Examining Birthweight Determinants: How Maternal Age, Education, Baby's Sex, and Gestational Period Shape Neonatal Outcomes*

An Analysis of Birthweight in the United States in 1998

Ruiyang Pang

November 29, 2024

This study examines factors influencing birth weight among infants born in the U.S. in 1998 using Ordinary Least Squares (OLS) and Bayesian regression models. The analysis shows that birth weight is positively associated with gestation length, maternal age, education, prenatal care timing, and infant gender, with gestation length having the most significant effect. However, the models explain less than 30% of the variation in birth weight, highlighting the limitations of the data in capturing unmeasured factors. These findings deepen our understanding of the complex determinants of infant health and offer insights for policies aimed at improving national infant health and fertility rates.

Table of contents

*Code and data are available at: https://github.com/RuiyangPang/birthweight.git.

1 Introduction

1.1 Overview

The challenge of aging populations continues to hinder national economic growth, prompting governments to promote higher birth rates (Christensen et al. 2009). However, modern pressures such as intense job competition, housing shortages, and high living costs have led many young adults to opt for child-free lifestyles or limit their families to a single child. Within this context, the health of newborns becomes a critical focus, with birth weight serving as a key indicator. Low birth weight is associated with an increased risk of diseases like ischemic heart disease and chronic conditions later in life. Studies from the early 1980s highlight the link between low birth weight and fetal malnutrition, which can have lasting developmental effects (Paneth 1995). While economic and technological advancements have reduced malnutrition, identifying contemporary factors influencing birth weight is crucial. This study investigates the determinants of birthweight in today's society, exploring how various socioeconomic and biological factors interplay to shape neonatal outcomes.

1.2 Estimand

My estimand focuses on the relationship between birth weight and various factors. These factors include:

- **Parental Attributes**: The age and education level of the parents and their marital status.
- **Pregnancy and Infant Factors**: Gestation length, prenatal care, the number of prior live births, and the infant's sex.

My goal is to determine which factors significantly affect birth weight, and I also want to understand how these factors contribute to the infant birth weight. This analysis aims to inform strategies for improving outcomes related to birth weight.

This study explored the relationship between various maternal, gestational, and infant factors and newborn birth weight using Ordinary Least Squares (OLS) regression and Bayesian analysis. The OLS model revealed a strong positive correlation between gestational length and birth weight, with each additional week of gestation associated with an average increase of approximately 114.06 grams in birth weight. Maternal age and education emerged as significant predictors, with older mothers tending to have infants with higher birth weights, potentially due to greater life experience and healthier lifestyle choices. Maternal education demonstrated an even stronger effect, indicating that higher educational levels equip mothers with knowledge conducive to practices that benefit fetal health. Marital status was another key factor, with single or divorced mothers having lower infant birth weights, likely reflecting greater socioeconomic and psychological stress. Male infants were generally heavier at birth than female infants, a pattern consistent with biological norms. Early initiation of prenatal care also positively influenced birth weight, underscoring the importance of timely and effective pregnancy management. The Bayesian model produced similar coefficients, reinforcing the robustness of these findings while addressing potential limitations of the OLS assumptions.

This study reveals that infant birth weight is not solely influenced by gestational length but also by other factors, highlighting its complex nature. It is insufficient to predict birth weight based solely on maternal age, education, or similar information. In the study's model, despite incorporating multiple predictors, the R-squared value remains below 30%, indicating that unmeasured variables may significantly contribute to variations in birth weight. The analysis has its limitations due to the data available. Future research could include additional variables, such as maternal health during pregnancy, as predictors to potentially improve the accuracy of birth weight predictions.

This paper is structured as follows: Section 2 provides a detailed overview of the dataset used in the study, including definitions of the outcome and predictor variables, as well as a discussion of key characteristics and limitations. Section 3 introduces the models employed, outlining the Ordinary Least Squares (OLS) and Bayesian models, along with their respective assumptions and methodologies. Section 4 presents the regression results, offering a detailed

explanation of how each predictor influences the outcome variable. Finally, Section 5 explores the broader implications of the findings, addresses study limitations, and proposes directions for future research.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2022), and with the help of several packages in R (R Core Team 2022) to explore these data and build regression models. These packages are include: tidyverse (Wickham et al. 2019),knitr (Xie 2023a),dplyr (Wickham et al. 2023), here (Müller 2020), tinytex (Xie 2023b),readr (Wickham, Hester, and Bryan 2024).

