A PROJECT REPORT ON,

STUDY OF "MATERNITY HEALTH"

Submitted to,

DEPARTMENT OF STATISTICS,

SHIVAJI UNIVERSITY, KOLHAPUR.

For the partial fulfillment of degree

M.Sc.-II in STATISTICS / APPLIED STATISTICS AND INFORMATICS 2022-2023

By,

MR. GODASE SANGRAMSINH DATTATRAY

MISS. POL SWATI UMESH

MR. KUCHEKAR AJIMKYA SHRIDHAR

MISS. LOHAR SHRADDHA BALKRISHNA

Under the Guidance of,

Mr. S. K. Ganjave Sir.

CERTIFICATE

This is to certify that this project report on,

"STATISTICAL ANALYSIS OF MATERNITY HEALTH"

Is submitted by,

Sr. No	Name of Students	PRN
1	MR. GODASE SANGRAMSINH DATTATRAY	2021000521
2	MISS. POL SWATI UMESH	2021071305
3	MR. KUCHEKAR AJINKYA SHRIDHAR	2021000264
4	MISS. LOHAR SHRADDHA BALKRISHNA	2021000446

As a partial fulfilment for the award of M.Sc., in statistics, under my supervision and guidance as per rules and regulations of the Shivaji University, Kolhapur. During the academic year 2022-2023 and submitted the same. This work represents the bonafide work of these students. To the best of my knowledge, the matter represented here in the project has not been submitted earlier.

Mr. S. K. Ganjave.

Project Guide,
Department of Statistics,
Shivaji University, Kolhapur.

Dr. Prof. S. B. Mahadik.

Head,
Department of Statistics,
Shivaji University, Kolhapur.

Date:

Place: Kolhapur.

INDEX

ACKNOWLEDGEMENT

We take great pleasure in submitting this project report on "MATERNITY HEALTH" as a practical fulfilment of M.Sc.-II.

We extend our heartfelt gratitude to Prof. Dr. S. B. Mahadik (Head of Department) for granting us permission to undertake this project.

We would like to express our deep appreciation to our project guide, Mr. S. K. Ganjave, who provided guidance from the inception of the project and offered invaluable suggestions for its systematic and presentable development. We would also like to extend special thanks to Prof. S. V. Rajguru (Assistant Professor, Department of Statistics) for providing valuable insights for our analysis.

We would like to express our highest regards to the Department of Statistics and acknowledge the contributions of our esteemed teachers, Prof. Dr. Mr. S. D. Pawar, Shri. S. K. Ganjave, Shri. S. V. Rajguru, and Mr. S. M. Patil. This project is dedicated to all the individuals whom we encountered, sought guidance from, interviewed, and learned from. We take this opportunity to sincerely thank each and every one of them as we submit this report.

PREFACE

Acquiring practical knowledge is of utmost importance in a student's academic journey. It allows students to understand real-life situations and challenges. While theoretical knowledge is crucial, practical knowledge holds equal significance. Projects serve as a means to impart practical knowledge to students.

This particular project serves as an opportunity and platform to comprehend the present situation and dynamics of the environment.

In the preparation of this project, I am profoundly grateful to all the faculty members, staff members of the respective organization, and every individual who assisted me in the creation of this project report.

INTRODUCTION

Recent research conducted by the World Health Organization (WHO) indicates that the potential to save lives through Caesarean Section (C-Section) procedures is expected to persist over the next decade. Globally, the rate of C-Section is projected to surpass the recommended population-based threshold of 10 to 15 percent.

However, studies conducted in several developed countries suggest that a significant proportion of these procedures may be unnecessary. In this analysis, we will delve into the factors influencing the escalating rates of C-Section and explore the underlying rationales behind this phenomenon.

Moreover, the increasing prevalence of C-Sections not only impacts individual health outcomes but also carries significant societal and economic implications. The overuse of C-Section interventions not only strains healthcare systems but also raises concerns regarding unwarranted medical interventions and associated risks. Understanding the factors driving this trend is essential for formulating strategies that promote evidence-based practices and ensure optimal health outcomes for both mothers and neonates. By addressing the root causes behind the rising rates of C-Sections, we can strive towards a more efficient and equitable approach to childbirth.

