Assignment 3

Silpa Soni Nallacheruvu (19980824-5287)

2025-01-14

Overview:

Research question : *Is there a causal effect of smoking during pregnancy on the risk of the child being born with low birth weight (below 2500 grams)?*

Analyse the data using logistic regression, include the right covariates in the model to be able to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight of the child by using the Directed Acyclic Graph (DAG) and interpret the results to answer the research question.

Let's start by loading the dataset and performing some initial data exploration.

But, before that let's clean and prepare the data to make sure that meaningful analysis can be performed on it.

Data Cleanup:

- The variable id was a sequence number randomly assigned to each woman, and is removed as it is not relevant for the analysis.
- There were NA values found in the birthweight_analysis dataset which were removed to ensure that the analysis is performed on complete data.
- There was a negative value found in the variable 'bp_sys' in the birthweight_original dataset which was removed as it was not possible to have a negative systolic blood pressure value.

Code:

```r
# Load the dataset
birthweight_original <- read.csv("birthweight_original.csv")
birthweight_analysis <- read.csv("birthweight_analysisdata.csv")

# Remove the id variable
birthweight_original <- birthweight_original[, -c(5)]
birthweight_analysis <- birthweight_analysis[, -c(5)]
# Remove any NA values
birthweight_original <- na.omit(birthweight_original)
birthweight_analysis <- na.omit(birthweight_analysis)

# Remove negative values in bp_sys
birthweight_original <- birthweight_original[birthweight_original$bp_sys > 0, ]
```

Description Statistics:

Now, let us preform the descriptive statistics of the dataset to understand the distribution of variables and identify any patterns or trends that may be relevant for the analysis.

The birth weight datasets contain information about the birth weight of children, smoking during pregnancy and various other variables such as birth order of the child, age of the mother, systolic blood pressure before pregnancy, birth weight of the child, preterm birth, and child's requirement of neonatal care.

In the birth weight analysis dataset, the variables are categorized as follows:

- The birth weight of the child is categorized as low if it is below 2500 grams.
- The high systolic blood pressure is categorized as high_bp if it is above 135 mmHg.

Code:

```r
# Define the variables
table1_vars <- c("birth", "smoke", "age", "bwt", "bp_sys", "preterm", "neocare")
table2_vars <- c("birth", "smoke", "age", "low", "high_bp", "preterm", "neocare")

# Create table for descriptive statistics: using table one package
table1 <- CreateTableOne(vars = table1_vars, data = birthweight_original,
                         factorVars = c("birth", "smoke", "preterm", "neocare"))

table2 <- CreateTableOne(vars = table2_vars, data = birthweight_analysis,
          factorVars = c("birth", "smoke", "low", "high_bp", "preterm", "neocare"))
```

Output:

Here's the Descriptive Statistics of Birth weight Original Data set:

```r
print(table1, quote = FALSE, noSpaces = TRUE)
```

```
##
##                        Overall
##   n                    486
##   birth (%)
##      1                 187 (38.5)
##      2                 188 (38.7)
##      3                 98 (20.2)
##      4                 13 (2.7)
##   smoke = 1 (%)        194 (39.9)
##   age (mean (SD))      26.40 (5.75)
##   bwt (mean (SD))      2839.49 (688.37)
##   bp_sys (mean (SD))   122.93 (15.07)
##   preterm = 1 (%)      55 (11.3)
##   neocare = 1 (%)      74 (15.2)
```

As we can see that the categorical variables are displayed with their factor percentage, and the continuous variables are displayed with their mean and standard deviation.

Here are some key observations from the descriptive statistics of original birth weight data set:

- **39.9 percent** of the mothers in the dataset smoked during pregnancy with the average age of the mothers being around **26.4 years**.

- **38.5 percent** of the children were first-order births, **38.7 percent** were second-order births, **20.2 percent** were third-order, and **2.7 percent** were fourth-order births.

- The average birth weight of the children was around **2839.49 grams** with a standard deviation of 688.37 grams.

- The average systolic blood pressure before pregnancy was around **122.93 mmHg** with a standard deviation of 15.07 mmHg.

- **11.3 percent** of the children were born preterm and **15.2 percent** of the children required neonatal care.

Here's the Descriptive Statistics of Birth weight Analysis Data set, after categorizing the birth weight as low if it is below 2500 grams and high systolic blood pressure as high_bp if it is above 135 mmHg:

```r
print(table2, quote = FALSE, noSpaces = TRUE)
```

```
##
##                    Overall
##   n                484
##   birth (%)
##      1             186 (38.4)
##      2             188 (38.8)
##      3             97 (20.0)
##      4             13 (2.7)
##   smoke = 1 (%)    193 (39.9)
##   age (mean (SD))  26.38 (5.73)
##   low = 1 (%)      151 (31.2)
##   high_bp = 1 (%)  122 (25.2)
##   preterm = 1 (%)  55 (11.4)
##   neocare = 1 (%)  74 (15.3)
```

Here are some key observations from the descriptive statistics of birth weight analysis data set after removing the presence of NA values in the dataset:

- **39.9 percent** of the mothers in the dataset smoked during pregnancy with the average age of the mothers being around **26.38 years**.
- **38.4 percent** of the children were first-order births, **38.8 percent** were second-order births, **20 percent** were third-order, and **2.7 percent** were fourth-order births.
- **31.2** percent of the children are born with low birth weight.
- **25.2** percent of the mothers had high systolic blood pressure before pregnancy.
- **11.4** percent of the children were born preterm and **15.3** percent of the children required neonatal care.

