Team 47 Final Report

[Github Repository - Team 47] /// *[Final Model]* /// *[Dataset]*

Robert Benedict, Andrew Dean, Bryon Jordan, Clay Watson

Factors Contributing to Infant Mortality in the US

I. Overview of Project

Infant mortality, the death of an infant before their first birthday, is a critical health indicator reflecting a nation's social, economic, and health conditions. In the United States, despite advanced healthcare systems, infant mortality rates remain a significant concern. Our project, inspired by our professional exposure to the healthcare industry, aims to dissect the multifaceted issue of infant mortality. With access to extensive public data provided by governmental bodies such as the Centers for Disease Control and Prevention (CDC), our analysis seeks to uncover underlying factors contributing to infant mortality. This endeavor is not only academically stimulating but also holds profound implications for healthcare policy, education, and practice.

The core objective of our investigation is to identify significant predictors of infant mortality in the U.S., thereby enabling more targeted and effective interventions. Our primary research question, "Are there significant contributing factors to infant mortality rates in the United States, and if so, what are they?" is supported by several auxiliary inquiries focusing on demographic influences, maternal and pregnancy risk factors, characteristics of the infant at birth, and the nature of delivery. Our hypothesis is that we will find significant factors contributing to infant mortality rates.

Business Justification

From a business perspective, especially within a large hospital system, the importance of understanding and mitigating infant mortality cannot be overstated. Enhanced knowledge and awareness can lead to improved training for healthcare providers, better resource allocation, and more informed patient care. Moreover, reducing infant mortality rates aligns with the goals of maximizing efficiency, minimizing costs, and upholding the hospital's reputation for providing exemplary neonatal care. In a broader sense, tackling infant mortality effectively can have ripple effects across the healthcare industry, influencing policy, healthcare delivery, and community health initiatives.

Our hypothesis points to certain variables emerging as significant predictors of infant mortality, with maternal risk factors anticipated to play a pivotal role. While we expect to identify variables with varying degrees of impact, our analysis is geared toward uncovering those with the most substantial influence on infant mortality rates.

II. Overview of Data

For our investigation, we tapped into an extensive dataset made publicly available by the CDC, drawing from several comprehensive databases, including birth records, period-linked birth-infant death records, and fetal stillbirth data from the year 2021. This wealth of data offered a multifaceted glimpse into an array of factors ranging from demographic specifics to comprehensive maternal and infant health indicators.

The variables under scrutiny in our model were diverse, encompassing maternal residency, levels of urbanization, various maternal and pregnancy-related characteristics, prenatal care, risk factors during pregnancy, attributes of labor and delivery, as well as numerous infant traits.

Out of the expansive dataset, which includes 3.78 million individual records, we distilled a set of 33 independent variables that were deemed critical for our predictive model. Many of these variables were categorical; we transformed these into dummy variables to better suit the requirements of logistic regression analysis. This conversion necessitated a detailed review of each variable to ensure completeness and to devise a strategy for handling any missing data that might skew our results.

Our initial intention was to analyze five years' worth of data sourced from the CDC. Although we were able to load the entire dataset into R Studio successfully, we encountered a significant hurdle when attempting to run the regression analysis: system memory limitations interfered with the completion of the model. This prompted a strategic shift in our approach, narrowing our focus exclusively to the most recent year of data, 2021, which allowed us to proceed within the confines of our computational resources.

Further Look into the Data Cleaning Process

Our dataset, provided by the CDC, required meticulous preparation to meet the analytical demands of our logistic regression. We undertook the following steps to ensure data quality and model reliability:

- *Column Name Standardization:* We replaced spaces in column headers with underscores to ensure consistency and avoid potential issues in referencing these columns during analysis.
- *Date Variable Creation:* We created a 'month_year' variable by concatenating the 'Year' and 'Month' fields. This new variable serves a dual purpose: enabling trend analysis over time and ensuring ease of access during our model development.