Our data is published in the Data and Story Library (DASL) (DASL 2008), an archive containing hundreds of datasets designed for teaching statistics and data science. I used the baby sample dataset from DASL, a randomly selected subset of 200 records from the 1998 Natality Public Dataset provided by the National Bureau of Economic Research (NBER) (Economic Research 2008). This sample was generated using set.seed(1) in R to ensure reproducibility (R Core Team 2022). Although the NBER dataset provides comprehensive information on all U.S. births in 1998, the full dataset includes over 3.9 million records and occupies nearly 200MB. Such a large dataset poses practical challenges. First, it significantly slows down computational processing. More importantly, in regression analysis, large sample sizes can lead to misleading results. Even minimal effect sizes may appear statistically significant but lack practical relevance, resulting in erroneous conclusions about predictor variables' importance. Following Alexander (2023), we consider random sampling can help us to get a valid dataset to represent the population. The selected dataset contains 7 predictors and 1 outcome variable, with 200 observations, which is sufficient for meaningful regression analysis. we effectively reduced the dataset size while maintaining its representatives of the population. After the data clean, Table 1 shows a preview of the cleaned data set.

Table 1: Preview of the cleaned 1998 birth data set

weight	MomAge	MomEduc	MomMarital	gestation	sex	prenatalstart
3175	35	17	1	39	F	1
3884	22	12	1	42	F	2
3030	35	15	1	39	F	2
3629	23	6	1	40	F	1
3481	23	13	1	42	F	2
3374	26	12	2	39	M	4

2.2 Measurement

The Natality Public Dataset is a comprehensive collection of U.S. birth records processed by the National Center for Health Statistics (NCHS). It includes detailed information on maternal and infant health, family demographics, and other birth-related variables. The dataset records births for both U.S. residents and non-residents but excludes births of U.S. citizens occurring outside the country.

In terms of measurement precision, the dataset replaces exact dates of birth with month and year to protect the privacy of newborns and parents. Geographic details are also simplified, making it less relevant for studies focusing on location-based factors. However, the focus of this dataset is on individual and family characteristics rather than geographic influence.

The dataset is collected officially through birth registration processes required for legal documentation, ensuring high coverage and minimal data missingness. Although the dataset is reliable, some variables, such as paternal age and education, have higher rates of missing data. This discrepancy is due to the lesser involvement of fathers in birth registration and some families lacking complete paternal information.

Despite these limitations, the dataset has low measurement error in variables like infant sex, gestational age, and maternal education. The 1998 dataset includes data for special areas like Northern Mariana Islands and American Samoa, providing an extensive range of records. Given its official source, the dataset serves as a robust foundation for analyzing maternal and infant health factors.

2.3 Outcome variables

Birthweight is reported in some areas in pounds and ounces rather than in grams. However, the metric system has been used in tabulating and presenting the statistics to facilitate comparison with data published by other groups. Equivalents of the gram weights in terms of pounds and ounces are all reported using the metric unit of grams.

A birth weight below 2500 grams is classified as low birth weight, while a weight below 1500 grams is considered very low birth weight (Health Statistics (NCHS) 2008). Newborns with low birth weight often require specialized care, such as admission to a Neonatal Intensive Care Unit (NICU).

Figure 1 shows the distribution of the birth weight in the sample data set. Among the observations, 2 cases are classified as low birth weight, representing approximately 1.09 % of the sample. The mean birth weight is 3318.22 grams, with the lowest recorded birth weight being 1729 grams. The distribution of birth weights appears to be approximately normal. Given that the data was collected through random sampling, a linear regression model can be appropriately applied to analyze the relationships between birth weight and the selected predictors.