METHODOLOGY

In order to collect the data required for this study, several methods were employed, primarily focusing on obtaining information from Maternity Hospitals. The following methods were utilized:

Primary Data Collection:

The primary data for this study was obtained from the files of admitted delivery patients. These files contain comprehensive information about the patients' medical history, delivery details, and relevant variables of interest.

Sample Size:

The total sample size for this study consisted of 146 participants. Through this sample, we aimed to investigate the factors influencing the mode of delivery, specifically focusing on normal deliveries and caesarean deliveries.

Maternity Hospitals:

Data collection took place in Maternity Hospitals, where the relevant information was sourced. These hospitals provided a suitable setting for accessing the required data, given their specialization in maternity care.

By utilizing primary data collection methods and working with a sample size of 146, we sought to analyse the impact of various factors on the choice between normal and caesarean deliveries. The data collection process was carried out in collaboration with Maternity Hospitals, ensuring access to comprehensive and pertinent information for our study.

OBJECTIVE

1. Prediction of Delivery Type using logistic regression.
2. Which variable has a greater impact on the occurrence of Caesarean delivery?
3. Survival analysis of gestation period.
4. What are the reasons for Caesarean delivery?

Using Software:

- Excel software
- R Programming
- Python
- SAS
- Minitab software

VARIABLES

Marriage_Age: The age at which the individual got married.

Current_Age: The current age of the individual.

Weight_BD: The weight of the individual before delivery.

Weight_AD: The weight of the individual after delivery.

Height: The height of the individual.

Systolic_BP: The systolic blood pressure of the individual.

Diastolic_BP: The diastolic blood pressure of the individual.

HB: The hemoglobin levels of the individual.

Gestation: The duration of pregnancy in weeks.

AC: Abdominal circumference of the individual.

Preference_delivery_type: The preferred mode of delivery indicated by the individual.

Previous_CD: History of previous Caesarean delivery.

Medical_Risk_Factor: Presence of any medical risk factors during pregnancy.

Breech: The position of the baby (breech or not breech).

Parity: The number of previous pregnancies carried to viable gestational age.

Exercise: Engagement in physical exercise during pregnancy.

Childbirth_Education: Participation in childbirth education or preparation classes.

Medical_Care: Level of medical care received during pregnancy.

Diet: Dietary habits during pregnancy.

Sleeping_hours: Increasing and Decreasing hours of sleep per month during pregnancy.

Qualification_P: Qualification of the individual.

Occupation_p: Occupation of the individual.

Weight_NB: The weight of the newborn baby.

Height_NB: The height of the newborn baby.

HC: Head circumference of the newborn baby.

Gender_NB: Gender of the newborn baby.

Twins: Presence of twins in the pregnancy.

Handicapped: Whether the baby is handicapped or not.

Residence: Place of residence of the individuals. (Urban or Rural)

Family: Information about the individual's family. (Joint or Nuclear)

Occupation_H: Occupation of the individual's spouse.

Qualification_H: Qualification of the individual's spouse.

Delivery_type: The mode of delivery (normal or Caesarean section).

TOOLS USED

Test of significance for Proportion:

The z-test of significance is a statistical test used to determine the significance of a proportion compared to a hypothesized value.

Correlation:

Correlation is a statistical measure that quantifies the relationship between two variables. It assesses the strength and direction of the linear association between variables, indicating how changes in one variable correspond to changes in another.

Shapiro-Wilk test:

The Shapiro-Wilk test is a statistical test used to assess the normality of a dataset. It examines whether the data follows a normal distribution or deviates significantly from it. The test is based on the null hypothesis that the data is normally distributed.

Mann-Whitney U test:

The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, is a nonparametric statistical test used to compare the distribution of two independent groups. It is employed when the assumptions of parametric tests, such as the t-test, are not met, particularly when the data is not normally distributed.

chi-square test for independence:

The chi-square test for independence is a statistical test used to determine if there is a relationship between two categorical variables. It examines whether the distribution of one variable is independent of the distribution of another variable.

Product-limit survival estimate:

The product-limit survival estimate provides an empirical estimation of the survival function, which represents the probability that an individual or a group will survive beyond a certain time point. It is a nonparametric approach that does not assume any specific distribution of survival times.