- *Time Variable Normalization:* The 'Time' variable, which we assumed represents the time of delivery, contained several irregular entries, including 9998, 9909, and missing values. We consolidated these into a single placeholder value (9999) to uniformly denote unknown or unrecorded times, thus simplifying subsequent analyses.
- *Facility Code Rectification:* For the 'Facility' variable, certain codes and missing values were reclassified to 9999. This step aids in addressing gaps in the dataset and maintaining the integrity of our analysis.
- *Race & Education Binary Variables*: We converted categorical data for 'Mothers.Race' and 'Mothers.Education' into binary dummy variables. This transformation allows logistic regression to interpret these categorical inputs correctly.
- *Risk Factor Binary Coding:* Several risk factors—'Diabetes', 'Gestational Diabetes', 'Hypertension', and others—were recoded into a binary format. This approach streamlines the representation of the presence or absence of a condition whilst also accommodating cases with unspecified information.
- *Assisted Reproductive Technology & Previous Cesareans:* We followed a similar binary coding strategy for 'Assisted Reproductive Technology' and 'Previous Cesareans'. This process includes handling missing information tactfully to prevent data loss and ensure meaningful statistical analysis.
- *Delivery Characteristics Coding:* Essential delivery-related variables, such as 'Fetal Presentation', 'Delivery Trial', and 'Delivery Method', were also transformed into dummy variables. This step captures the specifics of each category and provides clarity on the diversity of delivery circumstances.
- *Handling Extreme Cases:* For critical clinical variables like 'Ruptured Uterus' and 'ICU Admissions', binary coding was also applied. This step is crucial in highlighting extreme but significant cases that could heavily influence the outcome of interest.

Each of these steps was performed with a clear objective: to clean and prepare the data for a rigorous statistical examination. We excluded redundant variables, mitigated the risk of multicollinearity, and handled missing data with appropriate imputation strategies. As a result, our dataset was transformed into a format amenable to the powerful logistic regression techniques we employed, setting the stage for insightful and robust analysis.

We then set our base case as the following and looked to train our model:

(1)*Mothers.Race.White,* ***(2)****Mothers.Education.Unknown,* ***(3)****Risk.Factor.Diabetes.Unknown,* ***(4)****Risk.Factor.Gestational.Diabetes.Unknown,* ***(5)****Risk.Factor.Gestational.Hypertension.Unknown,* ***(6)****Delivery.trial.Unknown,* ***(7)****Risk.Factor.Hypertension.Eclampsia.Unknown,* ***(8)****Risk.Factor.Fertility.Drugs.Unknown,*

(9)*Asst.Reproductive_Technology.Unknown,* ***(10)****Previous.Ceasarians.Unknown,* ***(11)****Fetal.Presentation.Unknown,* ***(12)****Delivery.Method.NotStated,* ***(13)****Delivery.Route.Unknown,* ***(14)****Ruptured.Uterus.Unknown,* ***(15)****Mother.admitted.to.ICU.Unknown,* ***(16)****PrePregnancy.Hypertension.Unknown,*

III. Overview of Modeling

The team's objective was to construct a predictive model adept at classifying cases of infant mortality. Given the categorical nature of our response variable, it was unanimously decided that logistic regression would be the most appropriate modeling technique. We established a base case scenario and then divided our dataset, allocating 70% to training our model, with the remaining 30% reserved for testing. This splitting strategy enabled us to calibrate the model with the training data and subsequently assess its predictive accuracy using the test data, ensuring a robust evaluation of the model's performance that neither underfits nor overfits the data.

Model Optimization

Optimizing our logistic regression model involved testing various solvers and systematically eliminating insignificant variables to focus on the most impactful predictors. Our goal was to find an optimal balance between model complexity and predictive power, ensuring our model is both accurate and interpretable.

We initially identified 34 columns in the two data files that we downloaded from the CDC that were common between the file of newborns that survived and newborns that did not.