The descriptive statistics provide a comprehensive overview of the dataset, highlighting the distribution of variables and key characteristics of the study population.

Now, to gain a deeper understanding of the relationships between variables, we will construct a Directed Acyclic Graph (DAG) based on the available literature and domain knowledge, to identify the causal relationships between smoking during pregnancy and low birth weight in children.

DAG: Directed Acyclic Graph

The Directed Acyclic Graph (DAG) is a graphical representation of the causal relationships between variables in the dataset.

Motivation for the DAG:

The DAG is used to establish the total causal effect of smoking during pregnancy on the risk of low birth weight in children with covariates: maternal age, birth order, high systolic blood pressure before pregnancy, preterm birth, and neonatal care.

It ensures that confounders like maternal age and high systolic blood pressure are identified and adjusted for to block non-causal backdoor paths. Simultaneously, it prevents over adjustment by avoiding control of mediators (e.g., preterm birth) or downstream variables (e.g., neonatal care). This approach ensures that the estimated effect reflects the true total causal relationship between smoking and low birth weight.

The DAG is constructed based on the following motivations:

1. Smoke → Low Birth Weight: Maternal smoking in pregnancy was significantly associated with a higher risk of Low Birth Weight in offspring on a global scale. The risk of maternal smoking on infant LBW seems to be increasing over time, and was higher with longer smoking duration throughout pregnancy and more cigarettes smoked daily. Hence, we can conclude that the casual direction of effect is from smoking during pregnancy to low birth weight in children. [4]

2. Age → Low Birth Weight: Both young maternal age (< 20 years) and advanced maternal age (>= 35 years) are associated with a higher risk of LBW due to biological and socio-economic factors showing a U-shaped relationship. Young mothers may experience nutritional deficiencies and inadequate prenatal care, while older mothers face risks such as placental insufficiency and pregnancy complications. Hence, we can conclude that the casual direction of effect is from the maternal age to low birth weight in children. [12]

3. High BP → Low Birth Weight: Pre-pregnancy high systolic blood pressure is inversely associated with offspring birth weight, with higher blood pressure linked to smaller babies for gestational age. This suggests that maternal cardiovascular risk factors, even before conception, can influence fetal growth and may have long-term implications for both maternal and offspring health. Hence, we can conclude that the casual direction of effect is from pre-pregnancy high systolic blood pressure to low birth weight in children. [13]

4. Preterm → Low Birth Weight: Preterm birth (delivery before 37 weeks of gestation) is one of the strongest predictors of LBW because preterm infants have insufficient time for optimal growth in uterus. Most preterm infants weigh less than 2500 grams at birth, primarily due to their shorter gestation period. Hence, we can conclude that the casual direction of effect is from child born preterm to low birth weight in children and not the other way around. [6]

5. Age → Smoke: Smoking during pregnancy is more prevalent in younger mothers (< 20 years) and is associated with higher rates of low birth weight and preterm birth. Younger mothers (< 20 years) are more likely to smoke during pregnancy due to socio-economic factors such as stress, lack of education, and lower healthcare access. Older mothers (>= 35 years) are less likely to smoke during pregnancy, but may have other risk factors for adverse birth outcomes. Hence, we can conclude that the casual direction of effect is from the age of the mother to her smoking during pregnancy.[2]

6. Smoke → Preterm: Smoking during pregnancy is associated with an increased risk of preterm birth, with stronger effects observed for heavy smoking and exposure in late pregnancy. Passive smoking also increases the risk of early preterm birth. Quitting smoking during pregnancy is associated with improved birth outcomes compared to continued smoking. Hence, we can conclude that the casual direction of effect is from the smoking during pregnancy to child born preterm. [7]

7. Age → Preterm: Maternal age shows a U-shaped relationship with preterm birth risk, with the highest risks observed in mothers under 15 years (relative risk 1.569) and above 34 years (relative risk 1.572). The risk increases steeply for women over 40, with most preterm births among older mothers being "late" preterms (34–36 weeks). Women aged 20–34 have the lowest risk, serving as the control group. Hence, we can conclude that the casual direction of effect is from the age of the mother to child born preterm. [11]

8. Birth → Preterm: Birth order (parity) affects the likelihood of preterm birth. First pregnancies (primiparity) are associated with a higher risk of preterm birth due to biological factors and uterine readiness. In contrast, higher parity (multiple previous births) may also increase the risk due to uterine overdistension or scarring. Hence, we can conclude that the casual direction of effect is from the birth order of the child to child born preterm. [1]

9. High BP → Neocare and **High BP → Preterm:** High BP prior to or in early pregnancy was associated with adverse pregnancy outcomes and neonatal outcomes. High blood pressure before pregnancy increases the odds of preterm birth by 1.66 times and neonatal intensive care unit admission by 1.22 times. Hence, we can conclude that the casual direction of effect is from the pre-pregnancy high systolic blood pressure to child born preterm and child requiring neocare.[9]

10. Low → Neocare and **Preterm → Neocare:** Low birth weight is identified as the most critical factor contributing to neonatal mortality, accounting for 10% of deaths globally and necessitating specialized neonatal care for survival. Among LBW neonates, complications such as congenital anomalies (34%) and birth asphyxia (14%) further highlight the direct link between LBW and the need for neonatal care. However, Preterm neonates (gestational age < 37 weeks) had significantly higher mortality rates (32.5%) compared to term small-for-gestational-age neonates (18.9%). This highlights the vulnerability of preterm infants to complications like prematurity-related issues (43%), necessitating intensive neonatal care. Hence, we can conclude that the casual direction of effect is from the child born preterm and low birth weight of the child to child requiring neocare.[3]