Logistic regression:

Logistic regression is a statistical modeling technique used to analyze the relationship between a binary dependent variable and one or more independent variables. It is widely employed when the outcome of interest is binary or categorical, such as presence/absence, success/failure, or yes/no.

Word Cloud:

A word cloud is a visualization technique that represents the frequency or importance of words in a given text. It is a visual representation of textual data where the size and prominence of words are based on their frequency or significance within the text.

DATA ANALYSIS

Test of significance for single Proportion:

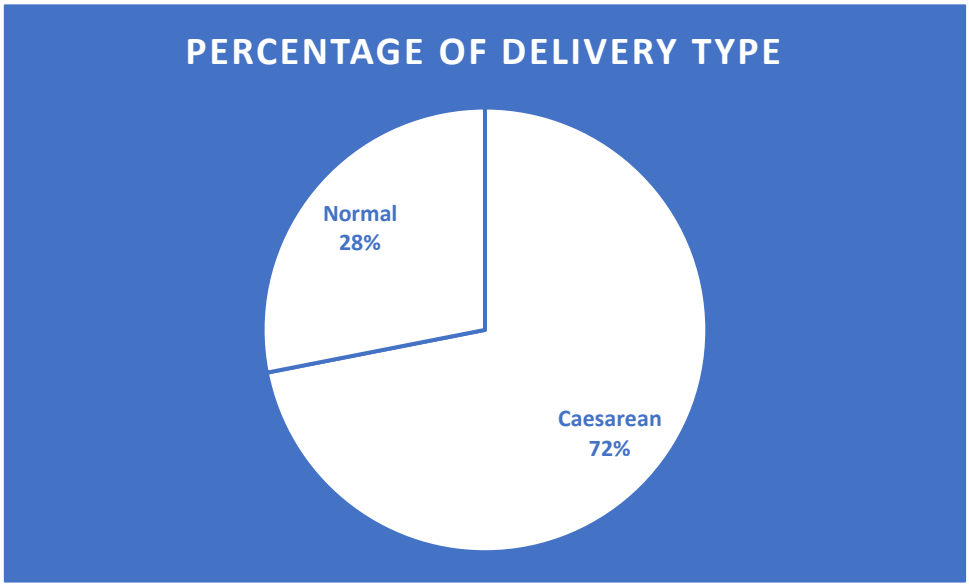
Test for the proportion of Normal and Caesarean delivery.

Hypothesis:

Null Hypothesis (H_0): Proportion of Caesarean delivery is greater than Normal delivery.

Alternative Hypothesis (H_1): Proportion of Caesarean delivery is less than Normal delivery.

Row Labels	Count of Delivery type	Proportion
Caesarean	105	0.71917808
Normal	41	0.28082192
Grand Total	146	



With large sample size, the Central Limit Theorem comes in to play.

According to the Central Limit Theorem, the sampling distribution of mean tends to approximate a Normal Distribution regardless of the underlying distribution of the data for large sample size.

One Sample Z Test for Proportions:

Left Tail Test:

A left-tailed hypothesis is one in which the proportion of the first sample is less than that of the second sample. This can be expressed as:

H_0: $p_1 >= p_2$

H_1: $p_1 < p_2$

Now,

Z_{cal}	P-Value
-5.8930	0.9999

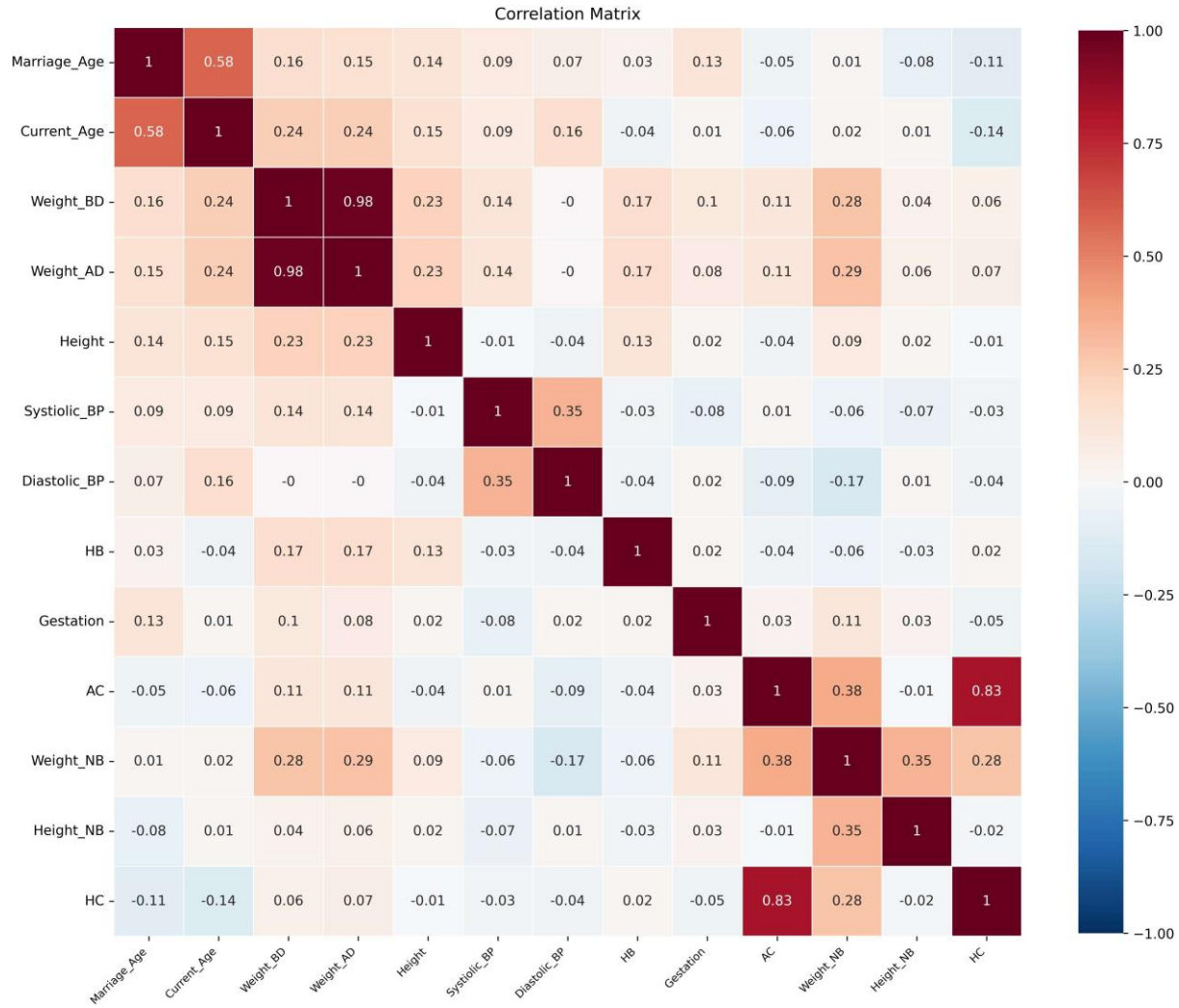
P-value for One Sample Z left tailed test is 0.999>0.05

We fail to reject H_0

Conclusion:

1) The percentage of caesarean delivery is 72% and normal delivery is 28%
2) Proportion of Caesarean delivery is greater than Normal delivery.

Correlation between variables:



Conclusion:

1. There is a strong correlation observed between the head circumference of newborn babies and the abdominal circumference of patients.
2. A strong relationship has been found between pre-delivery weight and post-delivery weight.

Normality test for continuous variable:

Shapiro–Wilk test:

	statistics	p_value
Marriage_Age	0.94	**0.000**
Current_Age	0.95	**0.000**
Weight_BD	0.97	**0.003**
Weight_AD	0.97	**0.006**
Height	0.94	**0.000**
SystioIic_BP	0.78	**0.000**
DiastoIic_BP	0.83	**0.000**
HB	0.99	**0.134**
Gestation	0.81	**0.000**
AC	0.94	**0.000**
Weight_NB	0.97	**0.003**
Height_NB	0.85	**0.000**
HC	**0.91**	**0.000**

From above table all variable except HB rejects null hypothesis i.e., only HB follows normal distribution.

Conclusion:

p-value for all variable is <0.05 except HB thus we use nonparametric test to compare means according to variable for CD and Normal delivery

The Mann-Whitney U test:

To test that there is effect of continuous variable on delivery type using Mann-Whitney U test for comparing mean of each variable with respect to delivery type.