We walked through all the categorical variables and identified any that had unknown or missing values outside of the scope of our guiding documentation. Fortunately, there were only a few of the variables that needed to account for missing data. We created a new category for those variables using the value “Unknown.”

We then built out indicator variables for the category variables for the logistic regression model. After building the new indicator variables, we removed the original category variables. The net gain of variables ballooned our model to 73 total variables.

Having 73 variables combined with the 3,785,309 data points led to some challenges with our model. We attempted to run a stepwise regression and LASSO regression to reduce the variables in the model, but we were not successful with the pure size of the data set and the memory limitation. This step may be something worth revisiting if this project was taken on by a large hospital system with the necessary computing power.

Pivoting from the problem we had with LASSO and Stepwise regression, we ran the initial regression tests just to get a basic understanding of the model. The sheer size of the model (horizontally and vertically) meant a single run took about 30 minutes.

There was an identifiable split of the variables into two groups within the regression summary (Appendix A.) There were 23 variables that had some sort of significance with p < 0.1 and the other 50 variables were significantly beyond the 0.1 threshold (all but one had a p-value greater than 0.2.) Given the tremendous gap between the two groups, we removed the variables that were beyond the threshold from the model.

Re-running the model using the remaining 23 variables, we managed to identify 3 more variables where the corresponding p-value was now beyond the 0.1 threshold. We then removed the three variables from the model, thus giving us the 19 variables used to build the model as shown in the "Model Results" section in Exhibit 1.1.

Below is a table showing each model iteration, variable count, significant variable count, and AIC. While model 1 had the lowest AIC, we ran the risk of overfitting our model and including far too many variables that had no significance. As a result, we built iteration 2 of the model with a reduced set of variables that were significant or close to significance. The AIC increased, but this is to be expected when reducing the variable count by 50 and eliminating much of the overfitting that was occurring. Still, model 2 had variables that were too insignificant to justify including. Model 3 was the final iteration and included a reduced set of variables that were significant in model 2, or close to significant. This resulted in a model that had a high level of significance among all variables and did not punish the AIC in a large way.

Model Version	**Total Variables**	**Significant Variables *(@ 5% sig.)***	**Insignficant *(@ 5% sig.)***	**AIC**
1	73	18	55	3,413
2	23	19	3	271,466
3 (Final)	19	17	2	273,284

Model Results

Below is a snapshot of the variables that we used in our model with their estimates and significance (Exhibit 1.1). We see that most of our variables are significant at a 5% confidence interval. For implementation, we would not recommend the variable 'genderU', as this is not remotely significant and would not be valuable to include in education, training, or staffing model creation.

Exhibit 1.1

```
Call:
glm(formula = Survived ~ ., family = binomial(), data = my_train_data_narrow,
    maxit = 20)

Coefficients:
                                       Estimate Std. Error  z value Pr(>|z|)
(Intercept)                           2.558e+00  7.976e-02   32.069  < 2e-16 ***
Time                                 -1.642e-04  7.395e-06  -22.202  < 2e-16 ***
Mothers_Age                          -1.868e-02  1.051e-03  -17.776  < 2e-16 ***
Interval_Since_Last_Birth            -1.579e-02  1.482e-03  -10.655  < 2e-16 ***
Monthly_PreNatal_Care_egan           -5.815e-01  3.956e-03 -146.974  < 2e-16 ***
Mothers_Height                       -2.211e-02  9.706e-04  -22.780  < 2e-16 ***
attendant                             1.429e-02  5.977e-03    2.391 0.016808 *
Plurality                             2.787e-01  1.983e-02   14.058  < 2e-16 ***
genderM                              -1.690e-01  1.142e-02  -14.797  < 2e-16 ***
genderU                              -3.374e+01  3.182e+01   -1.060 0.288979
Birthweight_in_grmas                  1.606e-03  5.438e-06  295.294  < 2e-16 ***
Mothers.Race.Black                   -1.470e-01  1.354e-02  -10.864  < 2e-16 ***
Mothers.Race.Nhopi                   -1.460e-01  7.601e-02   -1.921 0.054752 .
Mothers.Education.HighSchoolNoDiploma  2.356e-01  2.219e-02   10.614  < 2e-16 ***
Mothers.Education.HighSchoolGradOrGed -4.965e-02  1.448e-02   -3.430 0.000604 ***
Mothers.Education.SomeCollegeNoDegree  1.507e-01  1.687e-02    8.931  < 2e-16 ***
PrePregnancy.Hypertension.Yes        -3.594e-01  2.573e-02  -13.967  < 2e-16 ***
Fetal.Presentation.Cephalic           1.844e+00  1.961e-02   94.022  < 2e-16 ***
Delivery.Route.Spontaneous           -1.821e+00  1.849e-02  -98.484  < 2e-16 ***
Delivery.Route.Forceps               -2.166e+00  6.932e-02  -31.240  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 778966  on 2648456  degrees of freedom
Residual deviance: 273244  on 2648437  degrees of freedom
AIC: 273284
```

