A Project Report on

Risk Factors During Preconception Age of Women of Nashik District, Maharashtra.

A project report submitted in partial fulfillment of the requirements for the degree of M.Sc. Statistics (with specialization in Industrial Statistics)

Submitted by

Ms. Patil Priyanka Vasantsing (379371)

Ms. Chaudhari Divya sham (379350)

Mr. Kapadnis Vishal Kishor (379354)

Under the Guidance of

Dr. Rohan D Koshti

('A' Grade NAAC Re-Accredited)

Department of Statistics,

School of Mathematical Sciences,

Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon-425001

(Academic Year : 2023-2024)

CERTIFICATE

This is to certify that ***Ms. Patil Priyanka Vasantsing***, ***Ms. Chaudhari Divya Sham*** and ***Mr. Kapadnis Vishal Kishor*** are students of M.Sc. Statistics (with specialization in Industrial Statistics) at Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon have successfully completed their project entitled ***"Risk Factors During Preconception Age of Women of Nashik District, Maharashtra"*** as a part of M.Sc. Statistics degree program under my guidance and supervision during the academic year 2023-2024.

Place : Jalgaon
Date : 30/5/2024

Dr. Rohan D Koshti

(Project Guide)
Department Of Statistics
K.B.C North Maharashtra University,
Jalgaon.

ACKNOWLEDGEMENT

With the deep sense of gratitude, we express our sincere regards and acknowledgement to **Dr. R. L. Shinde**, Head of Department for his encouragement and motivation during the project work. We are highly thankful to **Dr. R. D. Koshti**, project guide for his guidance. They helped us a lot to overcome all the difficulties, that we faced while writing the project. It was a great pleasure to work with them and learn from **Dr. R. D. Koshti**.

We convey our heartfelt thanks to **Dr. K. K. Kamalja** and **Mr. M. C. Patil** of Department of Statistics for their constructive suggestions, answering our queries and providing us a friendly environment to learn new things. We are also thankful to Department of Statistics, KBCNMU for providing us all infrastructure facilities during project work. We would like to express deepest appreciation towards Dr. Rupeshkumar Deshmukh (Statistician & Asst. Prof.), Mayur Shinde (Statistician),CRPU, BVDU Medical College, Pune whose valuable guidance supported us in completing this project. Our heartfelt thanks and appreciation to our friends and classmates for being with us during the project.

Finally, we would like to thank our parents for their priceless support to us during our studies in KBCNMU, Jalgaon.

Place: Jalgaon
Date: 30/05/2024

Ms. Patil Priyanka Vasantsing (Seat No. 379371)

Ms. Chaudhari Divya Sham (Seat No. 379350)

Mr. Kapadnis Vishal Kishor (Seat No. 379354)

CONTENT Page no.

CHAPTER 1: Introduction

1.1 Motivation

During our course of M.Sc. Statistics with specialization in Industrial Statistics, we have one course of project for 6 credits. Some groups of students make the project on the basis of Industry and some of them like to do analysis on social work. We are interested to do a project on Risk Factors during preconception age of Women from Tribal and Non-Tribal blocks of Nashik District, India.

We were interested in analysing and understanding how presence of different factors during preconception age of women carries risks of bad events for the women and their forthcomimg progeny using different statistical tools and techniques.

1.2 Objectives of Project

Our primary Objectives are as follows :

- To measure the prevalence of various risks factors among women in the preconception period
- To compare the risks between study and comparison blocks .
- Identifying different risk facors for the Outcomes (abortion, still birth, Low birth weight, Pre term, Congenital defects, Neonatal Death).

1.3 General Information

Preconception phase in women though critical is comparatively ignored. The presence of risk factors affects well-being of the woman and her forthcoming progeny. The objectives of the study were to measure the prevalence of various risk factors among women and their comparison between blocks.

The data has been provided by the Bharati Vidyapeeth Deemed University Medical College, Pune, Maharashtra.

The study included women desiring conception within one year. Trained Accredited Social Health Activists collected information through house-to-house visits using a

validated interview schedule. They recorded anthropometric measures of women in a standard manner.

1.4 Scope of the Project

This project will be helpful to understand the Risk Factor present during childbearing age of women. Through this data we wish to predict the results for risk factors with use of Statistical Technique such as Logistic Regression. This project output might be helpful for us and we are confident that we will be able to analyse the surveyed data in future.

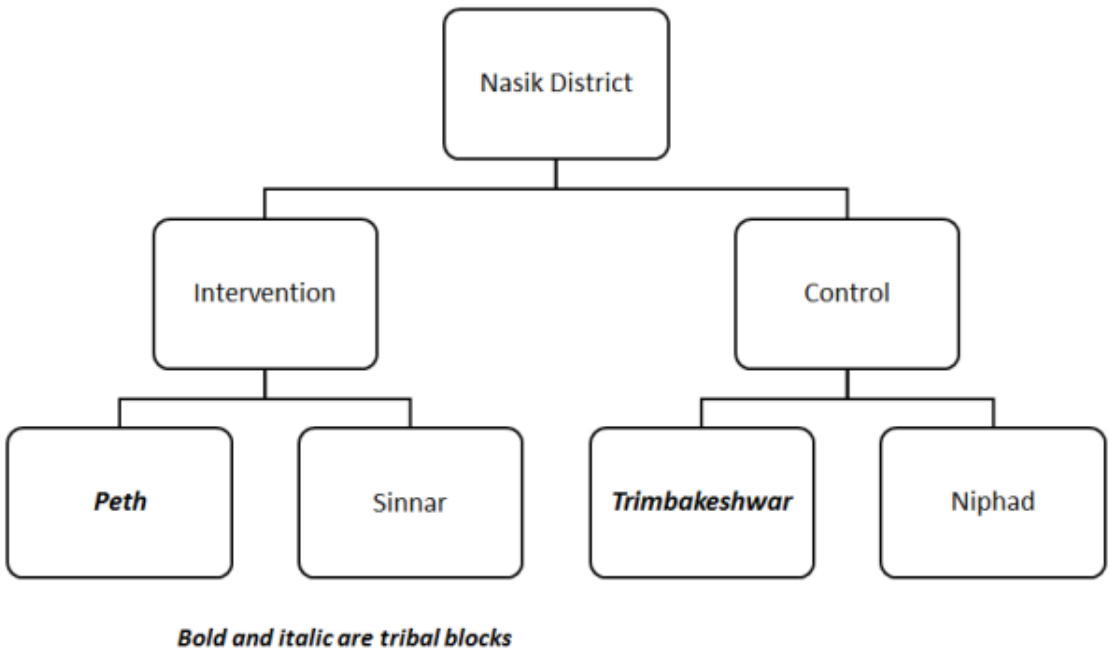
CHAPTER 2: Data Collection & Description

2.1 How Data is Collected ?

STUDY DESIGN: It was a cross-sectional study wherein women of reproductive age group were interviewed and among them desiring pregnancy within one year were further interrogated. This study was an initial phase of an intervention study.

The authors conducted this community-based study in the rural population of Nashik district, Maharashtra, India. In the district, about 57.5% population is rural, 25.6% is tribal. The government of Maharashtra had notified nine blocks out of 15 and 52 Primary Health Centers (PHCs) out of 106 as tribal. They randomly selected one tribal (Peint) and one non-tribal (Sinnar) block for proposed intervention and one adjacent tribal (Trimbakeshwar) plus one adjacent non-tribal (Niphad) block was purposively assigned for comparison (Fig 1). Their geographical location of these blocks is given in Fig 2. The study area covered 28 PHCs. The total population of these four blocks as per the last census was 1,127,902 .

Fig. 1



Tribal women 999 and Non-Tribal women 1244.

The study period was from April to December 2018. Actual enlisting and assessment of desiring women completed in two months, July to August 2018.

Specially trained Accredited Social Health Activists (ASHAs) collected the data through the house-to-house visit. ASHAs deposited the forms at the respective PHC and then PHC sent the forms to the nodal institution fortnightly. The authors created system of a unique identification code for each participant. The centrally trained supervisory staff reviewed and verified the forms for completeness, legibility, and accuracy before data entry. The operator concurrently carried out data cleaning, validity, and data entry.

Fig. 2

Geographical location of study blocks:

2.2 Data Description

For the data collection, women of reproductive age group were interviewed and an Enrollment form filled up by them (The Enrollment Form provided in the Appendix), the enrollment form contain total 28 questions and for the simplicity responses for each question is recorded in coding format as follows.

Q1. Village name, Sub Centre, PHC(Public Health Center), Block

Q2. Name (First name Husband's name Surname)

Q6. Religion

Religion	In Coding Format
Hindu	1
Muslim	2
Bouddha	3
Christian	4
Other	5

Q7. Caste

Caste	In Coding Format
Scheduled caste	1
Scheduled tribe	2
O.B.C.	3
Open	4

Q8. Type of family

Type of family	In Coding Format
Nuclear	1
Joint	2
Other	3

Q10. Education

Education of household head	In Coding Format
Post-graduate/professional degree	1
Graduate degree	2
HSC(12th)/ITI	3
SSC(10th)	4
Seventh pass	5
Less than 7th	6
Illiterate	7

Q11. Your occupation

Your occupation	In Coding Format
Professionals	1
Semi professional	2
Clerical	3
Shop owner	4
Farmer	5
Skilled worker	6
Semiskilled worker	7
Unskilled worker	8
Unemployed/Household	9

Q12. Education of household head

Education of household head	In Coding Format
Post-graduate/ professional degree	1
Graduate degree	2
HSC(12th)/ITI	3
SSC(10th)	4
Seventh pass	5
Less than 7th	6
Illiterate	7

Q13. Occupation of household head

Your occupation	In Coding Format
Professionals	1
Semi professional	2
Clerical	3
Shop owner	4
Farmer	5
Skilled worker	6
Semiskilled worker	7
Unskilled worker	8
Unemployed/Household	9

Q17. Marriage within relation

Marriage within relation	In Coding Format
Yes	1
No	2

Q21. Outcome of last event

Outcome of last event	In Coding Format
Live-birth	1
Still-birth	2
Abortion	3
Early neonatal death	4
Maternal death	5

Q23. Does child have or had congenital anomaly?