11. Age → High BP: Maternal age is positively associated with high systolic blood pressure before pregnancy, with an accelerated rise in systolic blood pressure observed as women age. This age-related increase is further exacerbated by menopause, contributing to heightened cardiovascular risks in middle-aged and elderly women. Hence, we can conclude that the casual direction of effect is from the maternal age to pre-pregnancy high systolic blood pressure. [14]

12. Smoke → Neocare: Maternal smoking during pregnancy is associated with a significant decrease in birthweight, with a stronger effect observed in boys compared to girls. Infants exposed to prenatal smoking are more likely to have low birthweight and require neonatal intensive care, highlighting the need for prenatal counseling to reduce smoking during pregnancy. Hence, we can conclude that the casual direction of effect is from the smoking during pregnancy to child requiring neocare.[15]

13. Age → Birth: Maternal age is inversely associated with birth order, with younger mothers more likely to have higher birth orders due to earlier childbearing. This relationship is influenced by socio-economic factors, cultural norms, and fertility preferences, with younger mothers often having more children compared to older mothers. The birth order will not affect the maternal age, but the maternal age will affect the birth order.

14. Birth → Low: First-order births are associated with a higher incidence of low birth weight compared to higher-order births, regardless of maternal age, with 60.45% of first-order births being low birth weight versus 48.79% for higher orders. This suggests that birth order influences birth weight, with subsequent births generally having a lower risk of low birth weight. [10]

Absence of arrows between the following is explained by the lack of direct causal relationship between them:

1. High BP and Smoke: High SBP before pregnancy can also arise independently of smoking, linked to broader metabolic and cardiovascular conditions.This meta-analysis on high BP and pregnancy outcomes does not identify smoking as a direct factor influencing pre-pregnancy SBP. Smoking is analyzed as a confounder rather than a causal factor. [9]

2. High BP and Birth: It is difficult to establish a direct causal relationship between high blood pressure before pregnancy and birth order, as high BP is a pre-existing condition that may not be influenced by the number of previous births. The relationship between high BP and birth order is likely mediated by other factors such as maternal age, genetic predisposition, and lifestyle factors. Hence, we do not include a direct arrow between high BP and birth order in the DAG.

3. Birth and Smoke: There may be no direct relationship between birth order and maternal smoking during pregnancy because smoking behavior is more likely influenced by individual factors such as socioeconomic status, education level, stress, and cultural norms, rather than the number of children a woman has. Birth order itself does not inherently impact a mother's decision to smoke.

4. Age and Neocare: Maternal age does not contribute significantly to major morbidity of preterm neonates at discharge from neonatal intensive care. [5]

5. Birth and Neocare: Neonatal care and birth order are not directly related because the observed variations in neonatal outcomes with birth order, such as neonatal death rates, are largely influenced by confounding factors like maternal age, parity-related biological changes, and individual susceptibility to adverse outcomes, rather than birth order itself. These effects are intertwined and cannot solely be attributed to the order of the child. [8]

Code:

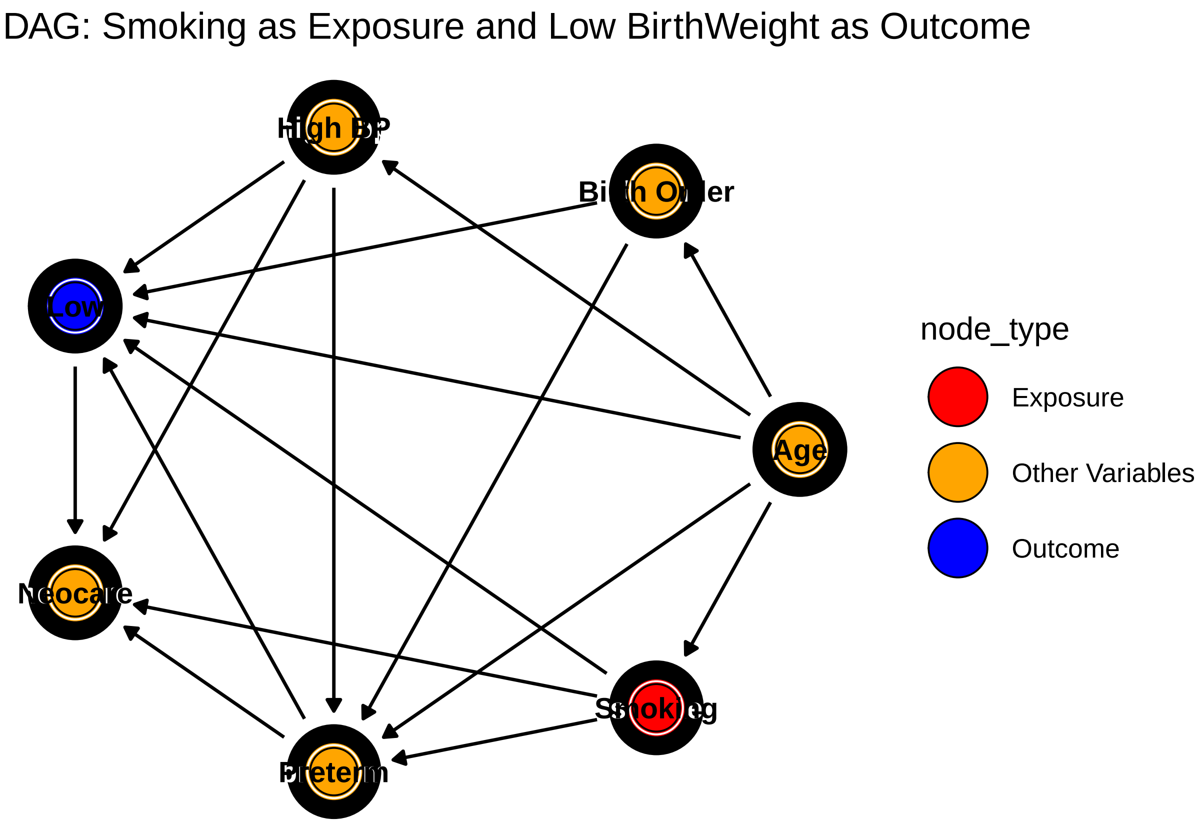
```
# Adjusted DAG String with smoking as exposure and low birth weight as the outcome
dag_string <- "dag {
  smoke [exposure];
  low [outcome];
  smoke -> low;
  age -> low;
  high_bp -> low;
  preterm -> low;
  age -> smoke;
  smoke -> preterm;
  age -> preterm;
  birth -> preterm;
  high_bp -> neocare;
  high_bp -> preterm;
  low -> neocare;
  preterm -> neocare;
  age -> high_bp;
  smoke -> neocare;
  age -> birth;
  birth -> low;
}"

# Generate the DAG
dag <- dagitty(dag_string)

# Extract DAG data for plotting
tidy_dag <- tidy_dagitty(dag, layout = "circle") %>%
  mutate(
    label = case_when(
      name == "birth" ~ "Birth Order",
      name == "smoke" ~ "Smoking",
      name == "age" ~ "Age",
      name == "low" ~ "Low",
      name == "high_bp" ~ "High BP",
      name == "preterm" ~ "Preterm",
      name == "neocare" ~ "Neocare",
      TRUE ~ name
    ),
    # Highlight exposure and outcome nodes
    node_type = case_when(
      name == "smoke" ~ "Exposure",
      name == "low" ~ "Outcome",
      TRUE ~ "Other Variables"
    )
  )

# Plot the DAG with custom labels and node highlighting
ggdag(tidy_dag) +
  geom_dag_node(aes(fill = node_type), shape = 21, size = 10) +  # Highlight nodes
  geom_dag_text(aes(label = label), size = 4, color = "black") +  # Add custom labels
  scale_fill_manual(
    values = c("Exposure" = "red", "Outcome" = "blue", "Other Variables" = "orange")
```

```r
) +  # Colors for exposure, outcome, and other nodes
theme_dag() +
ggtitle("DAG: Smoking as Exposure and Low BirthWeight as Outcome")
```



Based on the reviewed literature, maternal age is commonly categorized as young, middle, and old, rather than treated as a continuous variable. Similarly, birth order is often classified into groups such as first, second and third, and four or more. To align with these practices and simplify the analysis, we will perform data manipulation to categorize these variables accordingly.

Data Manipulation:

Categorization of Variables:

Age: The age variable was categorized into young (< 20 years), middle (20-34 years), and old (>= 35 years) to reflect the established thresholds for advanced maternal age and teenage pregnancy, both of which are associated with significant maternal and neonatal health risks. This categorization is supported by literature indicating distinct patterns of hypertension, preterm birth, and neonatal outcomes across these age groups. Additionally, it enables clearer identification of at-risk populations for targeted interventions.

Birth Order: The birth order variable was categorized into primiparity (first pregnancy), low parity (2-3 pregnancies), and high parity (4 or more pregnancies) to capture the differential risks associated with pregnancy outcomes. Literature indicates that primiparity is linked to higher risks of preterm birth due to uterine immaturity, while high parity is associated with risks like uterine overdistension and scarring, which can impact neonatal outcomes. This categorization ensures alignment with known clinical thresholds for understanding parity-related risks.

Code:

```r
# Categorise age into young, middle, and old
birthweight_analysis$age <- cut(birthweight_analysis$age, breaks = c(0, 20, 34, 100),
       labels = c("young", "middle", "old"), right = FALSE)
# Categorise birth order into primiparity, low parity, and high parity
birthweight_analysis$birth <- cut(birthweight_analysis$birth, breaks = c(0, 2, 4, 100),
       labels = c("primiparity", "low parity", "high parity"), right = FALSE)
```

Covariate Selection:

Covariate selection after constructing the DAG is essential to identify confounders that need adjustment to block backdoor paths, ensuring a valid estimate of the causal effect. It prevents overadjustment by avoiding control for mediators or downstream variables and avoids collider bias by not conditioning on colliders. This step ensures that the model focuses on the total causal effect without introducing unnecessary bias or complexity.

1. Smoking (smoke): Smoking during pregnancy is the primary exposure variable in this study and directly affects low birth weight (low) as per the DAG (smoke → low). Smoking also influences preterm birth (smoke → preterm) and high blood pressure (smoke → high_bp), indirectly contributing to low birth weight.

2. Low Birth Weight (low): Low birth weight is the outcome variable in the DAG and is directly affected by smoking (smoke → low), high blood pressure (high_bp → low), preterm birth (preterm → low), birth order (birth → low) and maternal age (age → low).

3. Age (age): Maternal age acts as a confounder, influencing smoking behavior (age → smoke), preterm birth (age → preterm), high blood pressure (age → high_bp), birth order (age → birth) and low birth weight (age → low). It is crucial to adjust for maternal age to estimate the total causal effect of smoking on low birth weight.