We then evaluated any multicollinearity issues within our model. To observe this, we calculated the Variance Inflation Factor for each variable (Exhibit 1.2).

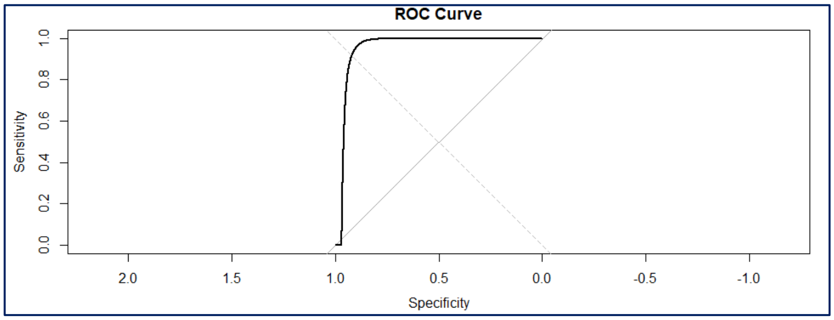
Exhibit 1.2

	GVIF	Df	GVIF^(1/(2*Df))
Time	1.003746	1	1.001871
Mothers_Age	1.275122	1	1.129213
Interval_Since_Last_Birth	1.111865	1	1.054450
Monthly_PreNatal_Care_egan	1.029481	1	1.014634
Mothers_Height	1.027972	1	1.013889
attendant	1.026567	1	1.013196
Plurality	1.114272	1	1.055591
gender	1.000446	2	1.000111
Birthweight_in_grmas	1.091028	1	1.044523
Mothers.Race.Black	1.057452	1	1.028325
Mothers.Race.Nhopi	1.004614	1	1.002304
Mothers.Education.HighSchoolNoDiploma	1.216544	1	1.102971
Mothers.Education.HighSchoolGradOrGed	1.372929	1	1.171721
Mothers.Education.SomeCollegeNoDegree	1.208010	1	1.099095
PrePregnancy.Hypertension.Yes	1.034736	1	1.017220
Fetal.Presentation.Cephalic	1.714111	1	1.309241
Delivery.Route.Spontaneous	1.793053	1	1.339049
Delivery.Route.Forceps	1.058418	1	1.028794

Based on the above, none of the factors are above 4 or below 0. This would likely indicate that our model’s variables exhibit little to no multicollinearity.

After checking the assumptions needed, we then generated an ROC curve to help us determine the optimal cutoff point for classification (Exhibit 1.3).

Exhibit 1.3



Keeping in mind of our goal to create a predictive model to flag deliveries that are at risk for infant mortality, we began iterating through multiple cutoff points and testing the accuracy, specificity, and sensitivity. It is imperative that our model does not overstate false negatives, as that would increase the probability that the staff would not be adequately prepared for deliveries. It is also important that we do not have a high degree of false positives, as that would cause unnecessary overstaffing, overeducation, and raise false alarms.