Congenital Anomly	In Coding Format
Yes	1
No	2

Q24 A. Do you consume tobacco (more than thrice a week)

Do you consume tobacco	In Coding Format
Yes	1
No	2

Q24 B. If yes, specify

If Yes, specify	In Coding Format
Smoking-cigarette/Bidi	1
Gutkha	2
Snuff	3
Mishri	4

Q25 A. Do you consume alcohol?

Do you consume alcohol?	In Coding Format
Yes	1
No	2

Q25 B. Type of alcohol

Type of alcohol	In Coding Format
Country	1
Tadi / Madi	2
Branded	3

Q26. Generally when do you have meal?

Generally when do you have meal?	In Coding Format
With husband	1
With family members	2
After men	3
Last	4

Q27. Average food quantity

Average food quantity	In Coding Format
Abundant	1
Enough	2
Lesser	3
Remaining	4

Q28. Food groups

Food groups	In Coding Format
Vegetarian	1
Non-vegetarian	2
Sometimes non-vegetarian	3

QUANTATIVE VARIABLES :

Q3. Age (in years)

Q4. Height (in meters)

Q5. Weight (in kg.)

Q14. Total number of family members (count)

Q15. Total family income per month (in 1000 Rs.)

Q16. Age at marriage (in years)

Q18. How many times were you pregnant ? (count)

Q19. How many times did you delivered ? (count)

Q20. Date of last event (Date)

Q22. In which month was the foetus aborted/child delivered ? (in month)

Q24 C. Since when are you consuming tobacco ? (in years & month)

Q25 C. Since when are you consuming alcohol ? (in years & month)

2.3 Some Important Terms

- **Abortion** : Abortion is the termination of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or "spontaneous abortion".
- **Still Birth** : A baby who dies after 28 weeks of pregnancy, but before or during birth, is classified as a stillbirth.
- **Low Birth Weight** : Low birth weight is defined by the World Health Organisation as a birth weight of an infant of 2,499 gm or less, regardless of gestational age.
- **Congenital anomaly** : Congenital anomalies, also known as birth defects, are structural or functional abnormalities (physical disability), including metabolic disorders, that are present from birth of baby.
- **Neonatal Death** : Neonatal death has been defined by World Health Organization (WHO) as "deaths among live births during the first 28 completed days of life" which can be further sub_divided into early neonatal deaths (deaths between 0 and 7 completed days of birth) and late neonatal deaths (deaths after 7 days to 28 completed days of birth).
- **Preterm Birth** : Preterm is defined as babies born alive before 37 weeks of pregnancy are completed.there are sub-categories of preterm birth, based on gestational age : extremely preterm (less than 28 weeks) very preterm (28 to less than 32 weeks) moderate to late preterm (32 to 37 weeks).

These are the outcome variables for our analysis where we are interested in understanding and evaluating the risk of presence of different factor in preconception age of women which results in occurrence of an outcome.

CHAPTER 3: Exploratory Data Analysis (EDA)

3.1 Introduction

Exploratory Data Analysis (EDA) is a crucial initial step in projects. It involves analyzing and visualizing data to understand its key characteristics, uncover patterns, and identify relationships between variables refers to the method of studying and exploring record sets to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables. EDA is normally carried out as a preliminary step before undertaking extra formal statistical analyses or modeling.

3.2 Analysis of Categorical Variables

In Chapter 2(Section 2.2 Data Description), recorded responses for each categorical variables are 2 or > 2, it may results in complexity during the analysis. Hence to avoid further complexity in the analysis each categorical variable is converted in 2 categories only and its frequency distribution is calculated which is giving simple summary of the variables as follows :

- Frequency Distribution of Categorical Variables:

Variable	*Category*	*Frequency*	*Percentage (%)*
Age < 20	1. YES	368	16.41
	2. NO	1854	82.66
Age > 35	1. YES	13	0.58
	2. NO	2209	98.48
BMI < 18.5	1. YES	765	34.11
	2. NO	1294	57.69
BMI > 25	1. YES	135	6.02
	2. NO	1924	85.78
Caste	1. ST	659	29.38
	2. Non-ST	1146	51.09
Family Type	1. Nuclear	303	13.51
	2. Non-Nuclear	1701	75.84
Education	1. < 10^{th}	812	36.20
	2. >= 10^{th}	1300	57.96
Occupation	1. Non-earning	839	37.41
	2. Earning	1257	56.04
Consanguinity Marriage	1. YES	447	19.93
	2. No	1577	70.31
Multiparity	1. YES (>2)	42	1.87
	2.NO	1119	49.89
Adverse Outcome Previous	1. YES	54	2.41
	2.NO	966	43.07
Tobacco or Smoking or alcohol	1. YES	116	5.17
	2. NO	2126	94.78
Caloriy Intake < 50 %	1. YES	59	2.63
	2.NO	2182	97.28
Protein Intake < 50 %	1. YES	64	2.85
	2.NO	2177	97.06
Abortion	0. NO	1379	61.48
	1. YES	78	3.48
Still Birth	0. NO	1445	64.42
	1. YES	12	0.53
Congenital Anomaly	0. NO	1356	60.45
	1. YES	11	0.49
Neonatal Death	0. NO	1361	60.68
	1. YES	6	0.27
Adverse Outcome	0. NO	1361	60.68
	1.YES	96	4.28
Preterm	0. NO	1185	52.83
	1.YES	182	8.11

3.3 Analysis of Continuous Variables :

Continuous Variables :

i) Age (yrs and mnth)
ii) Height (cm)
iii) Weight (kg)
iv) BMI
v) Family Income (Rs.)
vi) Age at Marriage (yrs and mnth)
vii) Baby Weight (kg)
viii) Gestational age of Mother (mnths)

I) Summary Statistics :

OBJECTIVE :

To understand all the continuous variable in summary format i.e how much responses available for each variable(N), what is average values(mean), CI, range (max-min), is there any missing values ?,...etc.

Variable	*N*	*Mean*	*Std. Dev.*	*Min*	*25th Percentile*	*75th Percentile*	*Max*
Age (yrs and mnth)	2222	23	3.6	14	20	25	40
Height (cm)	2102	152	21	15	147	155	551
Weight (kg)	2152	45	7.6	2.3	40	49	95
BMI	**2059**	**20**	**3.4**	**7.2**	**18**	**21**	**40**
Family Income	2013	10613	14283	4	3000	15000	210000
Age at marriage	2141	19	2.6	2	18	20	38
Baby weight	909	2.7	1.1	1.1	2.5	3	32
Gestational age of Mother	515	8.9	0.53	2	9	9	9.2

Summary Table

CONCLUSION :

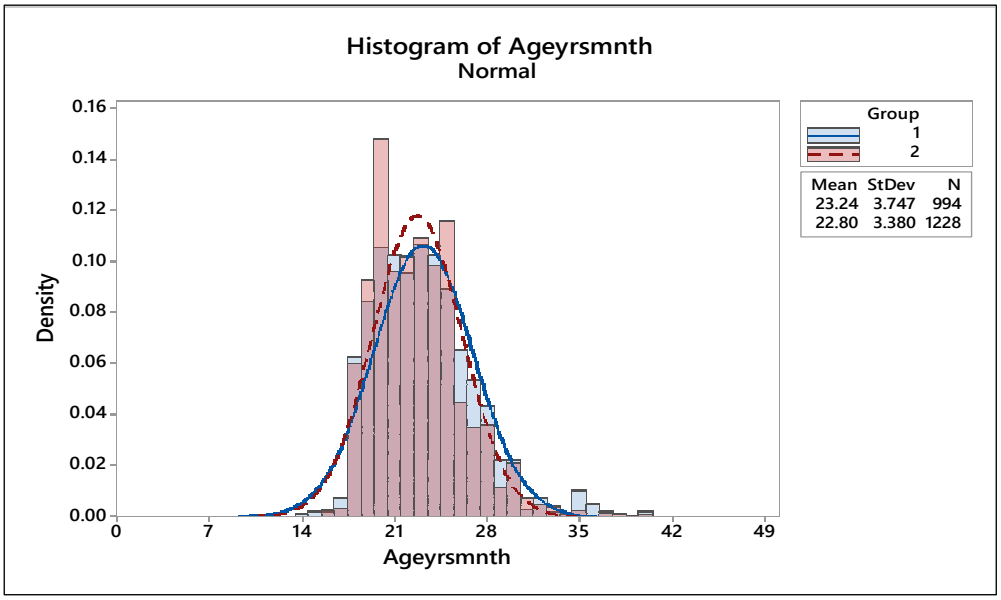
The summary table providing an idea about the variables, for eg. Out of total women in the study BMI of 2059 is available/recorded, in an average women have BMI 20 ($\pm$3*3.4)(women in both the blocks are included). Out of total 25% of women have BMI <18 and 75% of women have BMI < 21. Women with BMI 7.2 (minimum) & 40 (maximum) are also present in the study. Simillarly, the summary table provides conclusions for the other variables.

