such as BFACIL, 3 coded variables represented FRACE and MRACE, IP_GON – if gonorrhea was present during pregnancy, etc. The variables that were the remaining columns that could be used to predict the odds of having a cesarean birth. We parse the data frame and convert all necessary columns as factors and convert binary character variables to integers so that SAS can interpret them as class variables and have levels to the factor variable.

Finally, we must validate our observations and remove any observations with any NA's or any values that represent unknown/missing (9, 99, 999 etc. depending on the variable according the key). Doing so, we get our final birth dataset with 2,521,050 observations, 25 predictor variables, and our binary response variable CSECTION which we will use moving forward.

Data Set Name	WORK.BIRTHS	Observations	2521050
Member Type	DATA	Variables	26

In our final births dataset, there are 808,802 cesarean births which means about 32.1% of that dataset had C-section births.

Now that we have our dataset and CSECTION variable, we can proceed with our logistic regression in SAS. We know that our goal is to predict the odds of having a cesarean birth, and also that CSECTION is a binary variable with value 0 if not a C-section and 1 if it was a C-section. Then we can use logistic regression to set up a model of the form CSECTION = [subset of predictors].

In order to choose our predictors, we must use forward, backward, and stepwise selection to see which method leads to the best-fitting and most significant model. For all of these PROC LOGISTIC steps, we specify our class variables: CSECTION, MAR, MONTH, DAY, FEDUC, FRACE, MEDUC, ICU, MRACE, NO_RISKS, SEX with param=ref ref=first;

We will get summaries of our variable selection, as well as relevant null hypothesis testing, odds ratios, and goodness of fit tests for each of the methods that we can use to determine which best fits our data. Then we can use that model to make conclusions about our original questions of predicting the odds of having a C-section. Since there are over 2.5 million observations, the model is extremely over-determined and plotting is not possible but the full dataset is used instead of randomly sampling chunks.

Results:

We get the results of forward, backward, and stepwise selection shown in Appendix A. This shows the summaries of each of the methods as well as model fit statistics and global hypothesis tests. All methods lead to the same model which is shown below. First, we look at the model statistics:

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	3163997.3	2304916.0
SC	3164010.0	2305667.7
-2 Log L	3163995.3	2304798.0

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	859197.301	58	<.0001
Score	756301.356	58	<.0001
Wald	448446.907	58	<.0001

The model with the variables has a significantly lower AIC and SC, so therefore that model is better. From the global null hypothesis, we see that the p-values are <0.0001 so we reject the null that all betas are 0 and conclude that at least one of the predictors is not equal to 0. Then we look at the Type 3 analysis of variables and see that all the predictors are significant with p-values <0.05 therefore, the betas for all the predictors are not equal to 0.

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
BMI	1	4.3388	0.0373
CIG_0	1	205.4982	<.0001
DBWT	1	19085.3425	<.0001
MAR	1	328.2834	<.0001
MONTH	11	45.8530	<.0001
TIME	1	364.4686	<.0001
DAY	6	9075.7906	<.0001
FAGE	1	133.4603	<.0001
FEDUC	7	259.7187	<.0001
FRACE	5	77.9827	<.0001
MAGE	1	16059.5820	<.0001
MEDUC	7	338.2164	<.0001
ICU	1	1502.6681	<.0001
MRACE	5	314.6822	<.0001
M_HT_IN	1	1746.2307	<.0001
NO_RISKS	1	8344.9791	<.0001
PREVIS	1	115.1943	<.0001
PRIORDEAD	1	10.7352	0.0011
PRIORLIVE	1	80257.7178	<.0001
PREWEIGHT	1	361.1822	<.0001
CESAR_PREV	1	201910.019	<.0001
SEX	1	2676.0086	<.0001
WTGAIN	1	22347.4364	<.0001

We test the Goodness-of-Fit for this model next:

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
2213.8025	8	<.0001

We that the p-value is <0.0001 which means that we reject the null the model does not fit our data, but with over 2 million observations, this was to be expected and with all 3 selection methods selecting the same model, we can say that the final model predicting the odds of having a cesarean birth as a function of the 23 predictors from before. Starting with 55 variables, this model eliminates 22 other variables that are not included in the model. The influence plots could not be generated for 2 million data points and using the PLOTS(MAXPOINTS=NONE) option for PROC LOGISTIC in SAS causes a segmentation violation. The predictors were all significant in the model and although we reject the goodness of fit, we proceed with the model and make conclusions.