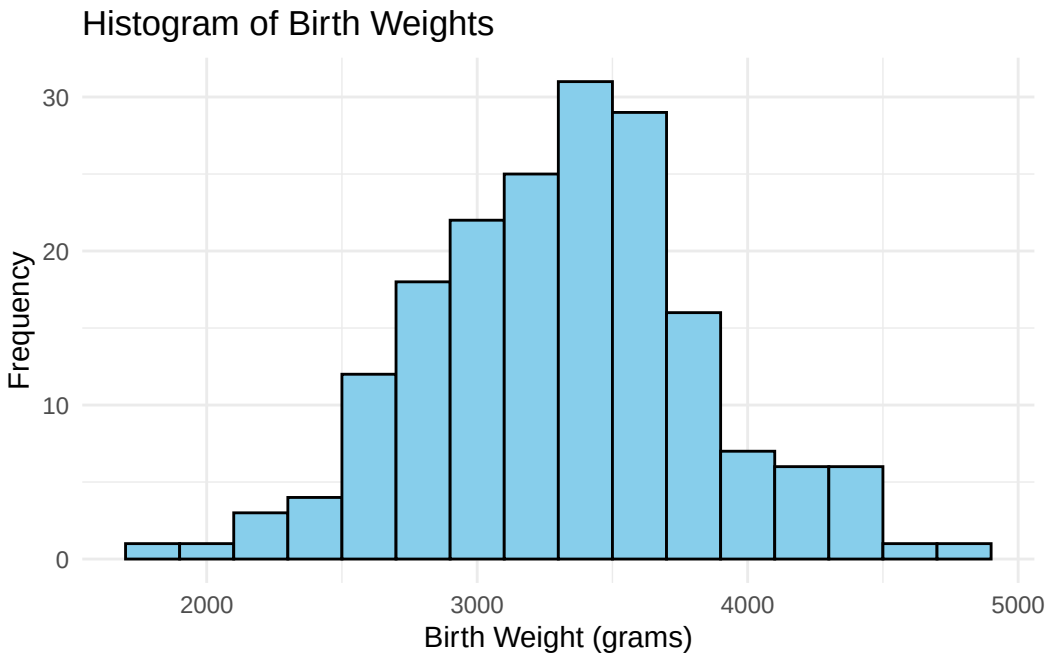

Figure 1: Birth weight of United States newborn in 1998

2.4 Predictor variables

There are 4 numerical and 2 categorical predictors in the sample. The summary of these numerical predictors are shown in Table 2 and their distribution are shown in Figure 2.

In this sample, maternal age ranges from 14 to 42 years, as shown in Figure 2. The distribution of maternal age is concentrated between 20 and 40 years, with the number of births decreasing steadily after age 30. Maternal age is obtained through official documentation, ensuring high accuracy. I predict that as mothers age, they are more likely to have encountered greater childbirth-related experience in life. These experiences help them develop better lifestyle habits and nutrition, which are beneficial for fetal development.

Additionally, maternal education is mostly concentrated at 12 years, corresponding to high school completion. This peak likely reflects that many mothers choose to enter the workforce after high school. Another peak occurs at 16 years, aligning with the typical length of a university education, as completing a four-year degree results in 16 years of education. Highly educated mothers have advantages in both income and knowledge. Therefore, I predict that higher education levels can promote fetal health, leading to an increase in birth weight.

The mode of gestational length is 40 weeks, with a sharp decline in births beyond 40 weeks. However, the probability of birth increases after 36 weeks, as 36 weeks marks the threshold for full-term infants, and delivery can occur anytime after this point. As the fetus develops in the

mother's womb, a longer gestation period allows more time for the fetus to receive nutrients, which in turn contributes to increased birth weight.

Prenatalstart refers to the timing of when prenatal care begins during pregnancy. Earlier initiation of prenatal care is essential for promoting fetal development and supporting maternal health. In this dataset, the mode of prenatalstart is 2, indicating that most mothers began prenatal care during their second month of pregnancy. Furthermore, the majority of these values are less than 5, reflecting that in the United States, parents are generally able to detect pregnancy early and provide timely and necessary prenatal care.

Table 2: Statistics summary of the numerical predictors

MomAge	MomEduc	gestation	prenatalstart
Min. :14.00	Min. : 2.00	Min. :33.00	Min. :0.000
1st Qu.:22.00	1st Qu.:12.00	1st Qu.:38.00	1st Qu.:2.000
Median :26.00	Median :12.00	Median :39.00	Median :2.000
Mean :27.01	Mean :12.79	Mean :39.11	Mean :2.393
3rd Qu.:32.00	3rd Qu.:15.00	3rd Qu.:40.00	3rd Qu.:3.000
Max. :42.00	Max. :17.00	Max. :47.00	Max. :9.000