II) Histogram & Boxplot :

i) Age (yrs & mnths) :

OBJECTIVE :

To understand the distribution of Ages of women in Intervention and Control group and to comapre the.



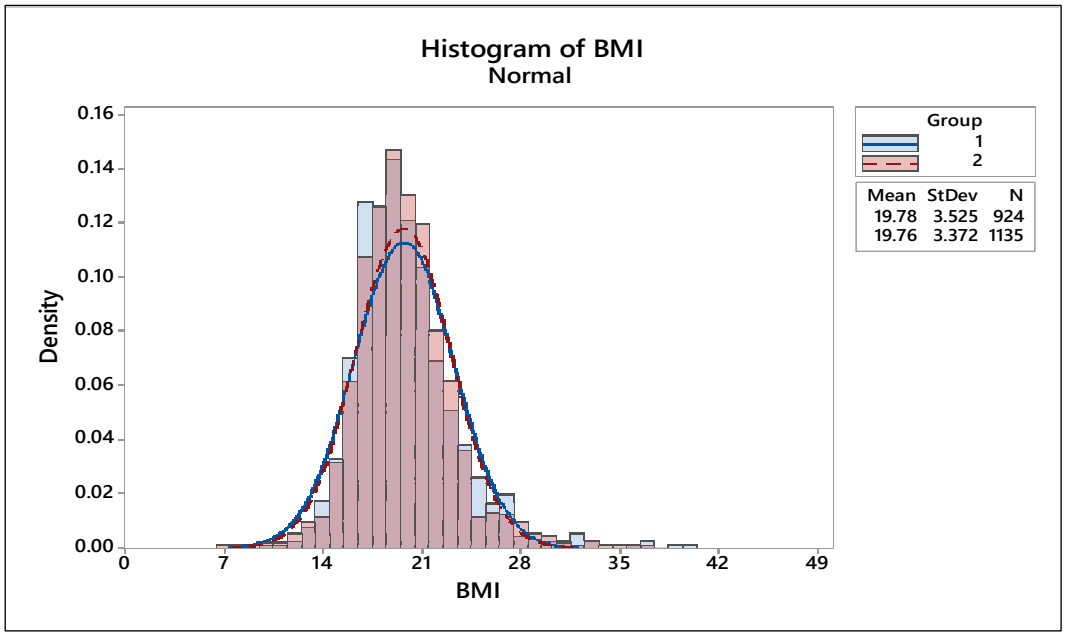
CONCLUSION :

By observing the histogram of Age, the distribition of ages in Intervention group as well as in control group seems to be approximately symmetric around 23.24 & 22.80 resp. Women in both Intervention and Control group have approximately same ages.

ii) BMI :

OBJECTIVE :

To understand the distribution of BMI of women in Intervention and Control group and to comapre them.



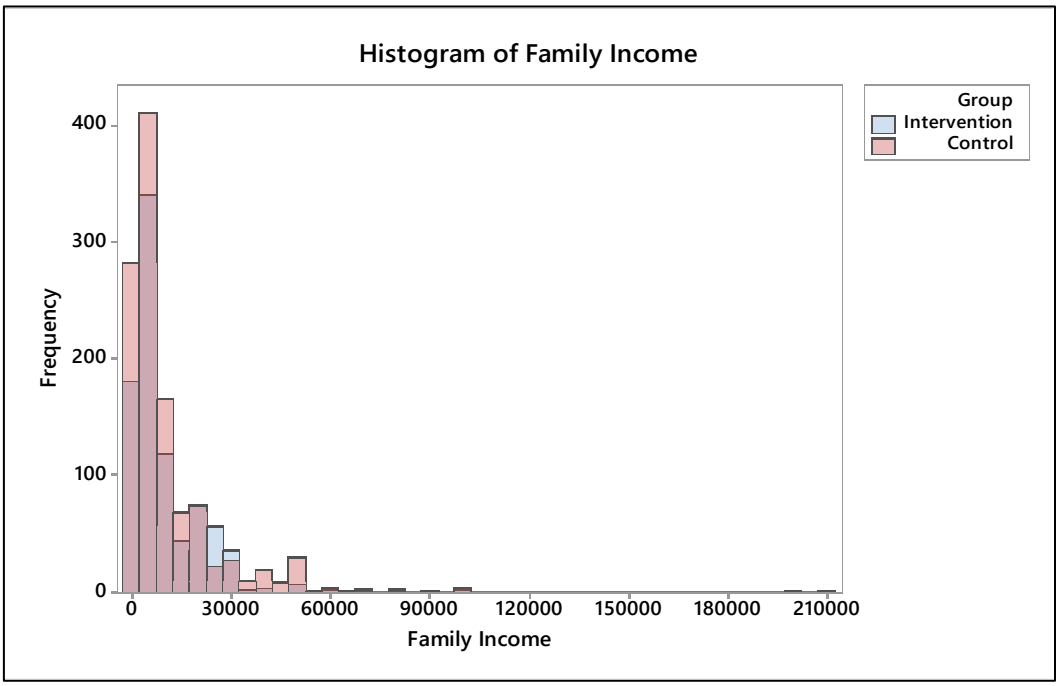
CONCLUSION :

Histogram of BMI for both intervention and Control group showing the distribution of BMI is approximately symmetric around 19.78 & 19.76 for Intervention and Control group resp.

iii) Family Income :

OBJECTIVE :

To find maximum women have family income in high or low group & to compare them in intervention and control group.

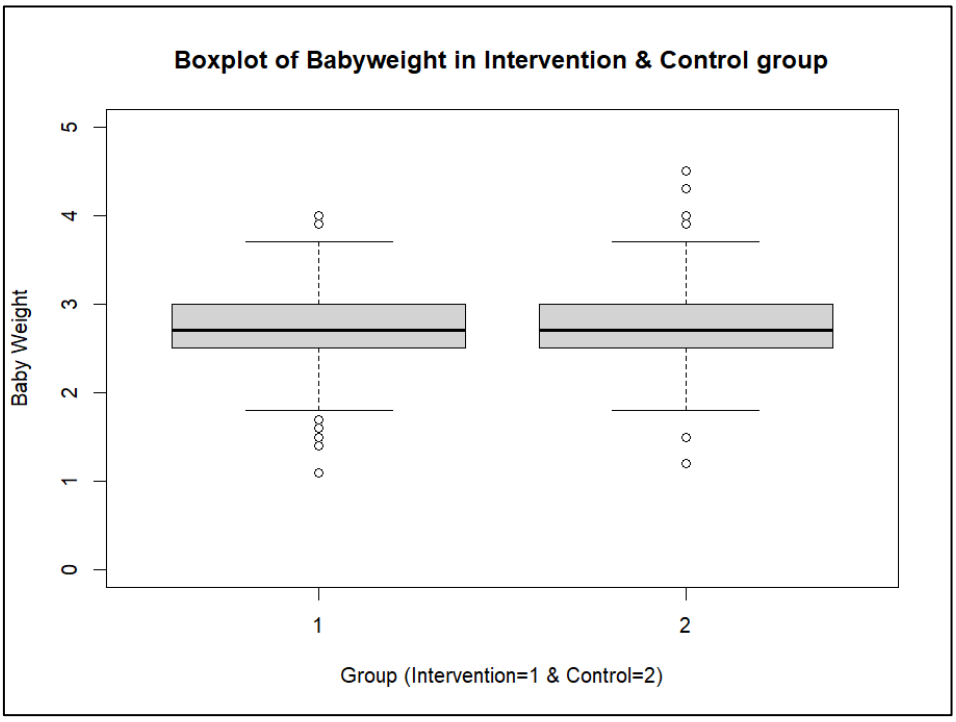


CONCLUSION :

The distribution for income we got is positively skewed indicating that maximum women have low family income where as very few are from families with high income, this result holds for both intervention and control group.

iv) Babyweight (Kg.):

OBJECTIVE : To compare weight of baby in two groups, Intervention & Control.



CONCLUSION : The boxplot for Babyweight is showing the disrtibution of babyweight is in both Intervention & Control group is more or less symmetric indicating average baby have weight approximately 2.7 (these results are given when there are some outliers present in the data).

3.4 Summary :

Performing EDA for different continuous and Categorical factors gave an idea about the data. For different continuous variables the distributions are more or less similar in Intervention & Control group i.e the variables do not follow any pattern if they are from intervention group or from control group. (Results are given approximately since some outliers are present in data)

(Analysis is done using R & Minitab softwares, required R codes provided in the Appendix)

CHAPTER 4: Use of Statistical Tests & Tools

4.1 Introduction

Understanding the relationships between categorical variables and assessing the risk associated with different exposures are crucial aspects of data analysis in many fields, including medical research, epidemiology, public health, and social sciences. In this chapter, we will explore the results obtained from two key statistical methods: the chi-square test and risk ratios. These methods provide insights into the association between variables and the relative risk of outcomes, aiding in the interpretation and decision-making processes based on our dataset.This chapter aims to present and interpret the results of these analyses, highlighting significant findings and their implications for our study.