4. High Blood Pressure (high_bp): High systolic blood pressure before pregnancy directly affects the risk of low birth weight (high_bp → low) and preterm birth (high_bp → preterm). It is also influenced by maternal age (age → high_bp) and mediates part of the pathway from age to low birth weight.

5. Preterm Birth (preterm): Preterm birth mediates the relationship between smoking and low birth weight in the DAG (smoke → preterm → low) and is also influenced by maternal age (age → preterm), high blood pressure (high_bp → preterm), and birth order (birth → preterm).

6. Birth Order (birth): Birth order directly affects low birth weight (birth → low) and indirectly affects low birth weight by influencing preterm birth (birth → preterm). It is connected to other variables like maternal age in the DAG.

7. Neonatal Care (neocare): Neonatal care is a downstream outcome of low birth weight (low → neocare) and preterm birth (preterm → neocare). It does not directly influence smoking or low birth weight and should not be adjusted in the analysis.

Here is the list of paths from smoke to low based on the provided DAG and justification for control:

Path	Justification
smoke → low	Direct path; we do not control for it as this is the causal effect of interest.
smoke → preterm → low	preterm is a mediator; we do not control for it to estimate the total causal effect of smoking on low birth weight.
smoke ← age → low	age is a confounder and a fork; we control for it to block the backdoor path.
smoke ← age → high_bp → low	age is a confounder and a fork; we control for it. high_bp is a mediator; we do not control for it.
smoke ← age → birth → low	age is a confounder and a fork; we control for it. birth is a mediator; we do not control for it.

Path	Justification
smoke ← age → high_bp → preterm → low	age and high_bp are confounders; we control for both. preterm is a mediator; we do not control for it.
smoke ← age → birth → preterm → low	age is a confounder and a fork; we control for it. birth and preterm are mediators; we do not control for them.
smoke → preterm → neocare ← low	preterm is a mediator; we do not control for it. neocare is a collider; we do not control for it.
smoke → neocare ← low	neocare is a collider; we do not control for it.
smoke → preterm ← high_bp → low	preterm is a collider; we do not control for it. high_bp is a confounder; we control for it.
smoke → preterm ← birth → low	preterm is a collider; we do not control for it. birth is a confounder; we control for it.

Summary:

- Control for:
  - age (confounder and fork).
  - high_bp (confounder).
  - birth (confounder).
- Do not control for:
  - preterm (mediator).
  - neocare (collider and downstream variable).
  - low (outcome).

This ensures proper estimation of the total causal effect of smoking on low birth weight while avoiding over adjustment or collider bias. Next, to quantify the causal effect of smoking during pregnancy on the risk of low birth weight in children, we will fit logistic regression models with the selected covariates.

Logistic Regression Model:

Let us test the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children, adjusting for maternal age and high systolic blood pressure before pregnancy. This statistical approach enables a robust assessment of the relationship between smoking and low birth weight, accounting for confounding factors and causal pathways.

First, we will look at the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children, adjusting for all the confounders: maternal age, high systolic blood pressure before pregnancy, and birth order.

Logistic regression model formula: $low \sim smoke + age + high\_bp + birth$

```
# Build the logistic regression model
model <- glm(low ~ smoke + age + high_bp + birth, data = birthweight_analysis,
             family = binomial)

summary(model)
```

```
##
## Call:
## glm(formula = low ~ smoke + age + high_bp + birth, family = binomial,
##     data = birthweight_analysis)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.46269    0.32918  -4.443 8.85e-06 ***
## smoke             0.85612    0.20284   4.221 2.44e-05 ***
```

```
## agemiddle         0.33908    0.34383   0.986    0.324
## ageold            0.59178    0.46441   1.274    0.203
## high_bp          -0.26714    0.23907  -1.117    0.264
## birthlow parity   0.05502    0.23083   0.238    0.812
## birthhigh parity  0.04844    0.64948   0.075    0.941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 600.82  on 483  degrees of freedom
## Residual deviance: 579.50  on 477  degrees of freedom
## AIC: 593.5
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate odds ratios and 95% confidence intervals
exp(cbind(Odds_Ratio = coef(model), confint(model)))
```

```
##                  Odds_Ratio     2.5 %    97.5 %
## (Intercept)       0.2316116 0.1177798 0.4310932
## smoke             2.3539996 1.5844402 3.5122002
## agemiddle         1.4036573 0.7282292 2.8240102
## ageold            1.8072097 0.7313738 4.5441697
## high_bp           0.7655689 0.4746703 1.2143404
## birthlow parity   1.0565576 0.6731954 1.6663232
## birthhigh parity  1.0496360 0.2750243 3.6626410
```

Interpretation of Odds Ratios and 95% Confidence Intervals:

- The **odds ratio for smoking during pregnancy is 2.35**, indicating that the odds of low birth weight are 2.35 times higher for children born to mothers who smoke during pregnancy compared to non-smokers, adjusting for maternal age, birth order and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (1.58, 3.51), suggesting a statistically significant association between smoking and low birth weight. This is supported by the p-value of 0.00002 ($<$ 0.05), indicating a significant effect of smoking on low birth weight at a 5% level of significance.

- The **odds ratio for middle-aged mothers (20-34 years) is 1.4**, suggesting that the odds of low birth weight are 1.4 times lower for middle-aged mothers compared to young mothers ($<$ 20 years), adjusting for smoking, birth order and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.72, 2.82) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.324 ($>$ 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.