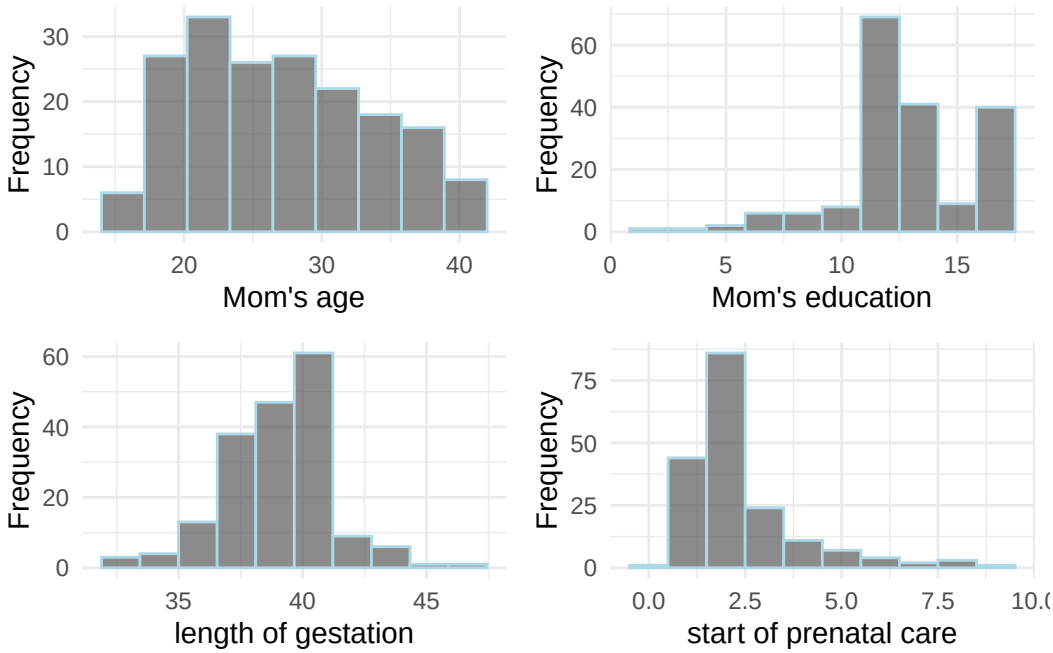

Figure 2: histogram of predictors

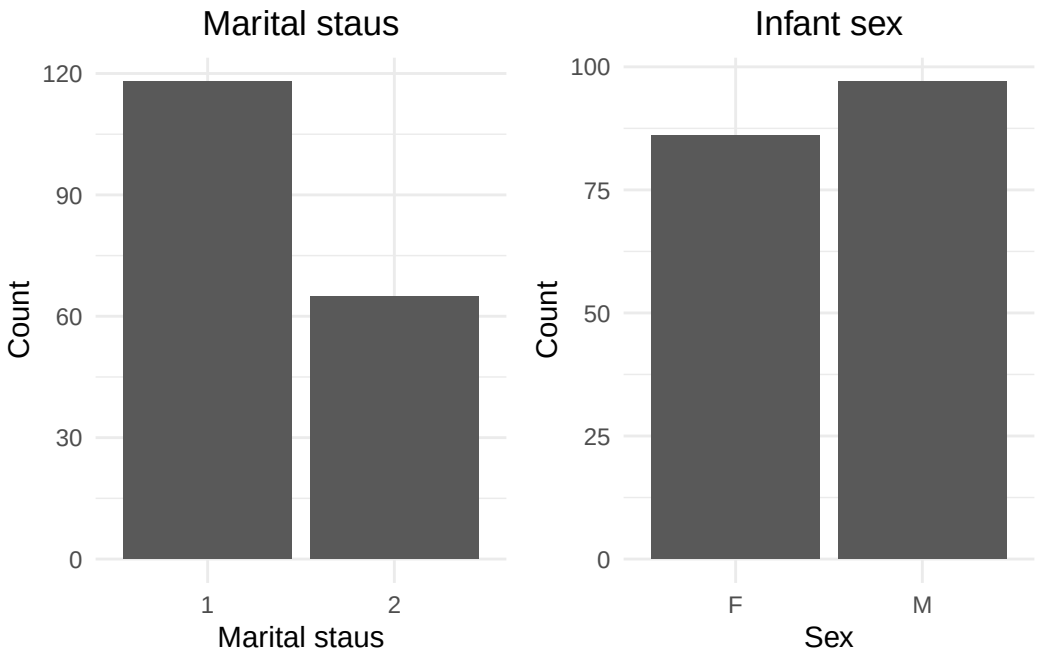

Figure 3: Barplot of predictors

The distribution of the two categorical variables are shown Figure 3. From the barplot of marital status, about of the 64.48% mom is married. The proportion of male infant is a little higher than female infant.

Figure 4 shows that as the length of gestation increases, birth weight also tends to rise. However, the correlation between gestation and birth weight is only 40.01%. This suggests that birth weight is not solely dependent on gestation but is also influenced by other factors. During the prenatal care phase, various parental attributes, pregnancy-related factors, and infant characteristics could all play a role in shaping birth weight. Therefore, in the following sections, I will use regression analysis to explore the impact of these factors on birth weight.

3 Model

The goal of our modeling strategy is twofold. First, we aim to investigate the influence of Parental Attributes (such as parental age and education) and Pregnancy and Infant Factors (like gestational age, timing of prenatal care, and infant sex) on newborn birth weight. This analysis seeks to identify the critical determinants of birth weight and their relative impacts. Second, we aspire to derive insights from these findings that can inform the public about fundamental aspects of childbirth. For instance, we explore the roles of parental characteristics, the gestation period, and prenatal care in fostering healthy birth weights. Moreover, by including