4.2 Objective

The primary objective of this chapter is to interpret the results of the chi-square test, risk ratios & two sample proportion Z-test obtained from our dataset. Specifically, we aim to:

Assess the Association Between Categorical Variables: Determine whether there is a significant association between any two variables using the chi-square test.

Evaluate the Relative Risk of Outcomes: Calculate and interpret the risk ratios for Outcome across different exposure groups defined by Risk Factors, identifying groups at higher risk & to compare the proportion of outcomes in Intervention and Control group.

4.3 Chisquare test

The chi-square test is a widely used statistical tool to examine the association between two categorical variables. It helps to determine whether the observed distribution of data deviates significantly from the expected distribution under the null hypothesis of independence or no association.

Key Outputs of the Chi-Square Test :

- Chi-Square Statistic (χ^2) :

This is the main output of the test. It quantifies the difference between the observed and expected frequencies. A higher χ^2 value indicates a greater difference between the observed and expected data.

- P-Value :

The p-value indicates the probability of obtaining a chi-square statistic at least as extreme as the one observed, assuming the null hypothesis is true. A low p-value (typically < 0.05) suggests that there is a significant association between the variables.

- Contingency Table :

This table displays the observed frequencies for each category. It helps in visualizing the data distribution across different categories.

4.3.1 Objective, Description & Conclusion of Table 1 :

Objective :

To study whether the predictors are associated or dependent with the groups intervention and control using chi-square test of association.

Description :

From Table 1 let us interpret for the Caste risk factor, the highlighted cross table Caste*Group indicates 343(38.93%) of women are from intervention group and having caste ST, 316(34.20%) women are from control group with caste ST, 538(61.07%) women are from intervention group with non-ST caste and 608(65.805) women are from control group with non-ST caste i.e there are total 659(36.51%) women with ST caste and 1146(63.49%) women with non-ST caste.

Similarlly description for risk factors Age Group, BMI group, Family Type, Education, Occupation, Consanguinity Marriage, Parity, Preadverse outcomes, Tobacco/Smoking/Alcohol Consumer, Calorie Intake and Protein Intake can be given.

Table 1

Chisquare test of association for different risk factors and group(Intervention & control)						
Predictors/Factors		**Group**		**Total**	**Chi-square value**	**p-value**
		Intervention	**Control**			
Age Group	<20	164 (16.49)	204 (16.61)	368 (16.56)	3.174	0.205
	20-35	821 (82.59)	1020 (83.06)	1841(82.85)		
	>35	9 (0.90)	4 (0.32)	13 (0.59)		
BMI group	<18.5	357 (38.64)	408 (35.95)	765 (37.15)	2.72	0.257
	18.5-25	502 (54.33)	657 (57.88)	1159 (56.29)		
	>25	65 (7.03)	70 (6.17)	135 (6.56)		
Caste	ST	343 (38.93)	316 (34.20)	659 (36.51)	4.15	**0.041**
	Non ST	538 (61.07)	608 (65.80)	1146 (63.49)		
Family type	Nuclear	160 (17.82)	143 (12.93)	303 (15.12)	8.85	**0.002**
	Non nuclear	738 (82.18)	963 (87.07)	1701 (84.88)		
Education	<10th Pass	357 (37.54)	455 (39.19)	812 (38.45)	0.53	0.465
	>10th Pass	594 (62.46)	706 (60.81)	1300 (61.55)		
Occupation	Non Earning	335 (35.19)	504 (44.06)	839 (40.03)	16.65	**<0.001**
	Earning	617 (64.81)	640 (55.94)	1257 (59.97)		
Consanguinity Marriage	Yes	181 (19.53)	266 (24.25)	447 (22.08)	6.24	**0.012**
	No	746 (80.47)	831 (75.75)	1577 (77.92)		
Parity	>2	18 (3.30)	24 (3.87)	42 (3.61)	0.126	0.722
	<2	526 (96.69)	596 (96.12)	1122 (96.39)		
Pre Adverse Outcome	Yes	24 (4.91)	30 (5.65)	54 (5.29)	0.151	0.698
	No	465 (95.09)	501 (94.35)	966 (94.71)		
Tobacco/ Smoking/Alcohol consumer	Yes	61 (6.11)	55 (4.42)	116 (5.17)	2.89	**0.089**
	No	937 (93.89)	1189 (95.58)	2126 (94.83)		
Calorie Intake	<50 Per cent	25 (2.51)	34 (2.74)	59 (2.63)	0.042	0.837
	>=50 Per cent	973 (97.49)	1209 (97.26)	2182 (97.38)		
Protein Intake	<50 Per cent	30 (3.01)	34 (2.74)	64 (2.85)	0.065	0.799
	>=50 Per cent	968 (96.99)	1209 (97.26)	2177 (97.14)		

Conclusion :

To check whether there is any association between caste and group or they are independent we have applied chisquare test of association and we got p-value as 0.041 i.e. we will reject our null hypothesis that caste being ST and non-ST is independent with group being Intervention or Control i.e these variables are dependent with group.

Similarlly conclusions for other predictors/factors can be given. The variables Family Type(Nuclear, Non-nuclear), Occupation(Non-Earning, Earning) and Consanguinity marriage(Yes, No) have p-values 0.002, <0.001, 0.012 resp, hence at 5% l.o.s we can say Family Type, Occupation and Consanguinity marriage are associated with group(Intervention, Control).

Simillarly, at 5% l.o.s the risk factors age group, BMI group, Education, Parity, Pre Adverse Outcome, Tobacco/Smoking/Alcohol Consumer, Calorie Intake and Protein Intake are not associated with group(intervention, Control). At 10% l.o.s risk factor Tobacco/Smoking/Alcohol Consumer is associated with group since p-value is <0.10.

4.4 Risk Ratio :

4.4.1 What is Risk Ratio & how it is calculated ?

The Relative Risk(RR) also known as Predicted Risk(PR) is the ratio of risk of an event in one group(eg. Exposed group) vs the risk of the event in the other group(eg. Nonexposed group).

To calculate PR :

Consider table:

Exposure	Outcome		
		Yes	No
	Yes	A	B
	No	C	D

$$RR=\frac{\left(\frac{A}{A+B}\right)}{\left(\frac{C}{C+D}\right)}$$

RR=1 : means the exposure does not affect the outcome

RR<1 : means that the risk of outcome is decresed by the exposure, which is a “protective factor”

RR>1 : means that the risk of outcome is incresed by the exposure, which is a “risk factor”

4.4.2 Objective, Description & Conclusion of Table 2 :

Objective :

To study whether the different predictors have the risk of an outcome(abortion, still birth, Low birth weight, Pre term, Congenital defects, Neonatal Death) in both Intervention group and Control group. Also to compare where(whether in Intervention or in Control group) women have more chances of risk of an event.

Description :

Table 2 shows crosstables for different risk factors and the events like abortion, still birth, Low birth weight, Pre term, Congenital defects, Neonatal Death. This table shows the outcomes for Intervention group(where the responses recorded for the women who have given some treatment during the preconceptional period).for example consider for Risk factor Family Type and outcome Abortion, here out of 61 women from Intervention group and having nuclear family 10 had abortion and remaining 51 do not had abortion also out of 377 women from Intervention group and having non-nuclear family 30 had abortion and remaining 347 do not had abortion.

Smilar results for other risk factors and outcomes(abortion, still birth, Low birth weight, Pre term, Congenital defects, Neonatal Death) are performed for control group.

(These results are calculated using R software & Risk Ratio calculator & R codes provided in the Appendix)

Conclusion :

For Intervention group as highlighted in Table 2, women with nuclear family type have double risk of abortion than women with non-nuclear family type. Also The women with caste ST are more protective in case of abortion than the non-ST women. Similarly women with less education($<10^{th}$) have more risk of still birth than more educated women but cast being ST is not a risk factor for still birth.

Table 2

Risk Ratio table for different Risk factors and Outcomes in Intervention Group									
		Abortion		Total	PR(CI)	Still birth		Total	PR
		Yes	No			Yes	No		
Age	<20	5	88	93	0.5579	2	91	93	2.0645
	20-35	37	347	384	1	4	380	384	1
	>35	0	2	2	0	0	2	2	0
Parity	>2	1	6	7	1.4047	0	7	7	0
	<2	24	212	236	1	3	233	236	1
Family Type	Nuclear	10	51	61	2.0601	1	60	61	1.2360
	Non-nuclear	30	347	377	1	5	372	377	1
Occupation	Non-Earning	14	143	157	1.0522	0	157	157	0
	Earning	25	270	295	1	6	289	295	1
Education	<10th Pass	10	165	175	0.5276	2	173	175	1.0552
	>10th Pass	30	247	277	1	3	274	277	1
Caste	ST	11	145	156	0.6711	1	155	156	0.3294
	Non-ST	27	230	257	1	5	252	257	1
Consanguineous Marriage	Yes	8	75	83	1.1566	2	81	83	4.3373
	No	30	330	360	1	2	358	360	1
Calorie Intake	<50%	2	9	11	2.1409	0	11	11	0
	>=50%	40	431	471	1	6	465	471	1
BMI	<18.5	15	165	180	0.8	2	178	180	0.6666
	18.5-25	25	215	240	1	4	236	240	1
	>25	0	27	27	0	0	27	27	0
Pre Adverse Outcome	Yes	2	6	8	2.8947	0	8	8	0
	No	19	201	220	1	2	218	220	1
Tobacco or Smoking or Alcohol consumer	Yes	0	27	27	0	0	27	27	0
	No	42	412	454	1	6	448	454	1
Protein Intake	< 50%	3	12	15	2.3948	0	15	15	0
	>=50%	39	428	467	1	6	461	467	1
Intervention	Yes	42	440	482	2.3599	6	476	482	2.0228
	No	36	939	975	1	6	969	975	1

In summary for **INTERVENTION** group we got risk factors for :

- ABORTION : Parity (having child >2), Family Type (Nuclear), Occupation (Non-Earning), Consangeneous marriage (Yes), Calorie Intake (<50%), Pre-adverse Outcome (Yes), Protein Inatake (<50%).
- STILL BIRTH : Age (<20), Family Type (Nuclear), Education (<10th), Consanguineous marriage (Yes).
- LOW BIRTH WEIGHT : Age (>35), Education (<10th), Caste (ST), Consanguineous marriage (Yes), BMI(<18.5), BMI (>25), Tobacco/Smoking/Alcohol consumer (Yes).
- PRE-TERM : Age (>35), Occupation (Non-Earning), Education(<10th), Caste (ST), Calorie Intake (<50%), BMI (>25), Pre-Adverse Outcome (yes), Tobacco/Smoking/Alcohol consumer (Yes), Protein Intake (<50%)
- CONGENIAL DEFECTS : Age (<20), Education (<10th), Caste (ST), Consanguineous marriage (Yes), BMI (<18.5).
- NEONATAL DEATH : Occupation (Non-Earning).