We can try to answer the question and our goal to predict the odds that a birth will be a C-section based on some subset of predictors from our original dataset. Using the 23 predictors, we can apply the statistics to the real world that the data represents.

Conclusions:

We can examine the odds ratios in order to make conclusions to real-world applications of the question we are trying to answer: can we predict the odds of having a cesarean birth based on 23 predictors? First we look at the most significant predictors:

CESAR_PREV	13.913	13.754	14.073
ICU 2 vs 1	5.509	5.054	6.006

NO_RISKS 1 vs 0	0.682	0.677	0.688
PRIORLIVE	0.581	0.579	0.583

Highest Predictor - CESAR_PREV: for every additional previous cesarean section a woman has had, the odds of the current pregnancy being a cesarean birth is 13.9 times more likely. This is predictable, as having a previous C-section is the most obvious predictor for if the birth is going to be a C-section. Another predictor with high odds is being admitted to the ICU (2) during delivery leads to the odds of the C-section to be 5.5 times more than not being admitted (1).

Lowest Odds - PRIORLIVE: for every additional live birth a woman has had in the past, the odds of this birth being a C-section are .581 times as high. This is showing that the more children a women has had, the odds of the current birth being a C-section are lowered by 42%. Similarly, odds of having a cesarean birth are 0.682 times lower for women with no risks (1) during pregnancy compared to risk experienced (0), since having no risks would mean a healthy pregnancy and no complications occurred during pregnancy so it is more likely to be a non cesarean birth.

MAGE	1.064	1.063	1.065
MEDUC 2 vs 1	1.075	1.041	1.109
MEDUC 3 vs 1	1.027	0.997	1.059
MEDUC 4 vs 1	1.007	0.977	1.039
MEDUC 5 vs 1	0.981	0.951	1.013
MEDUC 6 vs 1	0.955	0.926	0.985
MEDUC 7 vs 1	0.930	0.900	0.961
MEDUC 8 vs 1	0.899	0.867	0.931
FEDUC 2 vs 1	1.100	1.069	1.132
FEDUC 3 vs 1	1.136	1.105	1.167
FEDUC 4 vs 1	1.107	1.077	1.139
FEDUC 5 vs 1	1.093	1.061	1.125
FEDUC 6 vs 1	1.073	1.042	1.103
FEDUC 7 vs 1	1.049	1.018	1.081
FEDUC 8 vs 1	1.008	0.975	1.042

1	8th grade or less
2	9th through 12th grade with no diploma
3	High school graduate or GED completed
4	Some college credit, but not a degree.
5	Associate degree (AA,AS)
6	Bachelor's degree (BA, AB, BS)
7	Master's degree (MA, MS, MEng, MEd, MSW, MBA)
8	Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)

Education level of Bachelor's Degree and above (6-8) are only significant for mothers with odds of cesarean births compared to 8th grade or less (1) education being between 5-10% lower with lower odds for higher level of education. The odds for higher education levels is also much lower compared to the odds of high-school degree meaning more educated women have less likely odds to have a cesarean birth.

Father's education shows that odds of having a C-section are highest for High School/GED (3) level compared to (1) which cannot be explained but increasing education levels after 3 along with odds close to 1 for (6) – (8) meaning and not significant. This means that there is no real difference in the odds of having a C-section based on the father's education beyond a bachelor's degree.

MRACE 2 vs 1	1.149	1.130	1.168
MRACE 3 vs 1	0.985	0.943	1.028
MRACE 4 vs 1	1.056	1.033	1.079
MRACE 5 vs 1	0.984	0.903	1.072
MRACE 6 vs 1	0.975	0.954	0.997
FRACE 2 vs 1	1.037	1.021	1.053
FRACE 3 vs 1	0.946	0.905	0.988
FRACE 4 vs 1	0.942	0.920	0.964
FRACE 5 vs 1	0.944	0.869	1.025
FRACE 6 vs 1	0.954	0.933	0.976

1	White (only)
2	Black (only)
3	AIAN (only)
4	Asian (only)
5	NHOPI (only)
6	More than one race

Odds of black mothers having a C-section compared to white mothers is 14.9% higher, and black mother's have highest odds overall. Race plays small role in determining odds of having a C-section compared to other factors, most odds ratios are insignificant or affect the odds of having a C-section by 2% compared to each other.