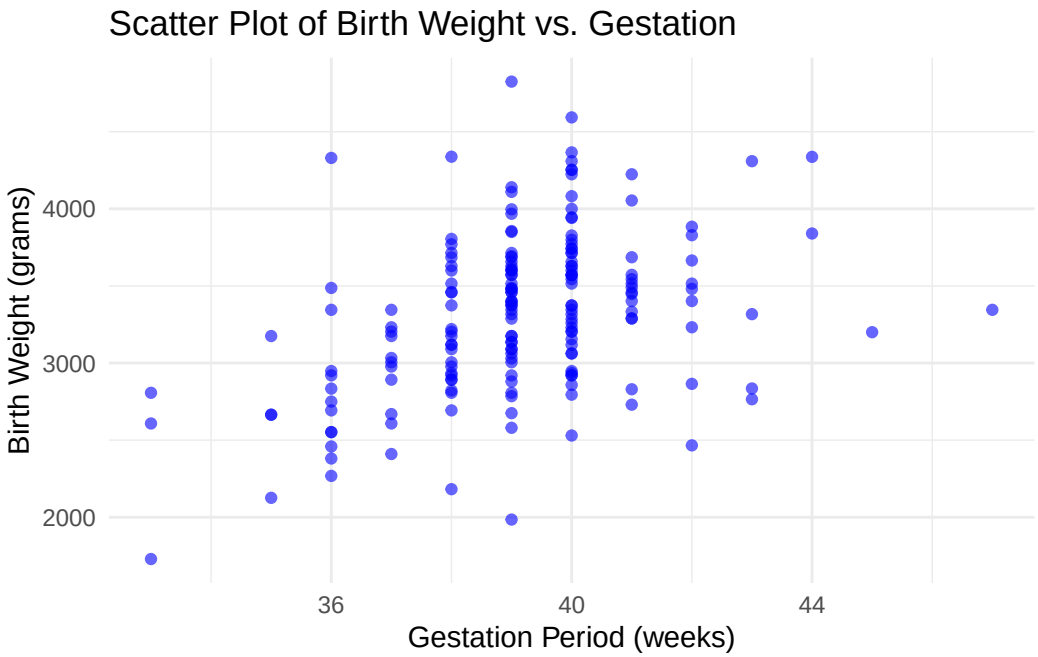

Figure 4: Correlation between birth weight and gestation

infant sex in the model, we aim to mitigate omitted variable bias, ensuring a more accurate understanding of predictor effects. This comprehensive approach not only contributes to the academic study of birth outcomes but also provides practical insights for improving maternal and neonatal health practices.

Linear regression models are often used to estimate relationships between continuous outcome variables and multiple predictors. However, the validity of linear regression can be compromised by various challenges, such as the assumption of linearity in predictors, homoscedasticity of errors, independence of observations, and the influence of outliers (Alexander 2023). To address the limitations of traditional linear regression models, this study employs a Bayesian analysis model.

In the Bayesian framework, model parameters are treated as random variables with prior distributions that reflect initial beliefs or knowledge about these parameters (Goodrich et al. 2022). By combining the information from sample data with prior distributions, we derive the posterior distributions of the parameters. This posterior distribution provides a more robust basis for inference and prediction, particularly in cases where data assumptions are violated or sample sizes are small. Using Bayesian analysis allows for integrating prior knowledge and updating it with observed data, yielding potentially more accurate and nuanced results.

3.1 Model set-up

I will use the OLS model and Bayesian multiple linear regression model to build the relationship between birth weight and other predictors. Define y_i as the birth weight in grams. The OLS model is shown as following:

$$y_i = \beta_0 + \beta_1 MomAge_i + \beta_2 MomEduc_i + \beta_3 gestation_i + \beta_4 prenatalstart_i + \epsilon_i \quad (1)$$
$$\epsilon_i \sim \text{Normal}(0, \sigma^2) \quad (2)$$

Then the Bayesian model is shown in the following:

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (3)$$
$$\mu_i = \beta_0 + \beta_1 MomAge_i + \beta_2 MomEduc_i + \beta_3 gestation_i + \beta_4 prenatalstart_i + \gamma_1 MomMarital_i + \gamma_2 sex_i \quad (4)$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (5)$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (6)$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \quad (7)$$
$$\beta_3 \sim \text{Normal}(0, 2.5) \quad (8)$$
$$\beta_4 \sim \text{Normal}(0, 2.5) \quad (9)$$
$$\gamma_1 \sim \text{Normal}(0, 2.5) \quad (10)$$
$$\gamma_2 \sim \text{Normal}(0, 2.5) \quad (11)$$
$$\sigma \sim \text{Exponential}(1) \quad (12)$$

We run the model in R (R Core Team 2022) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.2 Model justification

We expect a positive relationship between the length of gestation and birth weight, as longer gestation generally results in higher birth weights. Maternal age is another important factor influencing infant health. Older mothers, especially those over the age of 35, often face higher risks during pregnancy, which can impact fetal development (Frederiksen et al. 2018). Maternal education level also plays a crucial role, as more educated mothers are more likely to understand and follow scientifically supported prenatal care practices, benefiting fetal growth. Additionally, marital status can influence pregnancy outcomes. Married women may receive more support from their partners, leading to better nutrition and healthcare, which in turn

Table 3: Linear regressin model on the birth weight

	OLS model 1	OLS model 2
(Intercept)	−643.14	−1647.58
	(675.38)	(758.31)
gestation	101.28	114.06
	(17.24)	(17.47)
MomAge		7.70
		(6.45)
MomEduc		26.55
		(14.22)
MomMarital		−50.12
		(81.63)
as.factor(sex)M		43.90
		(70.77)
prenatalstart		0.56
		(23.78)
Num.Obs.	183	183
R2	0.160	0.214
R2 Adj.	0.155	0.187
AIC	2779.0	2776.9
BIC	2788.7	2802.6
Log.Lik.	−1386.517	−1380.470
RMSE	472.32	456.97

benefits fetal development. Furthermore, the earlier prenatal care begins, the more favorable it is for the fetus's health. Given these considerations, we have selected these factors as predictors for our model. To account for potential biases due to infant gender, we have included it in the model to reduce omitted variable bias.