- The **odds ratio for old-aged mothers ($>=$ 35 years) is 1.8**, suggesting that the odds of low birth weight are 1.8 times higher for old-aged mothers compared to young mothers ($<$ 20 years), adjusting for smoking, birth order and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.73, 4.54) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.203 ($>$ 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.

- The **odds ratio for high systolic blood pressure before pregnancy is 0.765**, suggesting that the odds of low birth weight are 0.765 times lower for mothers with high blood pressure compared to mothers without high blood pressure, adjusting for smoking, birth order and maternal age. The 95% confidence interval for the odds ratio is (0.47, 1.21) includes 1, indicating that the effect of high blood pressure on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.264 (> 0.05) which indicates that the effect of high blood pressure is not statistically significant at 5 % level of significance.

- The **odds ratio for low parity birth is 1.06**, suggesting that the odds of low birth weight are 1.06 times higher for children with birth order of second or third compared to first-order births, adjusting for smoking, maternal age and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.67, 1.66) includes 1, indicating that the effect of birth order on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.812 (> 0.05) which indicates that the effect of birth order is not statistically significant at 5 % level of significance.
- The **odds ratio for high parity birth is 1.05**, suggesting that the odds of low birth weight are 1.05 times higher for children with birth order of four or more compared to first-order births, adjusting for smoking, maternal age and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.27, 3.66) includes 1, indicating that the effect of birth order on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.941 (> 0.05) which indicates that the effect of birth order is not statistically significant at 5 % level of significance.

Although, the odds ratio results for low parity and high parity are in contradiction with the literature, which suggests that first-order births are associated with a higher incidence of low birth weight compared to higher-order births. This discrepancy may be due to the small sample size or other unmeasured confounders that were not accounted for in the model.

Now, let us check the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children, adjusting for maternal age and high systolic blood pressure before pregnancy.

Logistic regression model formula: $low \sim smoke + age + high\_bp$

```r
# Build the logistic regression model
model1 <- glm(low ~ smoke + age + high_bp, data = birthweight_analysis,
             family = binomial)
summary(model1)
```

```
##
## Call:
## glm(formula = low ~ smoke + age + high_bp, family = binomial,
##     data = birthweight_analysis)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4547     0.3274  -4.443 8.87e-06 ***
## smoke         0.8574     0.2025   4.234 2.30e-05 ***
## agemiddle     0.3657     0.3251   1.125    0.261
## ageold        0.6373     0.4140   1.540    0.124
## high_bp      -0.2678     0.2390  -1.121    0.262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 600.82  on 483  degrees of freedom
## Residual deviance: 579.56  on 479  degrees of freedom
## AIC: 589.56
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate odds ratios and 95% confidence intervals
exp(cbind(Odds_Ratio = coef(model1), confint(model1)))
```

```
##             Odds_Ratio     2.5 %    97.5 %
## (Intercept)  0.2334722 0.1191090 0.4329588
## smoke        2.3569566 1.5874364 3.5143788
## agemiddle    1.4414876 0.7780857 2.8041941
## ageold       1.8914099 0.8460402 4.3182166
## high_bp      0.7650415 0.4744287 1.2132575
```

Interpretation of Odds Ratios and 95% Confidence Intervals:

- The **odds ratio for smoking during pregnancy is 2.36**, indicating that the odds of low birth weight are 2.36 times higher for children born to mothers who smoke during pregnancy compared to non-smokers, adjusting for maternal age and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (1.59, 3.51), suggesting a statistically significant association between smoking and low birth weight. This is supported by the p-value of 0.00002 (< 0.05), indicating a significant effect of smoking on low birth weight at a 5% level of significance.
- The **odds ratio for middle-aged mothers (20-34 years) is 1.44**, suggesting that the odds of low birth weight are 1.44 times lower for middle-aged mothers compared to young mothers (< 20 years), adjusting for smoking and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.78, 2.8) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.261 (> 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.
- The **odds ratio for old-aged mothers (>= 35 years) is 1.89**, suggesting that the odds of low birth weight are 1.89 times higher for old-aged mothers compared to young mothers (< 20 years), adjusting for smoking and high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (0.85, 4.32) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.124 (> 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.
- The **odds ratio for high systolic blood pressure before pregnancy is 0.765**, suggesting that the odds of low birth weight are 0.765 times lower for mothers with high blood pressure compared to mothers without high blood pressure, adjusting for smoking and maternal age. The 95% confidence interval for the odds ratio is (0.47, 1.21) includes 1, indicating that the effect of high blood pressure on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.262 (> 0.05) which indicates that the effect of high blood pressure is not statistically significant at 5 % level of significance.

Now, let us check the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children, adjusting for maternal age.