MAR 2 vs 1	1.081	1.072	1.090

Marital Status 2 v 1 - The odds of having a cesarean birth for unmarried women (2) are 8.1% more than married women (1). So being married decreases the odds of having a C-section.

MONTH 2 vs 1	0.999	0.983	1.015
MONTH 3 vs 1	0.982	0.967	0.998
MONTH 4 vs 1	0.972	0.957	0.988
MONTH 5 vs 1	0.970	0.955	0.985
MONTH 6 vs 1	0.968	0.953	0.984
MONTH 7 vs 1	0.981	0.966	0.996
MONTH 8 vs 1	0.971	0.956	0.986
MONTH 9 vs 1	0.964	0.949	0.979
MONTH 10 vs 1	0.978	0.963	0.994
MONTH 11 vs 1	0.968	0.952	0.983
MONTH 12 vs 1	0.976	0.961	0.992

Having a baby in January or February has the highest odds of being a C-section, with every month having between 2-3% lower odds of having a C-section and odds in February not being significant compared to January. This means that babies born in the first two months of the year have the highest odds of being a cesarean birth, and being born in September has the lowest odds.

DAY 2 vs 1	1.574	1.552	1.596
DAY 3 vs 1	1.536	1.515	1.557
DAY 4 vs 1	1.491	1.470	1.511
DAY 5 vs 1	1.499	1.479	1.520
DAY 6 vs 1	1.564	1.542	1.586
DAY 7 vs 1	1.039	1.024	1.055

Odds are lowest for Sunday (1) births, every week day (2-6) has around 1.5 times higher odds of having a cesarean birth and Saturdays (7) have 3.9% more. This shows that the odds of a birth on a weekend being a cesarean birth are the lowest, and births during the week have the highest odds.

So all of these predictors overall affect the odds of having a cesarean birth by a percentage but within the levels of predictors, there may not be a difference in the odds. The predictor that will most influence the odds is number of previous C-sections which multiplies the odds by 13.9 for each additional one, and the predictor that most lowers the odds is number of prior births which multiples the odds by 0.581 for every additional previous birth.

Appendix:

Number of Observations Read	2521050
Number of Observations Used	2521050

Response Profile

Ordered Value	CSECTION	Total Frequency
1	1	808902
2	0	1712148

Probability modeled is CSECTION=1.

Classification table for class variables:

Class Level Information

Class	Value	Design Variables										
MAR	1	0										
	2	1										
MONTH	1	0	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	0	0	0	0	0	0
	4	0	0	1	0	0	0	0	0	0	0	0
	5	0	0	0	1	0	0	0	0	0	0	0
	6	0	0	0	0	1	0	0	0	0	0	0
	7	0	0	0	0	0	1	0	0	0	0	0
	8	0	0	0	0	0	0	1	0	0	0	0
	9	0	0	0	0	0	0	0	1	0	0	0
	10	0	0	0	0	0	0	0	0	1	0	0
	11	0	0	0	0	0	0	0	0	0	1	0
	12	0	0	0	0	0	0	0	0	0	0	1
DAY	1	0	0	0	0	0	0					
	2	1	0	0	0	0	0					

Class Level Information												
Class	**Value**	**Design Variables**										
	3	0	1	0	0	0	0					
	4	0	0	1	0	0	0					
	5	0	0	0	1	0	0					
	6	0	0	0	0	1	0					
	7	0	0	0	0	0	1					
FEDUC	**1**	0	0	0	0	0	0	0				
	2	1	0	0	0	0	0	0				
	3	0	1	0	0	0	0	0				
	4	0	0	1	0	0	0	0				
	5	0	0	0	1	0	0	0				
	6	0	0	0	0	1	0	0				
	7	0	0	0	0	0	1	0				
	8	0	0	0	0	0	0	1				
FRACE	**1**	0	0	0	0	0						
	2	1	0	0	0	0						
	3	0	1	0	0	0						
	4	0	0	1	0	0						
	5	0	0	0	1	0						
	6	0	0	0	0	1						
MEDUC	**1**	0	0	0	0	0	0	0				
	2	1	0	0	0	0	0	0				
	3	0	1	0	0	0	0	0				
	4	0	0	1	0	0	0	0				
	5	0	0	0	1	0	0	0				
	6	0	0	0	0	1	0	0				
	7	0	0	0	0	0	1	0				
	8	0	0	0	0	0	0	1				
ICU	**1**	0										
	2	1										
MRACE	**1**	0	0	0	0	0						
	2	1	0	0	0	0						