4 Results

Firstly, the OLS model results are summarized in Table 3. To validate the relationship between the length of gestation and birth weight, a simple linear regression (SLR) using only one variable is presented in the first column of Table 3. It shows that for every additional week of gestation, the birth weight increases by approximately 101.2752518 grams. The corresponding R-squared value for this SLR is 0.1600768, indicating that 16.0076818% of the variation in birth weight can be explained by gestation. This result aligns with expectations, confirming that full-term pregnancies reduce the likelihood of low birth weight.

Table 4: Bayesian regressin model on the birth weight

	Bayesian model
(Intercept)	−1637.58
MomAge	7.58
MomEduc	26.59
MomMarital	−51.97
gestation	114.22
as.factor(sex)M	42.34
prenatalstart	0.40
Num.Obs.	183
R2	0.223
R2 Adj.	0.145
Log.Lik.	−1381.764
ELPD	−1389.6
ELPD s.e.	11.5
LOOIC	2779.3
LOOIC s.e.	23.0
WAIC	2779.1
RMSE	457.07

Subsequently, additional predictors such as maternal age, education, marital status, infant gender, and the timing of prenatal care were added to the model (shown in the second column of Table 3). Maternal age (7.703764) positively impacts birth weight, suggesting that with age, mothers gain experience beneficial for infant development. Maternal education (26.5471448) has an even more pronounced effect on increasing birth weight, as knowledge enables mothers to adopt scientific methods to promote infant health. Marital status (-50.1223407) indicates that divorce negatively affects birth weight, as single mothers, lacking support from a spouse, face greater stress and challenges, leading to lower birth weights.It shows that for every additional week of gestation, the birth weight increases by approximately 114.0636259 grams. Considering the influence of infant gender (43.8992392), male infants tend to weigh more at birth. Lastly, an earlier start to prenatal care (0.560223) is shown to improve birth weight, emphasizing the importance of early pregnancy detection and timely preparation for childbirth.

The validity of OLS results depends on meeting critical assumptions. Therefore, this study adopts a Bayesian analysis model to reanalyze the data. By integrating prior knowledge and accounting for uncertainty, the Bayesian model provides a more robust and reliable explanation of the factors influencing the dataset. The coefficient table are shown in Table 4. Maternal age (7.5848524) and education (26.5857207) positively influence birth weight, with older and more educated mothers better supporting infant development. Marital status (-51.9652454) shows that divorce negatively impacts birth weight, as single mothers face greater stress and less

support. Male infants (114.2187918) typically weigh more, and earlier prenatal (42.3379247) care significantly improves birth weight, highlighting the need for timely pregnancy preparation.Comparing the Bayesian model results (Table 4) with OLS model results (Table 3) reveals that the coefficients obtained are similar to those from the OLS method. This alignment indicates that the model used in this study is robust and reliable.

5 Discussion

5.1 Whats is done

This study uses a simple random sampling (SRS) method to analyze data on newborns born in the United States in 1998. The research reveals a linear relationship between the length of gestation and birth weight. Additionally, maternal age, education level, prenatal preparation, and marital status are identified as significant factors influencing birth weight. Birth weight is closely tied to infant health. Contrary to societal concerns about risks associated with advanced maternal age, the findings show that increasing maternal age positively impacts birth weight. However, the data indicate a decline in childbirth among women over 30. To improve birth rates, it is crucial to address misconceptions about advanced maternal age and its impact on infant health, as these concerns may discourage women from having children after 30.

Regarding maternal education, higher levels of education correlate with increased birth weights, aligning with expectations. However, many women opt out of further education after high school graduation, potentially due to economic pressures. Since professional knowledge significantly influences the quality of prenatal care, measuring education solely in terms of years may introduce bias. Data on specific fields of study, which are often unavailable, could provide better insights. Future research might address this gap through survey-based questionnaires.

Similarly, the timing of prenatal care does not fully capture its quality. Despite including predictors such as maternal age, education, and prenatal care timing, the model explains less than 30% of the variation in birth weight, highlighting the need for additional factors to improve explanatory power.