Ultimately, we found that a cutoff value of 0.8 best met our goals. This cutoff point reduced False Negatives while containing False Positives. We believe this will create the best-case scenario in hospital care. After deciding on the cutoff point, we reviewed the Accuracy, Specificity, and Sensitivity of the outcome to evaluate our model (Exhibit 1.4).

Exhibit 1.4

```
         Actual
Predicted       0       1
        0   31124    6049
        1    7109 1090770
[1] "Accuracy: 0.988407579564637"
Confusion Matrix and Statistics

          Reference
Prediction       0       1
         0   31124    6049
         1    7109 1090770

               Accuracy : 0.9884
                 95% CI : (0.9882, 0.9886)
    No Information Rate : 0.9663
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8195

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.81406
            Specificity : 0.99448
         Pos Pred Value : 0.83727
         Neg Pred Value : 0.99352
             Prevalence : 0.03368
         Detection Rate : 0.02742
   Detection Prevalence : 0.03275
      Balanced Accuracy : 0.90427

       'Positive' Class : 0
```

As we can see from Exhibit 1.4, our model achieved an impressive accuracy rate of 98.84%, indicative of the development of a strong predictive tool. It is also of interest to note that our model has a high Specificity of 99.4%, indicating with high certainty its ability to predict the chances of an infant surviving. Sensitivity is at 81.4%, which would illustrate that there are some False Negatives still present in the logistic regression predictions. Given more resources and time, this may be of particular interest for improvement if a large hospital system desired to further optimize this model. Usage of it in its current state without further tuning to reduce the count of false negatives could potentially result in the hospital understaffing labor or delivery departments, and/or under-educating providers. In return, this would lead to a lower quality of care and potentially a higher infant mortality rate. If the hospital or healthcare system concludes that further research or investigation into predicting infant mortality is warranted, this is certainly an aspect of our model that would take precedence for refinement. Reducing false negatives and increasing our sensitivity is paramount to creating a model that produces the most accurate understanding of what is driving infant mortality.

Model Interpretation

Below is a list of our binary and ordinal significant variables along with their probabilities. (Exhibit 1.5). The probabilities represent, holding all else constant, the probability of survival for that given variable.

Exhibit 1.5

Variable	Coefficient Estimate	Rank	Probability	Variable Class
Fetal.Presentation.Cephalic	1.8440000	1	86.34%	Binary
Plurality	0.2787000	2	56.92%	Binary
Mothers.Education.HighSchoolNoDiploma	0.2356000	3	55.86%	Binary
Mothers.Education.SomeCollegeNoDegree	0.1507000	4	53.76%	Binary
attendant	0.0142900	5	50.36%	Binary
Birthweight_in_grmas	0.0016060	6	50.04%	Ordinal
Time	-0.0001642	7	50.00%	Ordinal
Interval_Since_Last_Birth	-0.0157900	8	49.61%	Ordinal
Mothers_Age	-0.0186800	9	49.53%	Ordinal
Mothers_Height	-0.0221100	10	49.45%	Ordinal
Mothers.Education.HighSchoolGradOrGed	-0.0496500	11	48.76%	Binary
Mothers.Race.Nhopi	-0.1460000	12	46.36%	Binary
Mothers.Race.Black	-0.1470000	13	46.33%	Binary
genderM	-0.1690000	14	45.79%	Binary
PrePregnancy.Hypertension.Yes	-0.3594000	15	41.11%	Binary
Monthly_PreNatal_Care_egan	-0.5815000	16	35.86%	Binary
Delivery.Route.Spontaneous	-1.8210000	17	13.93%	Binary
Delivery.Route.Foreceps	-2.1660000	18	10.28%	Binary
genderU	-33.7400000	19	0.00%	Binary