Variable	stat	P_val
Marriage_Age	2190.5	0.869530
Current_Age	2262.0	0.633662
Weight_BD	2489.0	0.142955
Weight_AD	2479.5	0.154941
Height	2375.5	0.330451
SystioIic_BP	1724.5	0.054409
DiastoIic_BP	2278.0	0.569518
HB	2386.5	0.308738
Gestation	1848.0	0.173693
AC	1809.0	0.135099
Weight_NB	2387.5	0.304907
Height_NB	2298.0	0.524606
HC	2237.0	0.714064

from The Mann-Whitney U test (for two independent samples) used in above table yhe variable Weight_BD, Weight_AD,SystioIic_BP,Gestation,AC, are dependent for delivery type at 20% level of significance.

Two Sample Z Test for Proportions:

According to the Central Limit Theorem, the sampling distribution of mean tends to approximate a Normal Distribution regardless of the underlying distribution of the data for large sample size.

Using Two Sample Z Test for Proportions for categorical data we have following results

Variable	Z_cal	P_val
MedicaI_Risk_Factor	-0.199530	0.841848
Breech	-1.948415	0.051365
Parity	1.677764	0.093393
Exercise	1.981570	0.047527
ChiIdbirth_Education	-0.315871	0.752101
MedicaI_Care	0.139088	0.889381
Diet	0.718006	0.472754
Gender_NB	-1.554134	0.120152
Twins	-0.889831	0.373557
Handicapped	-0.589254	0.555691
Residence	-0.567703	0.570237
Family	0.172338	0.863172

p- value for variable Breech, parity and Exercise is less than significance lavel 0.1 thus we reject null hypothesis for this variable and conclude that Breech pregnancy, parity and Exercise have impact on delivery type.

Conclusion:

Breech pregnancy, parity and Exercise have impact on delivery type.

chi-square test for independency:

To check dependency of categorical data having more than two categories

Variable	P_val
Preference_delivery_type	0.664051
Sleeping_hours	0.995704
QuaIification_P	0.999954
Occupation_p	0.999892
Occupation_H	0.999871
QuaIification_H	0.999995

p-value for all the variable are greater than significance value 0.05 thus we fail to reject null hypothesis and conclude that delivery type is independent of this variable.

Conclusion:

Delivery type is independent of occupation and qualification of patient and her husband. Also, it is independent of preference for delivery type and sleeping hours increases or decreases during pregnancy.

Survival Analysis for Gestation Period

Non-parametric Procedure

Product-limit survival estimates:

Product-limit survival estimates, also known as Kaplan-Meier estimates, are a statistical method used to estimate the survival probability over time in survival analysis.

Survival estimates for caesarean delivery:

Product-Limit Survival Estimates

The LIFETEST Procedure

Stratum 1: Type = 0

Gestation	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	105
32.0000	0.8857	0.1143	0.0310	12	93
34.0000	0.8571	0.1429	0.0341	15	90
35.0000	0.8476	0.1524	0.0351	16	89
36.0000	0.8000	0.2000	0.0390	21	84
37.0000	0.7048	0.2952	0.0445	31	74
38.0000	0.5619	0.4381	0.0484	46	59
39.0000	0.3048	0.6952	0.0449	73	32
40.0000	0.1333	0.8667	0.0332	91	14
41.0000	0.00952	0.9905	0.00948	104	1
42.0000	0	1.0000	.	105	0

Quartile Estimates				
		95% Confidence Interval		
Percent	Point Estimate	Transform	[Lower	Upper)
75	40.0000	LOGLOG	39.0000	40.0000
50	39.0000	LOGLOG	38.0000	39.0000
25	37.0000	LOGLOG	36.0000	38.0000

Mean	Standard Error
37.9905	0.2622

Survival estimates for Normal delivery:

Stratum 2: Type = 1

Product-Limit Survival Estimates

Gestation	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	41
36.0000	0.9756	0.0244	0.0241	1	40
37.0000	0.9512	0.0488	0.0336	2	39
38.0000	0.8049	0.1951	0.0619	8	33
39.0000	0.2195	0.7805	0.0646	32	9
40.0000	0.0244	0.9756	0.0241	40	1
41.0000	0	1.0000	.	41	0

Quartile Estimates				
		95% Confidence Interval		
Percent	Point Estimate	Transform	[Lower	Upper)
75	39.0000	LOGLOG	39.0000	40.0000
50	39.0000	LOGLOG	.	.
25	39.0000	LOGLOG	38.0000	39.0000

Mean	Standard Error
38.9756	0.1374

Test of Equality for survival experiences:

Null Hypothesis (H_0): There is no difference in the survival experiences between the groups.