5.2 What is something that we learn about the world?

The study results show that for every additional week a fetus stays in the mother's womb, the birth weight increases by 114.0636259 grams. This positive correlation suggests that, under normal circumstances, the fetus should remain in the womb as close to 40 weeks as possible for optimal birth weight. However, with current technological advances, induction and cesarean sections have become more common. Induction refers to using medication to speed up labor, while a cesarean section is an intervention to artificially set a delivery date, even in the absence

of signs of labor. Both induction and cesarean sections reduce the time the fetus spends in the womb. Therefore, doctors and expectant mothers should allow for natural delivery unless there are medical complications or emergencies.

However, gestation is only one of the factors influencing birth weight, explaining only 16% of the variation. This indicates the need for further research to explore other factors affecting birth weight.

In the literature, maternal weight gain during pregnancy also has a significant impact on birth weight. The fetus obtains nutrients through the umbilical cord, meaning the mother's nutrition indirectly affects the fetus's birth weight. In addition to diet, maternal mental health can also influence fetal growth. This explains why factors such as maternal age, education, and marital status also play a role in birth weight.

There has been a societal belief that older mothers are detrimental to their babies' health. However, our regression results show that as maternal age increases, birth weight slightly increases as well. This challenges the prevailing concerns about older mothers and suggests that further research is needed to verify the relationship between maternal age and birth weight.

5.3 Weaknesses and next steps

The aim of this study is to explore the factors influencing birth weight, using publicly available data from the National Center for Health Statistics (NCHS). The weakness is that although this dataset covers the entire population, it lacks specific physical metrics, which limits its explanatory power regarding birth weight. One of the benefits of using public data is its easy accessibility and the higher reliability due to its official collection. However, due to privacy concerns, certain detailed information is not available. For example, pregnancy duration is recorded in weeks rather than days, and the education level is only captured in years without considering the specific field of study. Different educational disciplines provide varying degrees of knowledge, and not all educational knowledge directly impacts prenatal care and lifestyle choices during pregnancy.

To improve the prediction of birth weight, it is necessary to collect more detailed personal information. One potential approach could be to collaborate directly with hospitals and extract data from their databases for research. However, hospitals may refuse to cooperate due to privacy issues. An alternative could be to conduct surveys with mothers to gather the necessary data. While this method can face issues such as selection bias, response bias, and response errors, designing a reasonable survey and securing sufficient funding for the project will be key areas for future research to ensure accurate data collection.

This study has practical implications for government policies. For vulnerable groups such as single mothers or those with limited education, the government can provide targeted subsidies

to support maternal and infant health. Additionally, further research into the potential negative effects of advanced maternal age on infant health could help verify the accuracy of such concerns.

Table 5: Description of Predictor Variables

Variable	Description
MomAge	Mother's Single Year of Age
MomEduc	Mother's Education in year
MomMarital	Mother's Marital Status (1 = Yes, 2 = No).
gestation	Gestation – Detail in Weeks,The primary measure used to determine the gestational age of the newborn is the interval between the first day of the mother's last normal menstrual period (LMP) and the date of birth.
sex	Sex of Infant, M male,F Female.
prenatalstart	Month Prenatal Care Began in Month

Appendix

A Surveys, sampling, and observational data

A.1 Population, Frame and sample

Population: Population refers to all births in the United States, including birth statistics for all states and the District of Columbia based on information in the General Record File. This information is received on computer data tapes coded by state and provided to the National Center for Health Statistics (NCHS) through the Vital Statistics Cooperative Program. NCHS receives data for this file from registries in all states, the District of Columbia, and New York City. Birth data for Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands (CNMI) are included in the Public Use File as separate data sets.

Birth data for the United States are limited to births to U.S. residents and nonresidents within the United States. All statistical tables by residence exclude births to nonresidents. This file does not include births to U.S. citizens outside the United States. Birth data for Puerto Rico, the Virgin Islands, Guam, American Samoa, and the CNMI are limited to births within their respective territories.

Frame: The frame is the list of birth records by the NCHS in 1998.

Sample: The sample data is collected by simple random sampling by using set.seed(1) in R (R Core Team 2022), the sample size is 200. The sample data is published on the DASL (DASL 2008). The Description of Predictor Variables are shown in Table 5.

A.2 Simulating Scheme

Given the observed positive correlation between birth weight and gestation, while other variables show less noticeable relationships with birth weight, I designed a simulation to model this relationship. A linear connection between these two variables was specified, with random noise added to simulate birth weight. I set the marginal effect of gestation on birth weight to 82.3, meaning that for each additional week of gestation, the infant's birth weight increases by 82.3 on average.

In the simulation, maternal education was assumed to follow a normal distribution, which might not accurately reflect reality, as people often finish schooling at milestones like high school or college graduation. Prenatal care, which most often begins in the second month, was simulated using a Poisson distribution to reflect its skewed timing. During the testing phase, I ensured all variables were correctly typed and verified that their ranges were within reasonable scales, ensuring the integrity of the simulation.