In summary for **CONTROL** group we got risk factors for :

- ABORTION: Family Type (Nuclear), Occupation (non-Earning), Education (<10th), Caste (ST), Protein Intake (<50%)
- STILL BIRTH : Family Type (Nuclear), Education (<10th), Caste (ST), Pre-Adverse outcome (Yes).
- LOW BIRTH WEIGHT : Age (<20), Family Type (Nuclear), Occupation (Non-Earning), Consanguineous marriage (Yes), BMI (<18.5), Pre-Adverse Outcome (yes).
- PRE-TERM : Parity (>2), Occupation (non-Earning), Education (<10th), BMI (<18.5), Pre-Adverse Outcome (Yes), Tobacco/Smoking/Alcohol cosumer (Yes).
- CONGENIAL DEFECTS : Age (<20), Education (<10th), Caste (ST), Consanguineous marriage (Yes).
- NEONATAL DEATH : Age (<20), Consanguineous marriage (Yes).

Objective and Conclusion for Table 3 :

OBJECTIVE :

The main objective for constructing the table is to find that is being in intervention group have the minimum risk of the outcomes (abortion, Still Birth, Low Birth Wt, Pre term,Congenial Anomly, Neonatal Death) for the women than being in control group.

Table 3

		Abortion			Still Birth			Low birth Weight	
		Yes	No		Yes	No		Yes	No
Intervention	No	36	939		6	969		86	795
	Yes	42	440		6	476		35	385
	Total	78	1379		12	1445		121	1180
	PR	0.4237	1		0.4943	1		1.1914	1
		Pre term			Congenial Anomly			Neonatal Death	
		Yes	No		Yes	No		Yes	No
Intervention	No	129	804		5	958		4	929
	Yes	53	381		6	428		2	432
	Total	182	1185		11	1356		6	1361
	PR	1.1322	1		0.3876	1		0.9309	

CONCLUSION :

From the table we can conclude that for the women in control group (i.e when no any treatment is provided from preconception age) has almost 20% higher risk of Low birth weight for their forthcoming progeny than the women in intervention group (i.e when some treatment is provided from preconception age).

4.5 Two sample Z-Proportion Test

4.5.1 About the Test :

Two sample Z-test of proportions is the test to determine whether the two populations differ significantly on specific characteristics. It used to compare the proportion of two different populations that have some single characteristic.

ASSUMPTIONS : To perform this test following assumptions must satisfied :

i) The data are simple random values from both populations

ii) Both population follow binomial distribution

iii) Samples are independent of each other

4.5.2 Objective, Description & Conclusion of Table 4 :

Objective :

Here the objective is to evaluate the proportion of occurrence of outcomes (Abortion, Still Birth, Low Birth Weight, Pre Term, Congenial Defects, Neonatal Death) in Intervention and control group is same or differs significantly.

HYPOTHESIS :

H0 : The proportion of outcomes in Intervention group is same as the proportion of outcomes in Control group. ($p_i = p_i'$)

H1 : The proportion of outcomes in Intervention group is less than the proportion of outcomes in Control group. ($p_i < p_i'$)

Description :

From table 3 : Let us interprete for Outcome Abortion, N=482 indicates there are total 482 women in the Intervention group who have responded for the outcome Abortion (Yes/No) & n=42 indicates out of 482 responses 42 women had their abortion i.e the proportion of abortion in intervention group is 8.71% . Simillarly in the Control group

N=975 women responded for outcome Abortion (Yes/No) out of those n=36 had the event of abortion i.e the proportion of abortion in Control group is 3.69% (The results are based on the sample data). Similar description for other outcomes can be given from the table

(These results are calculated using R Software & R codes are provided in the Appendix)

Table 4

Z-test difference between two proportions							
Outcomes	Intervention			Control			p-value
	N	n	Proportion (%) (p_i)	N	n	Proportion (%) (p_i')	
Abortion	**482**	**42**	**8.71**	**975**	**36**	**3.69**	**<0.01**
Still Birth	482	6	1.24	975	6	0.62	0.1089
Low birth weight	417	35	8.39	881	86	9.76	0.2139
Pre term	434	53	12.21	933	129	13.83	0.2059
Congenial defects	434	6	1.38	933	5	0.54	0.0528
Neonatal Death	434	2	0.46	933	4	0.43	0.4688

Conclusion :

By observing the p-values for each outcome variable obatained through two sample proportion test we can conclude, the proportion of Abortion in Intervention group is significantly different (lesser) from the proportion of abortion in the Control group. And the proportion of Still Birth, Low Birth Weight, Pre term Baby, Congenial Defects, Neonatal Death in Intervention are not significantly different from Control group.

In summary women in Intervention group have more protection in the case of abortion than the women in Control group (i.e providing some treatment to women during preconception age lessers the risk of bad outcomes for the women and their forthcoming progeny).

CHAPTER 5: Identification & Evaluation of Risk Factors

5.1 Introduction

In this chapter, we present the results of our logistic regression analysis aimed at identifying factors associated with outcome variables. We conducted both univariate and multivariate logistic regression analyses to determine the impact of various predictors on the likelihood of outcomes. Logistic regression is appropriate for our study because it allows us to model the relationship between a binary outcome variable and several predictor variables (which are also binary here), providing insights into the factors that significantly influence outcomes.

5.2 Objective

To perform Univariate Logistic Regression to examine the effect of each predictor on outcome individually

5.3 Logistic Regression & Odds Ratio :

LOGISTIC REGRESSION is a statistical method used to model and analyze datasets in which the outcome variable is binary (i.e., it takes on two possible outcomes). Unlike linear regression, which predicts a continuous outcome, logistic regression predicts the probability of a categorical outcome. Logistic regression uses Maximum Likelihood Estimation to estimate the parameters. It derives the relationship between a set of variables (independent) and a categorical variable (dependent).

An **ODDS RATIO (OR)** is a measure of association between an exposure and an outcome. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure. Odds ratios are most commonly used in case-control studies, however they can also be used in cross-sectional and cohort study designs as well.

OR=1 Exposure does not affect odds of outcome OR>1 Exposure associated with higher odds of outcome OR<1 Exposure associated with lower odds of outcome

5.3.1 Univariate Logistic Regression :

We have performed univariate logistic regression for each Outcome variable (Abortion, Still Birth, Low Birth Weight, Pre term Baby, Congenial Defects, Neonatal Deaths) using each Predictor variable (Age < 20, Age > 35, BMI < 18.5, BMI > 20, Caste, Family Type, Education, Occupation, Cansanguinity Marriage, Parity, Pre-Adverse Outcome, Tobacco/Smoking/Alcohol, Protein Intake < 50%, Calorie Intake < 50%) to know the effect of each predictor on outcome individually (For Intervention & Control both groups).

Using Statistical Software R models are fitted and results are provided (R codes are provided in the Appendix).