When building education, training, and staffing models around deliveries, it is important to break down the coefficients and justify how this would impact mortality rates. We can do this in two ways, look at the highest contributors of infant survival rates and the highest contributors of mortality rates. Let's first look at the highest survival probabilities which are:

1) *Fetal Presentation Cephalic* ***2)*** *Plurality* ***3)*** *Mother's Education High School Diploma No* ***4)*** *Mother's Education Some College No Degree*

Interpreting these variables, some of these are a little surprising. For example, looking at the highest contributors for survival rate, two instances of Mother's Education occur. However, we would have anticipated that having a lower education would lead to a lower chance of survival for the baby. One idea here to be explored further is to observe if there is correlation between Mother's Age and Education. If a mother is younger, the baby has a higher chance of survival based on our model. So, if a young mother must leave school early, the baby has a higher chance of survival even with the lower education.

Similarly, it is interesting to note Plurality also has a higher chance of survival. We would have expected it to be a higher risk. One explanation that we could conjecture is that maybe there is more care and focus on a mother birthing more than one child. Therefore, since they are receiving a higher level of care, that could mean there is a higher chance of survival. Again, the hospital system would want to look at this, but one explanation of what might be happening here.

Conversely, we can look at the highest mortality rate contributors. We removed 'genderU' from the below because it was insignificant:

1) *Delivery Route Foreceps* ***2)*** *Delivery Route Spontaneous* ***3)*** *Monthly Prenatal Care Begin* ***4)*** *Pre-Pregnancy Hypertension Yes*

Looking at the four biggest risks for a newborn, we were not as surprised by these outcomes and seem to align with what we expected. We were surprised that some other variables, such as Diabetes, did not come into play, but overall, the top 4 mortality rate contributors make sense.

A hospital system could consider these when caring for the mother and having further education around these areas to be better prepared. For example, delivering by forceps is the most dangerous method of delivery. A hospital system would want to educate their staff in the appropriate methods and the associated risks with it.

IV. Conclusion

Does the Model Answer Our Questions?

We believe we succeeded in proving our hypothesis correct. The insights gathered from our model suggest that certain variables serve as significant indicators, providing a hospital system with the foresight to anticipate and mitigate the risks associated with infant mortality. By utilizing these predictors, healthcare facilities could optimize staffing, enhance training programs, and allocate resources more effectively, ultimately contributing to the preservation of infant lives.

Key predictor variables identified in Exhibit 1.1 stand out as instrumental factors that are significant in understanding what drives increased rates of infant mortality. While not all these variables are influenceable, they are all vital in building proper education and training around infant care, what conditions or attributes to be aware of, and how to tailor care to a specific individual. For instance, we cannot adjust the time interval since the last birth, but by understanding what the interval is, we would be able to inform a healthcare provider that longer intervals are lower risk and extra precautions need to be taken.

Using the methodology and understanding noted above, all coefficients effect can be interpreted and used as variables to explain what is driving infant mortality, where resources should be focused for education and training, and when more prepared staff is needed.

Keeping the significance in mind, it is crucial to acknowledge the limitations of our current model. The process of variable selection, while thorough, did not capture every potential factor. Thus, there may exist additional determinants of infant mortality that were not included in the original dataset. These variables could exert a significant influence on the outcomes in question. We also observed some potential further deep dives that likely need to be done on some of the positive contributors to infant survival rates, two of such being Mother's education and Plurality that would need to be examined further.

Other Key Consideration

There are inherent differences between healthcare systems operating in diverse environments. For instance, a hospital in the bustling urban landscape of New York City would always differ in its dynamics from one in the tranquil rural setting of Stevensville, Montana. Socioeconomic factors, such as the affluence or deprivation of the surrounding area, could also impact the model's relevance. As a result, healthcare providers might find value in tailoring this model to reflect specific regional characteristics, thus enhancing its precision and utility in different contexts.