Logistic regression model formula: $low \sim smoke + age$

```r
# Build the logistic regression model
model2 <- glm(low ~ smoke + age, data = birthweight_analysis, family = binomial)
summary(model2)
```

```
##
## Call:
## glm(formula = low ~ smoke + age, family = binomial, data = birthweight_analysis)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  -1.5180     0.3228  -4.702 2.58e-06 ***
## smoke         0.8361     0.2012   4.156 3.24e-05 ***
## agemiddle     0.3672     0.3247   1.131   0.2581
## ageold        0.6836     0.4115   1.661   0.0966 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 600.82  on 483  degrees of freedom
## Residual deviance: 580.84  on 480  degrees of freedom
## AIC: 588.84
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate odds ratios and 95% confidence intervals
exp(cbind(Odds_Ratio = coef(model2), confint(model2)))
```

```
##             Odds_Ratio     2.5 %    97.5 %
## (Intercept)  0.2191543 0.1127092 0.4024392
## smoke        2.3074319 1.5578111 3.4307414
## agemiddle    1.4436982 0.7798970 2.8065397
## ageold       1.9809522 0.8907843 4.5025216
```

Interpretation of Odds Ratios and 95% Confidence Intervals:

- The **odds ratio for smoking during pregnancy is 2.31**, indicating that the odds of low birth weight are 2.31 times higher for children born to mothers who smoke during pregnancy compared to non-smokers, adjusting for maternal age. The 95% confidence interval for the odds ratio is (1.55, 3.43), suggesting a statistically significant association between smoking and low birth weight. This is supported by the p-value of 0.00003 (< 0.05), indicating a significant effect of smoking on low birth weight at a 5% level of significance.
- The **odds ratio for middle-aged mothers (20-34 years) is 1.44**, suggesting that the odds of low birth weight are 1.44 times lower for middle-aged mothers compared to young mothers (< 20 years), adjusting for smoking. The 95% confidence interval for the odds ratio is (0.78, 2.8) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.258 (> 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.
- The **odds ratio for old-aged mothers (>= 35 years) is 1.98**, suggesting that the odds of low birth weight are 1.98 times higher for old-aged mothers compared to young mothers (< 20 years), adjusting for smoking. The 95% confidence interval for the odds ratio is (0.89, 4.5) includes 1, indicating that the effect of age on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.097 (> 0.05) which indicates that the effect of age is not statistically significant at 5 % level of significance.

Now, let us check the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children, adjusting for high systolic blood pressure before pregnancy.

Logistic regression model formula: $low \sim smoke + high\_bp$

```r
# Build the logistic regression model
model3 <- glm(low ~ smoke + high_bp, data = birthweight_analysis, family = binomial)
summary(model3)
```

```
##
## Call:
## glm(formula = low ~ smoke + high_bp, family = binomial, data = birthweight_analysis)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0865     0.1449  -7.496 6.59e-14 ***
## smoke         0.8497     0.2016   4.214 2.51e-05 ***
## high_bp      -0.3011     0.2366  -1.273    0.203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 600.82  on 483  degrees of freedom
## Residual deviance: 581.99  on 481  degrees of freedom
## AIC: 587.99
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate odds ratios and 95% confidence intervals
exp(cbind(Odds_Ratio = coef(model3), confint(model3)))
```

```
##             Odds_Ratio     2.5 %    97.5 %
## (Intercept)  0.3373893 0.2522842 0.4457054
## smoke        2.3389677 1.5778925 3.4812369
## high_bp      0.7399825 0.4608777 1.1677571
```

Interpretation of Odds Ratios and 95% Confidence Intervals:

- The **odds ratio for smoking during pregnancy is 2.34**, indicating that the odds of low birth weight are 2.34 times higher for children born to mothers who smoke during pregnancy compared to non-smokers, adjusting for high systolic blood pressure before pregnancy. The 95% confidence interval for the odds ratio is (1.58, 3.48), suggesting a statistically significant association between smoking and low birth weight. This is supported by the p-value of 0.00002 (< 0.05), indicating a significant effect of smoking on low birth weight at a 5% level of significance.
- The **odds ratio for high systolic blood pressure before pregnancy is 0.739**, suggesting that the odds of low birth weight are 0.739 times lower for mothers with high blood pressure compared to mothers without high blood pressure, adjusting for smoking. The 95% confidence interval for the odds ratio is (0.46, 1.17) includes 1, indicating that the effect of high blood pressure on low birth weight is not statistically significant in this model. This is also suggested by the p-value of 0.203 (> 0.05) which indicates that the effect of high blood pressure is not statistically significant at 5 % level of significance.

Finally, let us check the logistic regression model to estimate the total causal effect of smoking during pregnancy on the risk of low birth weight in children without adjusting for any covariates.

Logistic regression model formula: $low \sim smoke$

```r
# Build the logistic regression model
model4 <- glm(low ~ smoke, data = birthweight_analysis, family = binomial)
summary(model4)
```

```
##
## Call:
## glm(formula = low ~ smoke, family = binomial, data = birthweight_analysis)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.1497     0.1372  -8.382  < 2e-16 ***
## smoke         0.8256     0.2002   4.124 3.73e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 600.82  on 483  degrees of freedom
## Residual deviance: 583.65  on 482  degrees of freedom
## AIC: 587.65
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate odds ratios and 95% confidence intervals
exp(cbind(Odds_Ratio = coef(model4), confint(model4)))
```

```
##             Odds_Ratio     2.5 %    97.5 %
## (Intercept)  0.3167421 0.2404626 0.4120208
## smoke        2.2832908 1.5442816 3.3879480
```

Interpretation of Odds Ratios and 95% Confidence Intervals:

- The **odds ratio for smoking during pregnancy is 2.28**, indicating that the odds of low birth weight are 2.28 times higher for children born to mothers who smoke during pregnancy compared to non-smokers. The 95% confidence interval for the odds ratio is (1.54, 3.39), suggesting a statistically significant association between smoking and low birth weight. This is supported by the p-value of 0.00003 (< 0.05), indicating a significant effect of smoking on low birth weight at a 5% level of significance.