For Intervention Group :

The tables provides a brief interpretation of the logistic regression results for each Outcome (The table include the predictors with the corresponding estimates of parameters in fitted model and the odds ratios) for Intervention group (i.e. women who had received some treatment during preconception age).

i) ABORTION :

Result of logistic model when Abortion is an outcome variable :

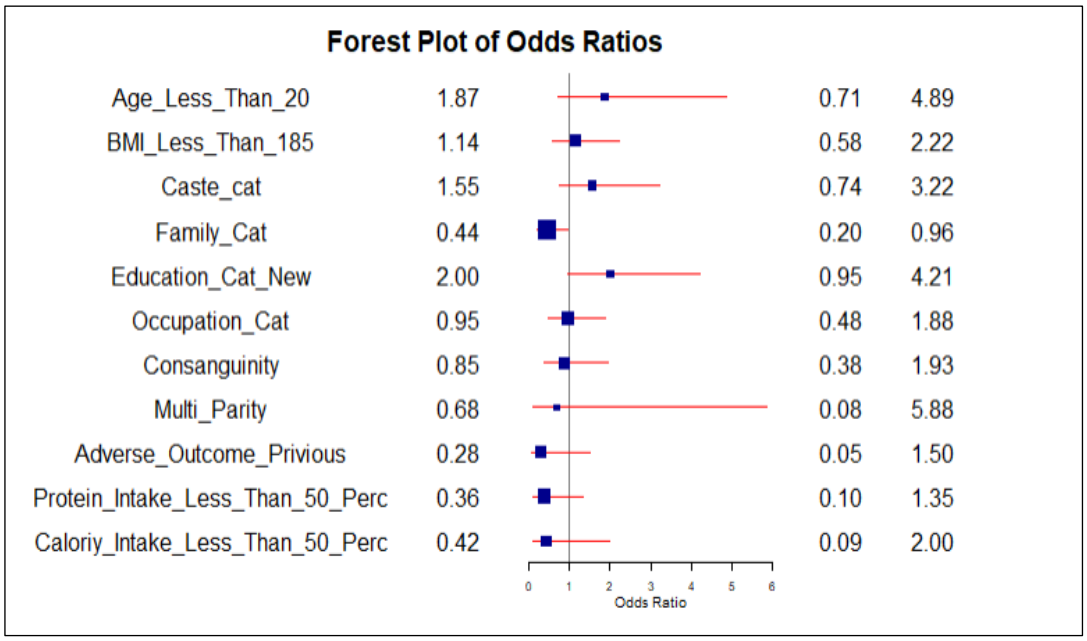
Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-3.49	1.87
	Age < 20 (Yes)	0.62	
BMI < 18.5	(Intercept)	-2.53	1.14
	BMI < 18.5 (Yes)	0.13	
Caste	(Intercept)	-3.02	1.55
	Caste (ST)	0.44	
Family Type	(Intercept)	-0.81	0.45
	Family Type (Nuclear)	-0.82	
Education	(Intercept)	-3.50	2.01
	Education (<10^{th})	0.70	
Occupation	(Intercept)	-2.27	0.95
	Occupation (Non-Earning)	-0.06	
	(Intercept)	-2.08	0.86

Consanguinity marriage	Consanguinity marriage (Yes)	-0.16	
Multi Parity	(Intercept)	-1.40	0.68
	Multi Parity (Yes)	-0.39	
Adverse Outcome Previous	(Intercept)	0.16	0.29
	Adverse Outcome Previous (Yes)	-1.26	
Protein Intake < 50 %	(Intercept)	-0.38	0.37
	Protein Intake < 50 % (Yes)	-1.01	
Calorie Intake < 50 %	(Intercept)	-0.63	0.42
	Calorie Intake < 50 % (Yes)	-0.87	

INTERPRETATION : Woman with age < 20 have 1.87 times i.e approximately double risk of getting abortion than women having age ≥ 20. Lly women with Caste (ST), BMI (<18.5), Education (<10th) have more risk of abortion than women with Caste (non-ST), BMI (≥ 18.5), Education (≥10th).

FOREST PLOT :

This plot is providing us an visual idea of estimated results (Odds ratio with corresponding Confindence Intervals)



ii) STILL BIRTH :

Result of logistic model when Still Birth is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-3.08	0.48
	Age < 20 (Yes)	-0.74	
BMI < 18.5	(Intercept)	-4.79	1.36
	BMI < 18.5 (Yes)	0.30	
Caste	(Intercept)	-6.17	3.08
	Caste (ST)	1.12	
Family Type	(Intercept)	-3.88	0.81
	Family Type (Nuclear)	-0.22	
Education	(Intercept)	-4.41	0.95
	Education (<10th)	-0.05	
Consanguinity marriage	(Intercept)	-2.22	0.23
	Consanguinity marriage (Yes)	-1.49	

INTERPRETATION : Woman with Caste (ST) have 3.08 times i.e approximately 3 times risk of Still Birth than women having Caste (non-ST). Lly women with BMI (<18.5), have more risk of Still Birth than women with BMI (≥ 18.5).

iii) LOW BIRTH WEIGHT

Result of logistic model when Low Birth Weight is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-3.10	1.49
	Age < 20 (Yes)	0.39	
Age < 35	(Intercept)	2.41	0.09
	Age < 35 (Yes)	-2.41	
BMI < 25	(Intercept)	-2.59	1.17
	BMI < 25 (Yes)	0.15	
BMI < 18.5	(Intercept)	-1.80	0.73
	BMI < 18.5 (Yes)	-0.32	
Caste	(Intercept)	-0.83	0.36
	Caste (ST)	-1.04	
Family Type	(Intercept)	-3.17	1.56
	Family Type (Nuclear)	0.44	

Education	(Intercept)	-1.85	0.73
	Education (<10th)	-0.32	
Occupation	(Intercept)	-3.05	1.50
	Occupation (Non-Earning)	0.40	
Consanguinity Marriage	(Intercept)	-2.36	0.98
	Consanguinity Marriage (Yes)	-0.02	
Tobacco/Smoking	(Intercept)	-0.39	0.36
	Tobacco/Smoking (Yes)	-1.05	

INTERPRETATION : Woman with age < 20 have 1.47 times i.e approximately 49% more risk of Low Birth Weight than women having age ≥ 20. Lly women with BMI (<25), Family Type (Nuclear), Occupation (non-Earning) have more risk of Low Birth Weight than women with BMI (≥ 25), Family Type (non-Nuclear), Occupation (Earning).

iv) PRE-TERM

Result of logistic model when Pre Term is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-2.11	1.09
	Age < 20 (Yes)	0.08	
Age > 35	(Intercept)	1.98	0.14
	Age > 35 (Yes)	-1.98	
BMI > 25	(Intercept)	-1.00	0.62
	BMI > 25 (Yes)	-0.48	
BMI < 18.5	(Intercept)	-2.19	1.18
	BMI < 18.5 (Yes)	0.16	
Caste	(Intercept)	-1.20	0.63
	Caste (ST)	-0.46	
Family Type	(Intercept)	-2.37	1.20
	Family Type (Nuclear)	0.18	
Education	(Intercept)	-1.11	0.55
	Education (<10th)	-0.60	
Occupation	(Intercept)	-1.86	0.93
	Occupation (Non-Earning)	-0.08	
Consanguinity Marriage	(Intercept)	-2.24	1.17
	Consanguinity Marriage (Yes)	0.15	
	(Intercept)	-1.49	0.89

Adverse Outcome Previous	Adverse Outcome Previous (Yes)	-0.12	
Tobacco/Smoking	(Intercept)	-1.42	0.76
	Tobacco/Smoking (Yes)	-0.28	
Protein Intake < 50 %	(Intercept)	-0.19	0.41
	Protein Intake < 50 % (Yes)	-0.91	
Calorie Intake < 50 %	(Intercept)	-0.51	0.48
	Calorie Intake < 50 % (Yes)	-0.74	

INTERPRETATION : Age < 20 (Yes), BMI < 18.5 (Yes), Family Type (Nuclear), Consanguinity Marriage (Yes) are the risk factors for Pre-Term baby. (i.e presence of these predictors may carry more risk of event Pre Term baby.

v) CONGENIAL DEFECTS

Result of logistic model when Congenial Defects is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-3.03	0.50
	Age < 20 (Yes)	-0.71	
BMI < 18.5	(Intercept)	-3.59	0.69
	BMI < 18.5 (Yes)	-0.38	
Caste	(Intercept)	-2.40	0.32
	Caste (ST)	-1.16	
Education	(Intercept)	-3.57	0.67
	Education (<10th)	-0.41	
Occupation	(Intercept)	-5.96	2.75
	Occupation (Non-Earning)	1.01	
Consanguinity Marriage	(Intercept)	-1.61	0.22
	Consanguinity Marriage (Yes)	-1.54	

INTERPRETATION : By this table, women with Occupation (non-Earning) has 2.75 times more risk of Congenial defects than women with Occupation (earning).

vi) NEONATAL DEATH

Result of logistic model when Neonatal Death is an outcome variable :

Predictor	Estimate	Odds Ratio

Occupation	(Intercept)	-4.34	0.54
	Occupation (Non-Earning)	-0.62	

INTERPRETATION : By this model, Occupation (non-Earning) carries 46% less risk of outcome Neonatal Death.