Final Statement

In summary, while our model serves as a robust starting point, it also opens the door to further exploration and refinement. It provides a foundation upon which healthcare systems can build and adapt to create targeted strategies for improving infant health outcomes across various communities. We are pleased with the model results and think the outcomes are interesting to dive deeper into based on what a hospital system may want to dig into further.

Works Cited

Original Data Repository (CDC): https://wonder.cdc.gov/natality.html

Sub-site (CDC): https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm#Births

Team 47 Repository: Github Repository - Team 47

Appendix A

```
Call:
glm(formula = Survived ~ ., family = binomial(), data = train_data)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-3.724   0.000   0.000   0.000   2.976

Coefficients: (13 not defined because of singularities)
                                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                                -1.720e+02  1.305e+03  -0.132 0.895112
Year                                               NA         NA      NA       NA
Month                                      -2.143e-01  2.133e+00  -0.100 0.919995
Time                                       -2.049e-04  4.015e-05  -5.103 3.34e-07 ***
Weekday                                    -3.398e-02  2.309e-02  -1.471 0.141205
Facility                                   -1.236e+01  5.411e+01  -0.228 0.819294
Mothers_Age                                -4.631e-02  7.955e-03  -5.822 5.82e-09 ***
Interval_Since_Last_Birth                  -4.335e-02  1.177e-02  -3.682 0.000232 ***
Monthly_PreNatal_Care_egan                 -9.204e-02  1.179e-02  -7.803 6.03e-15 ***
Cigarettes_before_pregnancy                -3.449e-03  1.781e-03  -1.937 0.052768 .
Mothers_Height                              4.700e-02  6.951e-03   6.763 1.36e-11 ***
Mothers_BMI                                -1.573e-03  2.716e-03  -0.579 0.562566
Previous_Cesarians_Number                   2.054e-03  1.542e-01   0.013 0.989372
attendant                                   2.974e-01  3.640e-02   8.169 3.11e-16 ***
Plurality                                  -1.538e+00  2.584e-01  -5.953 2.63e-09 ***
genderM                                    -2.390e-01  8.742e-02  -2.734 0.006263 **
genderU                                    -2.375e+01  1.201e+03  -0.020 0.984219
Birthweight_in_grmas                        3.224e-04  2.111e-05  15.273  < 2e-16 ***
month_year                                  9.802e-03  7.009e-02   0.140 0.888786
Mothers.Race.Black                         -4.710e-01  1.095e-01  -4.302 1.70e-05 ***
Mothers.Race.Aian                           4.562e-01  4.808e-01   0.949 0.342684
Mothers.Race.Asian                         -1.710e-01  1.895e-01  -0.902 0.366873
Mothers.Race.Nhopi                         -1.312e+00  5.448e-01  -2.409 0.016004 *
Mothers.Race.multi                         -1.649e-01  2.697e-01  -0.612 0.540848
Mothers.Education.LessThan8th              -5.058e-02  2.779e-01  -0.182 0.855574
Mothers.Education.HighSchoolNoDiploma      -8.982e-01  2.045e-01  -4.392 1.12e-05 ***
Mothers.Education.HighSchoolGradOrGed      -1.034e+00  1.630e-01  -6.343 2.25e-10 ***
Mothers.Education.SomeCollegeNoDegree      -7.082e-01  1.780e-01  -3.979 6.92e-05 ***
Mothers.Education.AssociateDegree          -1.301e-01  2.177e-01  -0.598 0.550020
Mothers.Education.BachelorScience          -1.350e-01  1.867e-01  -0.723 0.469663
Mothers.Education.Masters                   4.263e-01  2.233e-01   1.909 0.056256 .
Mothers.Education.Doctorate                 1.993e-01  3.731e-01   0.534 0.593323
Risk.Factor.Diabetes.Yes                   -2.524e+00  4.061e+05   0.000 0.999995