A.3 Link to Literature

The birth rate of a country significantly impacts its future development. Nations with higher birth rates often have better long-term growth prospects. Historically, early German studies on infant health frequently focused on specific hospitals and their records over defined periods. For example, UCSF's Fetal Hospital conducted research involving 2,946 live births at 37 weeks of gestation to examine how maternal weight gain influenced infant birth weight (Abrams and Laros Jr 1986).

However, relying on regional data can introduce selection bias. For instance, in affluent areas, concerns about fetal malnutrition may be minimal, making such samples unrepresentative of the broader U.S. newborn population. Additionally, hospital-collected data often suffer from recording errors and a higher prevalence of missing values. In contrast, the data used in this study come from official sources, offering greater reliability and accuracy for analysis.

B Model details

B.1 Diagnostic of OLS model

Figure 5 illustrates the residual plot of the OLS model. The residual plot indicates that the assumptions of the OLS model are generally met. Due to this adherence to assumptions, the OLS results show a high degree of similarity to the Bayesian model outcomes, further validating the methods used in this study.

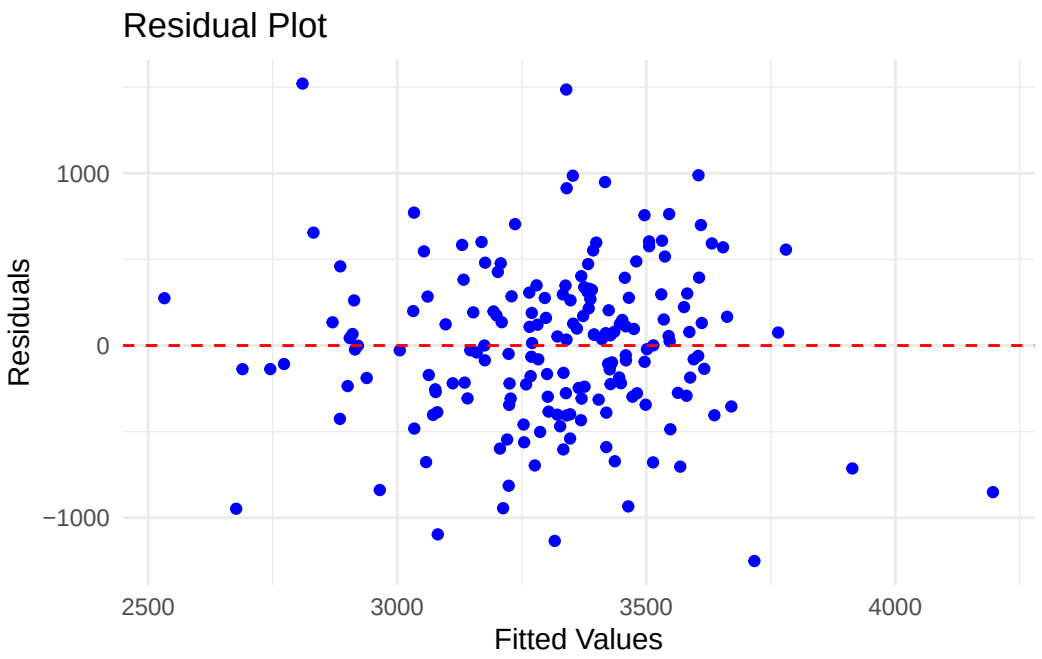

Figure 5: Diagnostic of OLS model

References

Abrams, Barbara F, and Russell K Laros Jr. 1986. "Prepregnancy Weight, Weight Gain, and Birth Weight." *American Journal of Obstetrics and Gynecology* 154 (3): 503–9.

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Christensen, Kaare, Gabriele Doblhammer, Roland Rau, and James W Vaupel. 2009. "Ageing Populations: The Challenges Ahead." *The Lancet* 374 (9696): 1196–1208.

DASL. 2008. "Birthweight Dataset." Online. https://dasl.datadescription.com.

Economic Research, National Bureau of. 2008. "NBER Natality Data Project." https://data.nber.org/natality/ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/.

Frederiksen, Line Elmerdahl, Andreas Ernst, Nis Brix, Lea Lykke Braskhøj Lauridsen, Laura Roos, Cecilia Høst Ramlau-Hansen, and Charlotte Kvist Ekelund. 2018. "Risk of Adverse Pregnancy Outcomes at Advanced Maternal Age." *Obstetrics & Gynecology* 131 (3): 457–63.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Health Statistics (NCHS), National Center for. 2008. "User Guide to the 2008 Natality Public Use File." National Center for Health Statistics, Centers for Disease Control; Prevention. https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Paneth, Nigel S. 1995. "The Problem of Low Birth Weight." *The Future of Children*, 19–34.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023a. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

———. 2023b. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents*. https://github.com/rstudio/tinytex.