Class Level Information												
Class	**Value**	**Design Variables**										
	3	0	1	0	0	0						
	4	0	0	1	0	0						
	5	0	0	0	1	0						
	6	0	0	0	0	1						
NO_RISKS	**0**	0										
	1	1										
SEX	**1**	0										
	2	1										

Backward Selection:

Summary of Backward Elimination					
Step	**Effect Removed**	**DF**	**Number In**	**Wald Chi-Square**	**Pr > ChiSq**
1	**PRIORTERM**	1	24	1.6743	0.1957
2	**PRECARE**	1	23	2.9706	0.0848

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	3163997.3	2304916.0
SC	3164010.0	2305667.7
-2 Log L	3163995.3	2304798.0

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	859197.301	58	<.0001
Score	756301.356	58	<.0001
Wald	448446.907	58	<.0001

Forward Selection:

Summary of Forward Selection					
Step	**Effect Entered**	**DF**	**Number In**	**Score Chi-Square**	**Pr > ChiSq**
1	**CESAR_PREV**	1	1	539134.702	<.0001
2	**PRIORLIVE**	1	2	68210.3286	<.0001
3	**BMI**	1	3	51276.5468	<.0001
4	**MAGE**	1	4	33254.1795	<.0001
5	**DBWT**	1	5	22110.7786	<.0001
6	**WTGAIN**	1	6	23052.9311	<.0001
7	**M_HT_IN**	1	7	9204.2087	<.0001
8	**DAY**	6	8	9069.8402	<.0001
9	**NO_RISKS**	1	9	8699.2572	<.0001
10	**SEX**	1	10	2735.7951	<.0001
11	**MRACE**	5	11	1856.6319	<.0001
12	**ICU**	1	12	1708.4676	<.0001
13	**MEDUC**	7	13	1394.8401	<.0001
14	**MAR**	1	14	488.4440	<.0001
15	**PREWEIGHT**	1	15	381.6632	<.0001
16	**TIME**	1	16	360.5597	<.0001
17	**FEDUC**	7	17	273.8192	<.0001
18	**CIG_0**	1	18	213.4121	<.0001
19	**FAGE**	1	19	133.9560	<.0001
20	**PREVIS**	1	20	113.4147	<.0001
21	**FRACE**	5	21	77.8381	<.0001
22	**MONTH**	11	22	45.8274	<.0001
23	**PRIORDEAD**	1	23	10.7389	0.0010

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	3163997.3	2304916.0
SC	3164010.0	2305667.7
-2 Log L	3163995.3	2304798.0

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	859197.301	58	<.0001
Score	756301.356	58	<.0001
Wald	448446.907	58	<.0001

Stepwise Selection:

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	CESAR_PREV		1	1	539134.702		<.0001
2	PRIORLIVE		1	2	68210.3286		<.0001
3	BMI		1	3	51276.5468		<.0001
4	MAGE		1	4	33254.1795		<.0001
5	DBWT		1	5	22110.7786		<.0001
6	WTGAIN		1	6	23052.9311		<.0001
7	M_HT_IN		1	7	9204.2087		<.0001
8	DAY		6	8	9069.8402		<.0001
9	NO_RISKS		1	9	8699.2572		<.0001
10	SEX		1	10	2735.7951		<.0001
11	MRACE		5	11	1856.6319		<.0001
12	ICU		1	12	1708.4676		<.0001
13	MEDUC		7	13	1394.8401		<.0001
14	MAR		1	14	488.4440		<.0001
15	PREWEIGHT		1	15	381.6632		<.0001
16		BMI	1	14		3.1716	0.0749
17	TIME		1	15	360.5623		<.0001
18	FEDUC		7	16	272.8844		<.0001
19	CIG_0		1	17	213.0555		<.0001
20	FAGE		1	18	134.0967		<.0001
21	PREVIS		1	19	113.4549		<.0001
22	FRACE		5	20	77.8242		<.0001
23	MONTH		11	21	45.7961		<.0001
24	PRIORDEAD		1	22	10.7204		0.0011
25	BMI		1	23	4.3403		0.0372

Analysis of Max Likelihood Estimates:

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	2.8909	0.1601	326.1356	<.0001
BMI		1	0.00580	0.00278	4.3388	0.0373
CIG_0		1	0.00530	0.000370	205.4982	<.0001
DBWT		1	-0.00041	2.987E-6	19085.3425	<.0001
MAR	**2**	1	0.0775	0.00428	328.2834	<.0001
MONTH	**2**	1	-0.00141	0.00828	0.0290	0.8648
MONTH	**3**	1	-0.0180	0.00805	4.9987	0.0254
MONTH	**4**	1	-0.0281	0.00816	11.8497	0.0006
MONTH	**5**	1	-0.0305	0.00801	14.5193	0.0001
MONTH	**6**	1	-0.0323	0.00807	15.9752	<.0001
MONTH	**7**	1	-0.0194	0.00799	5.8919	0.0152
MONTH	**8**	1	-0.0295	0.00790	13.9981	0.0002
MONTH	**9**	1	-0.0367	0.00805	20.7517	<.0001
MONTH	**10**	1	-0.0218	0.00798	7.4581	0.0063
MONTH	**11**	1	-0.0329	0.00812	16.4493	<.0001
MONTH	**12**	1	-0.0243	0.00813	8.9566	0.0028
TIME		1	-0.00005	2.59E-6	364.4686	<.0001
DAY	**2**	1	0.4535	0.00711	4070.7280	<.0001
DAY	**3**	1	0.4292	0.00702	3735.4029	<.0001
DAY	**4**	1	0.3992	0.00704	3218.6637	<.0001
DAY	**5**	1	0.4051	0.00702	3327.7332	<.0001
DAY	**6**	1	0.4472	0.00704	4034.4496	<.0001
DAY	**7**	1	0.0387	0.00778	24.7254	<.0001
FAGE		1	0.00425	0.000368	133.4603	<.0001
FEDUC	**2**	1	0.0958	0.0146	43.1282	<.0001
FEDUC	**3**	1	0.1274	0.0139	83.9268	<.0001
FEDUC	**4**	1	0.1020	0.0143	51.2512	<.0001
FEDUC	**5**	1	0.0885	0.0150	34.7974	<.0001
FEDUC	**6**	1	0.0700	0.0145	23.3229	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
FEDUC	**7**	1	0.0481	0.0153	9.8433	0.0017
FEDUC	**8**	1	0.00779	0.0169	0.2124	0.6449
FRACE	**2**	1	0.0362	0.00810	19.9994	<.0001
FRACE	**3**	1	-0.0560	0.0224	6.2299	0.0126
FRACE	**4**	1	-0.0600	0.0118	25.9522	<.0001
FRACE	**5**	1	-0.0580	0.0422	1.8862	0.1696
FRACE	**6**	1	-0.0471	0.0114	16.9347	<.0001
MAGE		1	0.0622	0.000491	16059.5820	<.0001
MEDUC	**2**	1	0.0719	0.0162	19.6377	<.0001
MEDUC	**3**	1	0.0270	0.0155	3.0328	0.0816
MEDUC	**4**	1	0.00725	0.0157	0.2120	0.6452
MEDUC	**5**	1	-0.0187	0.0163	1.3160	0.2513
MEDUC	**6**	1	-0.0460	0.0160	8.3086	0.0039
MEDUC	**7**	1	-0.0726	0.0165	19.3497	<.0001
MEDUC	**8**	1	-0.1070	0.0184	33.8944	<.0001
ICU	**2**	1	1.7064	0.0440	1502.6681	<.0001
MRACE	**2**	1	0.1391	0.00849	268.5994	<.0001
MRACE	**3**	1	-0.0154	0.0222	0.4816	0.4877
MRACE	**4**	1	0.0543	0.0113	23.2710	<.0001
MRACE	**5**	1	-0.0159	0.0438	0.1321	0.7163
MRACE	**6**	1	-0.0249	0.0113	4.8459	0.0277
M_HT_IN		1	-0.1037	0.00248	1746.2307	<.0001
NO_RISKS	**1**	1	-0.3826	0.00419	8344.9791	<.0001
PREVIS		1	0.00456	0.000425	115.1943	<.0001
PRIORDEAD		1	-0.0340	0.0104	10.7352	0.0011
PRIORLIVE		1	-0.5428	0.00192	80257.7178	<.0001
PREWEIGHT		1	0.00894	0.000471	361.1822	<.0001
CESAR_PREV		1	2.6328	0.00586	201910.019	<.0001
SEX	**2**	1	0.1712	0.00331	2676.0086	<.0001
WTGAIN		1	0.0173	0.000116	22347.4364	<.0001