Risk.Factor.Diabetes.No                    -1.903e+00  4.061e+05   0.000 0.999996
Risk.Factor.Gestational.Diabetes.Yes        3.906e-01  2.187e-01   1.786 0.074128 .
Risk.Factor.Gestational.Diabetes.No                NA         NA      NA       NA
```

```
PrePregnancy.Hypertension.Yes                  -6.485e-01  2.725e-01  -2.380 0.017310 *
PrePregnancy.Hypertension.No                           NA         NA      NA       NA
Risk.Factor.Gestational.Hypertension.Yes        2.782e-01  2.223e-01   1.252 0.210727
Risk.Factor.Gestational.Hypertension.No                NA         NA      NA       NA
Risk.Factor.Hypertension.Eclampsia.Yes                 NA         NA      NA       NA
Risk.Factor.Hypertension.Eclampsia.No                  NA         NA      NA       NA
Risk.Factor.Infertility.Treatment.Unknown              NA         NA      NA       NA
Risk.Factor.Infertility.Treatment.Yes           2.144e+00  2.053e+00   1.044 0.296427
Risk.Factor.Infertility.Treatment.No                   NA         NA      NA       NA
Risk.Factor.Fertility.Drugs.Yes                -4.244e+00  2.217e+00  -1.914 0.055585 .
Risk.Factor.Fertility.Drugs.No                 -9.400e+00  8.082e+00  -1.163 0.244813
Risk.Factor.Fertility.Drugs.NotApplicable              NA         NA      NA       NA
Asst_Reproductive_Technology.Yes                6.970e+00  7.798e+00   0.894 0.371398
Asst.Reproductive_Technology.No                        NA         NA      NA       NA
Asst.Reproductive_Technology.NotApplicable             NA         NA      NA       NA
Previous.Ceasarians.Yes                         5.993e-01  4.061e+05   0.000 0.999999
Previous.Ceasarians.No                          4.807e+00  4.061e+05   0.000 0.999991
Previous.Ceasarians.NotApplicable                      NA         NA      NA       NA
Fetal.Presentation.Cephalic                    -1.017e+00  4.359e-01  -2.334 0.019595 *
Fetal.Presentation.Breech                       1.281e+01  7.587e+02   0.017 0.986525
Fetal.Presentation.Other                        1.836e+01  1.792e+03   0.010 0.991826
Delivery.trial.Yes                             -1.010e+01  3.960e+04   0.000 0.999797
Delivery.trial.No                              -6.540e+00  3.961e+04   0.000 0.999868
Delivery.trial.NotApplicable                    2.285e+10  2.554e+11   0.089 0.928719
Delivery.Method.Vaginal                        -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Method.VaginalAfterPreviousCSection   -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Method.PrimaryCSection                -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Method.RepeatCSection                 -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Method.VaginalUnknownPrevCSection     -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Method.CSectionUnknownPrevCSection    -2.285e+10  2.554e+11  -0.089 0.928719
Delivery.Route.Spontaneous                     -9.186e-01  3.885e-01  -2.365 0.018052 *
Delivery.Route.Foreceps                        -2.251e+00  5.696e-01  -3.951 7.77e-05 ***
Delivery.Route.Vacuum                                  NA         NA      NA       NA
Delivery.Route.Cesarean                         2.285e+10  2.554e+11   0.089 0.928719
Ruotured.Uterus.Yes                             2.393e+01  1.154e+04   0.002 0.998346
Ruotured.Uterus.No                              2.513e+01  2.058e+02   0.122 0.902817
Mother.admitted.to.ICU.Yes                     -5.163e+01  8.813e+03  -0.006 0.995326
Mother.admitted.to.ICU.No                      -5.230e+01  1.063e+03  -0.049 0.960750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 778986.5  on 2648456  degrees of freedom
Residual deviance:   3291.4  on 2648396  degrees of freedom
AIC: 3413.4

Number of Fisher Scoring iterations: 25
```