Conclusion:

The logistic regression analysis was conducted to address the research question: *Does smoking during pregnancy have a causal effect on the risk of the child being born with low birth weight (below 2500 grams)?*

The results from all the above models consistently show that smoking during pregnancy significantly increases the risk of low birth weight in children. After adjusting for confounders like maternal age, birth order of the child and pre-pregnancy systolic blood pressure, the **odds of low birth weight were found to be approximately 2.3 as high for children born to mothers who smoked during pregnancy**. These findings underscore the importance of smoking cessation interventions during pregnancy to reduce the risk of adverse birth outcomes.

The inclusion of covariates in the model helps to estimate the total causal effect of smoking on low birth weight while accounting for potential confounding factors. The results provide valuable insights for public health initiatives aimed at improving maternal and child health outcomes.

Hence, we conclude that **smoking during pregnancy has a significant causal effect on the risk of low birth weight in children.**

References

[1] Cande V. Ananth, Cande V. Ananth, and Anthony M. Vintzileos. "Epidemiology of preterm birth and its clinical subtypes". In: *The Journal of Maternal-Fetal & Neonatal Medicine* 19.12 (Jan. 2006), pp. 773–782. DOI: 10.1080/14767050600965882. URL: http://dx.doi.org/10.1080/14767050600965882.

[2] Sven Cnattingius. “The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes”. In: *Nicotine & Tobacco Research* 6 (Apr. 2004), pp. 125–140. DOI: 10.1080/14622200410001669187. URL: http://dx.doi.org/10.1080/14622200410001669187.

[3] Smitha D’Sa et al. “Effect of low birth weight on neonatal mortality in preterm and small for gestational age babies in a tertiary neonatal intensive care unit in India”. In: *International Journal of Contemporary Pediatrics* (2016), pp. 735–738. DOI: 10.18203/2349-3291.ijcp20162018. URL: http://dx.doi.org/10.18203/2349-3291.IJCP20162018.

[4] Hong-Kun Di et al. “Maternal Smoking Status During Pregnancy and Low Birth Weight in Offspring: Systematic Review and Meta-Analysis of 55 Cohort Studies Published from 1986 to 2020”. In: *World Journal of Pediatrics* 18.3 (2022), pp. 176–185. DOI: 10.1007/s12519-021-00501-5.

[5] Smadar Eventov-Friedman et al. “Maternal age and outcome of preterm infants at discharge from the neonatal intensive care unit”. In: *International Journal of Gynecology & Obstetrics* 132.2 (Oct. 4, 2015), pp. 196–199. DOI: 10.1016/j.ijgo.2015.06.052. URL: http://dx.doi.org/10.1016/j.ijgo.2015.06.052.

[6] Robert L. Goldenberg et al. “Epidemiology and Causes of Preterm Birth”. In: *The Lancet* 371.9606 (2008), pp. 75–84. DOI: 10.1016/s0140-6736(08)60074-4.

[7] Vincent W. V. Jaddoe et al. “Active and Passive Maternal Smoking During Pregnancy and the Risks of Low Birthweight and Preterm Birth: The Generation R Study”. In: *Paediatric and Perinatal Epidemiology* 22.2 (2008), pp. 162–171. DOI: 10.1111/j.1365-3016.2007.00916.x.

[8] WILLIAM H. JAMES. “Neonatal death and birth order”. In: *Annals of Human Genetics* 33.4 (May 1970), pp. 385–394. DOI: 10.1111/j.1469-1809.1970.tb01665.x. URL: http://dx.doi.org/10.1111/j.1469-1809.1970.tb01665.x.

[9] Ming Jin et al. “Association of pre-/early pregnancy high blood pressure and pregnancy outcomes: a systemic review and meta-analysis”. In: *The Journal of Maternal-Fetal & Neonatal Medicine* 37.1 (Dec. 27, 2023). DOI: 10.1080/14767058.2023.2296366. URL: http://dx.doi.org/10.1080/14767058.2023.2296366.

[10] N Maitra, B Patel, and M Hazra. “Birth outcomes by birth order and maternal nutritional status”. In: *Indian J Matern Child Health* (1995).

[11] G. Mousiolias et al. “Maternal Age as a Predictive Factor of Pre-Term Birth. An Epidemiological Study from 1999 to 2008 in Greece”. In: *Journal of Obstetrics and Gynaecology* 33.1 (2012), pp. 28–31. DOI: 10.3109/01443615.2012.730078.

[12] Nancy E. Reichman and Deanna L. Pagnini. “Maternal Age and Birth Outcomes: Data from New Jersey”. In: *Family Planning Perspectives* 29.6 (1997), p. 268. DOI: 10.2307/2953415.

[13] P. R. Romundstad et al. “Associations of Prepregnancy Cardiovascular Risk Factors with the Offspring’s Birth Weight”. In: *American Journal of Epidemiology* 166.12 (2007), pp. 1359–1364. DOI: 10.1093/aje/kwm272.

[14] T Rosenthal and S Oparil. “Hypertension in women”. In: *Journal of Human Hypertension* 14.10-11 (Oct. 1, 2000), pp. 691–704. DOI: 10.1038/sj.jhh.1001095. URL: http://dx.doi.org/10.1038/sj.jhh.1001095.

[15] Francis A. Tayie and Christina Powell. “Sex Differences in the Association between Prenatal Smoking and Decreased Birthweight, and Intensive Health Care of the Neonate”. In: *Behavioral Medicine* 38.4 (Oct. 2012), pp. 138–142. DOI: 10.1080/08964289.2012.703977. URL: http://dx.doi.org/10.1080/08964289.2012.703977.