For Control Group :

The tables provides a brief interpretation of the logistic regression results for each Outcome (The table include the predictors with the corresponding estimates of parameters in fitted model and the odds ratios) for Control group (i.e. women who did not received any treatment during preconception age).

i) ABORTION :

Result of logistic model when Abortion is an outcome variable :

Predictor	**Estimate**		**Odds Ratio**
Age < 20	(Intercept)	-2.71	0.74
	Age < 20 (Yes)	-0.30	
BMI > 25	(Intercept)	-4.77	**2.23**
	BMI > 25 (Yes)	0.80	
BMI < 18.5	(Intercept)	-4.26	**1.85**
	BMI < 18.5 (Yes)	0.61	
Caste	(Intercept)	-2.09	0.55
	Caste (ST)	-0.61	
Family Type	(Intercept)	-1.79	0.47
	Family Type (Nuclear)	-0.78	
Education	(Intercept)	-2.74	0.74
	Education (<10^{th})	-0.31	
Occupation	(Intercept)	-2.95	0.82
	Occupation (Non-Earning)	-0.21	
Consanguinity Marriage	(Intercept)	-3.71	**1.34**
	Consanguinity Marriage (yes)	0.29	
Protein Intake < 50%	(Intercept)	-3.18	0.96
	Protein Intake < 50% (Yes)	-0.04	
Calorie Intake < 50%	(Intercept)	-3.41	**1.08**
	Calorie Intake < 50% (Yes)	0.07	

INTERPRETATION : Woman with BMI > 25 have 2.23 times i.e approximately double risk of getting abortion than women having BMI ≤ 25. Lly women with BMI (<18.5), Consanguinity marriage (Yes) & calorie Intake (<50%) have more risk of abortion than women with BMI (≥ 18.5), Consanguinity marriage (No) & calorie Intake (>50%)

ii) STILL BIRTH :

Result of logistic model when Still Birth is an outcome variable :

Predictor	Estimate		Odds Ratio
Caste	(Intercept)	-3.38	0.35
	Caste (ST)	-1.06	
Family Type	(Intercept)	-2.77	0.30
	Family Type (Nuclear)	-1.23	
Education	(Intercept)	-3.30	0.32
	Education (<10th)	-1.17	
Occupation	(Intercept)	-5.52	1.37
	Occupation (Non-Earning)	0.31	
Adverse Outcome Previous	(Intercept)	1.90	0.03
	Adverse Outcome Previous (Yes)	-3.85	

INTERPRETATION : Here, women with Occupation (non-Earning) have 1.37 times more risk of Outcome Still birth.

iii) LOW BIRTH WEIGHT

Result of logistic model when Low Birth Weight is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-1.72	0.76
	Age < 20 (Yes)	-0.28	
BMI > 25	(Intercept)	-1.24	0.62
	BMI > 25 (Yes)	-0.48	
BMI < 18.5	(Intercept)	-2.05	0.94
	BMI < 18.5 (Yes)	-0.07	
Caste	(Intercept)	-3.30	1.90
	Caste (ST)	0.64	
Family Type	(Intercept)	-2.15	0.95

	Family Type (Nuclear)	-0.06	
Education	(Intercept)	-2.88	1.54
	Education (<10th)	0.43	
Occupation	(Intercept)	-2.08	0.96
	Occupation (Non-Earning)	-0.05	
Consanguinity Marriage	(Intercept)	-1.83	0.81
	Consanguinity Marriage (Yes)	-0.22	
Adverse Outcome Previous	(Intercept)	-3.62	1.86
	Adverse Outcome Previous (Yes)	0.62	
Protein Intake < 50 %	(Intercept)	-3.78	2.20
	Protein Intake < 50 % (Yes)	0.79	

INTERPRETATION : Caste (ST), Education(<10th), Adverse utcome Previous(Yes) are the risk factors for Low Birth Weight. (i.e presence of these predictors may carry more risk of event Low Birth Weight).

iv) PRE-TERM

Result of logistic model when Pre term is an outcome variable :

Predictor	Estimate		Odds Ratio
Age < 20	(Intercept)	-1.95	1.07
	Age < 20 (Yes)	0.06	
BMI > 25	(Intercept)	-3.24	2.09
	BMI > 25 (Yes)	0.74	
BMI <18.5	(Intercept)	-1.68	0.93
	BMI <18.5 (Yes)	-0.08	
Caste	(Intercept)	-2.22	1.28
	Caste (ST)	0.25	
Family Type	(Intercept)	-2.03	1.12
	Family Type (Nuclear)	0.11	
Education	(Intercept)	-1.52	0.85
	Education (<10th)	-0.17	
Occupation	(Intercept)	-1.34	0.75
	Occupation (Non-Earning)	-0.30	
Consanguinity Marriage	(Intercept)	-1.81	1.01
	Consanguinity Marriage (Yes)	0.00	
MultiParity	(Intercept)	-1.58	0.91
	MultiParity (Yes)	-0.10	

Adverse Outcome Previous	(Intercept)	-1.63	0.86
	Adverse Outcome Previous(Yes)	-0.16	
Tobacco/Smoking/ Alcohol	(Intercept)	-1.97	1.08
	Tobacco/Smoking/Alcohol (Yes)	0.07	
Protein Intake < 50%	(Intercept)	-2.16	1.19
	Protein Intake < 50 % (Yes)	0.17	

INTERPRETATION : Age < 20 (Yes), Caste (ST), Family Type (Nuclear), Consanguinity Marriage (Yes), Tobacco/Smoking/Alcohol (Yes), Protein Intake < 50 % (Yes) are the risk factors for Pre-Term baby. (i.e presence of these predictors may carry more risk of event Pre Term baby.

v) CONGENIAL DEFECTS

Result of logistic model when Congenial Defects is an outcome variable :

Predictor	**Estimate**		**Odds Ratio**
Age < 20	(Intercept)	-5.20	1.00
	Age < 20 (Yes)	-0.01	
BMI < 18.5	(Intercept)	-4.83	0.83
	BMI < 18.5 (Yes)	-0.19	
Caste	(Intercept)	-4.08	0.51
	Caste (ST)	-0.69	
Education	(Intercept)	-5.03	0.93
	Education (<10th)	-0.08	
Consanguinity Marriage	(Intercept)	-4.01	0.53
	Consanguinity marriage (Yes)	-0.65	

INTERPRETATION : By the above table,for women with age < 20 has no effect on outcome Congenial defects, while remaining predictors also carries less risk of Congenial defects.

vi) NEONATAL DEATH

Result of logistic model when Neonatal Death is an outcome variable :

Predictor	**Estimate**		**Odds Ratio**
Age < 20	(Intercept)	-4.91	0.75

	Age < 20 (Yes)	-0.30	
BMI < 18.5	(Intercept)	-6.23	1.67
	BMI < 18.5 (Yes)	0.51	
Caste	(Intercept)	-5.88	1.52
	Caste (ST)	0.42	
Education	(Intercept)	-6.42	1.85
	Education (<10th)	0.62	
Occupation	(Intercept)	-6.57	2.05
	Occupation (Non-Earning)	0.72	
Consanguinity marriage	(Intercept)	-3.60	0.35
	Consanguinity marriage (Yes)	-1.06	

INTERPRETATION : BMI (<18.5), Caste (ST),Caste (ST),Education (<10th), Occupation(non-Earning) are the risk factors for Neonatal Death (i.e presence of these predictors may carry more risk of event Neonatal Death).

CONCLUSIONS

- Performing EDA for different continuous and Categorical factors gave an idea about the data. For different continuous variables the distributions are more or less similar in Intervention & Control group i.e the continuous variables do not follow any pattern if they are from intervention group or from control group.
- Performing different statistical test like chisquare test of association and tools like Risk Ratios & odds ratios we got that the factors Parity (having child >2), Family Type (Nuclear), Occupation (Non-Earning), Consangeneous marriage (Yes), Calorie Intake (<50%), Pre-adverse Outcome (Yes), Protein Inatake (<50%), Age (<20), Education (<10^{th}), Consanguineous marriage (Yes), Age (>35), Caste (ST), BMI(<18.5), BMI (>25), Tobacco/Smoking/Alcohol consumer (Yes) can carry the risk of bad outcomes (Abortion, Still-Birth, Low Birth Wt, Pre-Term, Congenial Anomly, Neonatal Death) in the women so thease factors can considered as the risk factors for the women & their forthcoming progeny.
- Also the Risk Ratio and two sample proportion Z test indicates that being in control group can carry the risk of bad events like Abortion, Low-Birth Wt & preterm baby for the women.
- In overall summary, from the data we can conclude that providing some treatment to women during preconception age lessers the risk of bad outcomes for the women and their forthcoming progeny.

LIMITATIONS

- There are many missing values present in the data, some women are selected in study but some what reason they do not share their information with data collector.(i.e. ASHA Workers) (The reasons might include many fcators like if a woman was not pregnant, she has not provided the responses for outcome like Abortion, Still-Birth, Pre Term baby, Congenial defects, Neonatal deaths.
- Due to the presence of the missing values in the data we were unable to perform some machine learning algorithms.

REFERENCES

- Research Paper : Prevalence of Preconception Risk Factors Among Women From Tribal and Non-Tribal Blocks in Nashik District, India : A Cross-Sectional Study https:// www.researchsquare.com/article/rs-298120/v1.pdf
- 3 EDn logistic-regression-a-self-learning-text by David G. Kleinbaum.MitchelKlein
- An Introduction to statistical Learning with Applications in R by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani Springer Texts in Statistics.
- Relative Risk Calculator https://www.omnicalculator.com/statistics/relative-risk
- Softwares Used:
    - R Software
    - MS-Excel
    - Minitab
- https://www.goggle.com

APPENDIX

i)Enrollment Form :

A study of implementation of pre-conception care for preventing adverse pregnancy outcome in Nashik district, Maharashtra.

1	Village name:	Sub Centre PHC Block
2	Name: First name	Husband's name Surname
3	Age	……years
4	Height (in meters)	…….mt
5	Weight (in kg.)	…….kg
6	Religion	1.Hindu, 2.Muslim, 3.Bouddha, 4. Christian, 5.Other
7	Caste	1.Scheduled caste, 2.Scheduled tribe, 3.O.B.C. 4.Open
8	Type of family	1.Nuclear, 2.Joint, 3.Other
9	Contact number	| | | | | | | | | | | |
10	Your education	1.Post-graduate/professional degree, 2.Graduate degree, 3.HSC (12^{th})/ ITI, 4.SSC (10^{th}), 5.Seventh pass, 6. less than 7^{th}, 8.Illiterate
11	Your occupation	1.Professionals, 2.Semi professional, 3.Clerical, 4.Shop owner, 5.Farmer, 6.Skilled worker, 7.Semiskilled worker, 8.Unskilled worker, 9.Unemployed/ Household
12	Education of household head	1.Post-graduate/professional degree, 2.Graduate degree, 3.HSC (12^{th})/ ITI, 4.SSC (10^{th}), 5.Seventh pass, 6. less than 7^{th}, 8.Illiterate
13	Occupation of household head	1.Professionals, 2.Semi professional, 3.Clerical, 4.Shop owner, 5.Farmer, 6.Skilled worker, 7.Semiskilled worker, 8.Unskilled worker, 9.Unemployed
14	Total number of family members	
15	Total family income per month	Rs.
16	Age at marriage	….years
17	Marriage within relation	1.Yes, 2.No (If yes) prior relation to husband:
18	How many times were you pregnant?	
If primi, go to question 24		
19	How many times did you delivered?	
20	Date of last event	Date: DD/MM/YYYY

21	Outcome of last event	1.Live-birth: birthweight:…….kg, gestational age in months:……., 2.Still-birth, 3.Abortion 4.Early neonatal death, 5.Maternal death
In case of abortion, please do not ask question 23		
22	In which month was the foetus aborted/ child delivered?	
23	Does child have or had congenital anomaly?	
24A	Do you consume tobacco (more than thrice a week)	1.Yes, 2. No
If answer is no go to question 25		
24B	If yes, specify	1.Smoking -cigarette/Bidi 2.Gutkha 3.Snuff 4.Mishri
24C	Since when are you consuming tobacco?	…..yrs…..months
25A	Do you consume alcohol?	1.Yes, 2. No
If answer is no go to question 26		
25B	Type of alcohol	1.Country 2.Tadi/ Madi 3. Branded
25C	Since when are you consuming alcohol?	…..yrs
26	Generally when do you have meal?	1.With husband, 2.With family members, 3.After men, 4. Last
27	Average food quantity	1.Abundant, 2.Enough, 3.Lesser, 4.Remaining
28	Food groups	1.Vegetarian, 2.Non-vegetarian 3.Sometimes non-vegetarian

ii) R Codes :

```
rm(list=ls());library(readxl)

data<- read_excel("C:/Users/shree/Downloads/MSC PROJECT WORK/Data_for_Project.xlsx")

#View(data);

dim(data); length(data$Ageyrsmnth)

#Summary Statistics

library(vtable)  #To calculate summary table

st((data.frame(data$Ageyrsmnth,data$Heightcm,data$Weightkg,data$BMI, data$FamilyIncome,data$Ageatmarriage,data$Babyweight,data$GestationalageofMother)))

#Data for Intervention Group

I2 = data[data$Group==1,];dim(I2)
```

```
#For chisquare table :
length(which(I2$Age_Less_Than_20==1))
age=ifelse(I2$Ageyrsmnth<=35 & I2$Ageyrsmnth>=20,1,0)
length(which(age==TRUE))
length(which(I2$Age_More_Than_35==1)); length(which(I2$Caste_cat==1))
length(which(I2$Caste_cat==2));length(which(I2$Multi_Parity==1))
.
.
##To find Association (Chisquare test)
AG=chisq.test(matrix(c(164,204,821,1020,9,4),byrow=T,ncol=2))
BG=chisq.test(matrix(c(357,408,502,657,65,70),byrow=T,ncol=2))
Cast=chisq.test(matrix(c(343,316,538,608),byrow=T,ncol=2))
#To find ODDs ratios
#For Abortion
A.1=table(I2$Multi_Parity,I2$Abortion)
names(dimnames(A.1))=c("parity","abortion");A.1
A.2=table(I2$Family_Cat,I2$Abortion)
names(dimnames(A.2))=c("famType","abortion");A.2
.
.
#Data for Control Group
C2 = data[data$Group==2,];dim(C2)
#For chisquare table :
length(which(C2$Age_Less_Than_20==1))
age_=ifelse(C2$Ageyrsmnth<=35 & C2$Ageyrsmnth>=20,1,0)
length(which(age_==TRUE))
length(which(C2$Age_More_Than_35==1)); length(which(C2$Caste_cat==1))
```

```
length(which(C2$Caste_cat==2)); length(which(C2$Multi_Parity==1))

length(which(C2$Multi_Parity==2))

.

.

#To find ODDs ratios

#For Abortion

A..1=table(C2$Multi_Parity,C2$Abortion)

names(dimnames(A..1))=c("parity","abortion");A..1

A..2=table(C2$Family_Cat,C2$Abortion)

names(dimnames(A..2))=c("famType","abortion");A..2

.

.

# For Two sample Proportion Table

#Intervention

#Eg. Abortion(N)=length(which(I2$Abortion==1))+length(which(I2$Abortion==0))

#(i.e Total no of women who have given response for abortion whether it's yes/no)

#Abortion(n)=length(which(I2$Abortion==1)) (out of total responses no of yes responses)

length(which(I2$Abortion==1));length(which(I2$Abortion==0))

.

.

#Control

length(which(C2$Abortion==1)); length(which(C2$Abortion==0))

.

.

###Z Proportion test

prop.test(x=c(42,36),n=c(482,975),p=NULL,alternative="less",correct=TRUE)

prop.test(x=c(6,6),n=c(482,975),p=NULL,alternative="greater",correct=TRUE)
```

.

.

#Code for Logistic Regression & Forest plots

```
rm(list=ls())

#Neonatal_Death

# Load necessary libraries
library(readxl)
library(dplyr)
library(broom)
library(forestplot)

# Load data
data1 <- read_excel("D:/Bhavishya/d1.xlsx", sheet = "Sheet3")

# Define a function to extract summary statistics and odds ratios
extract_summary <- function(model) {
  tidy_model <- tidy(model)
  tidy_model <- tidy_model %>%
    mutate(
      Odds_Ratio = exp(estimate),
      CI_Lower = exp(estimate - 1.96 * std.error),
      CI_Upper = exp(estimate + 1.96 * std.error),
      Significance = case_when(
        p.value < 0.001 ~ "***",
        p.value < 0.01 ~ "**",
        p.value < 0.05 ~ "*",
        p.value < 0.1 ~ ".",
        TRUE ~ " "
      )
    )
  tidy_model
}

# Perform logistic regressions and extract summaries
models <- list(
  Age_Less_Than_20 = glm(Neonatal_Death  ~ Age_Less_Than_20, data = data1,
family = "binomial"),
  Age_More_Than_35 = glm(Neonatal_Death ~ Age_More_Than_35, data = data1,
family = "binomial"),
  BMI_More_Than_25 = glm(Neonatal_Death  ~ BMI_More_Than_25, data = data1,
family = "binomial"),
  BMI_Less_Than_185 = glm(Neonatal_Death  ~ BMI_Less_Than_185, data = data
1, family = "binomial"),
  Caste_cat = glm(Neonatal_Death ~ Caste_cat, data = data1, family = "bino
mial"),
  Family_Cat = glm(Neonatal_Death  ~ Family_Cat, data = data1, family = "b
```

```
inomial"),
  l7 = glm(Neonatal_Death ~ Education_Cat_New, data = data1, family = "bin
omial"),
  l8 = glm(Neonatal_Death  ~ Occupation_Cat, data = data1, family = "binom
ial"),
  l9 = glm(Neonatal_Death ~ Consanguinity, data = data1, family = "binomia
l"),
  l10 = glm(Neonatal_Death ~ Multi_Parity, data = data1, family = "binomia
l"),
  l11 = glm(Neonatal_Death ~ Adverse_Outcome_Privious, data = data1, famil
y = "binomial"),
  l12 = glm(Neonatal_Death ~ Tobacco_or_Smoking, data = data1, family = "b
inomial"),
  l13 = glm(Neonatal_Death ~ Protein_Intake_Less_Than_50_Perc, data = data
1, family = "binomial"),
  l14 = glm(Neonatal_Death  ~ Caloriy_Intake_Less_Than_50_Perc, data = dat
a1, family = "binomial")
)

# Create a list to store summary data frames
summary_list <- lapply(models, extract_summary)

# Combine all summaries into a single data frame with model names
summary_df <- do.call(rbind, lapply(names(summary_list), function(model_na
me) {
  cbind(Model = model_name, summary_list[[model_name]])
}))

# Create a data frame for forest plot
forest_df <- summary_df %>%
  filter(term != "(Intercept)") %>%
  select(Model, term, Odds_Ratio, CI_Lower, CI_Upper, p.value, Significanc
e) %>%
  mutate(Model = factor(Model, levels = names(models)))

# Filter out rows with non-finite CI_Upper values
forest_df <- forest_df %>%
  filter(is.finite(CI_Upper))

# Print the combined summary data frame
print(summary_df)

# Create the forest plot
forestplot::forestplot(
  labeltext = cbind( forest_df$term, sprintf("%.2f", forest_df$Odds_Ratio)
, sprintf("%.2f", forest_df$CI_Lower), sprintf("%.2f", forest_df$CI_Upper)
),
  mean = forest_df$Odds_Ratio,
  lower = forest_df$CI_Lower,
  upper = forest_df$CI_Upper,
```

```
  title = "Forest Plot of Odds Ratios",
  xlab = "Odds Ratio",
  col = forestplot::fpColors(box = "darkblue", lines = "red", zero = "gray
50"),
  zero = 1,
  xticks = seq(0, ceiling(max(forest_df$CI_Upper, na.rm = TRUE)), by = 1),
graph.pos = 3,
  align = "c",
  new_page = TRUE
)
```