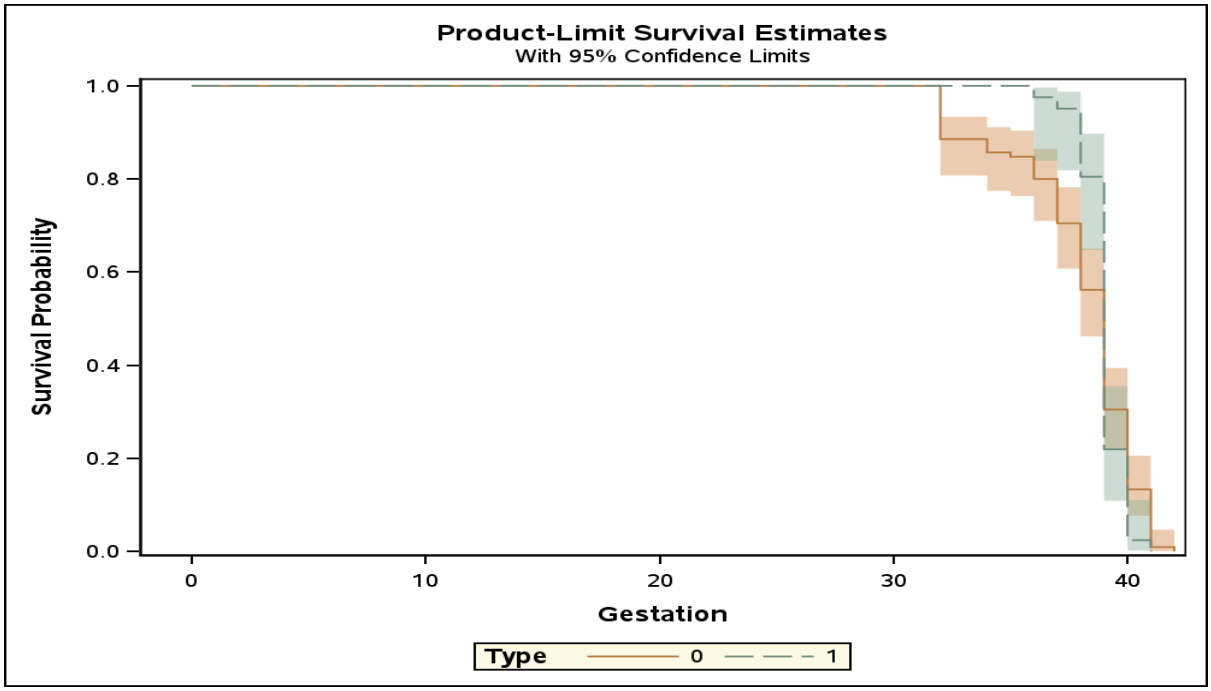
Alternative Hypothesis (H_1): There is a difference in the survival experiences between the groups.

Rank Statistics		
Type	**Log-Rank**	**Wilcoxon**
0	1.2234	609.00
1	-1.2234	-609.00

Covariance Matrix for the Log-Rank Statistics		
Type	**0**	**1**
0	17.4529	-17.4529
1	-17.4529	17.4529

Covariance Matrix for the Wilcoxon Statistics		
Type	**0**	**1**
0	215042	-215042
1	-215042	215042

Test of Equality over Strata			
Test	**Chi-Square**	**DF**	**Pr > Chi-Square**
Log-Rank	0.0858	1	0.7696
Wilcoxon	1.7247	1	0.1891
-2Log(LR)	0.0194	1	0.8892



Delivery_type	count	mean	std	min	25%	50%	75%	max
0	105.0	37.990476	2.687203	32.0	37.0	39.0	40.0	42.0
1	41.0	38.975610	0.879994	36.0	39.0	39.0	39.0	41.0

The survival analysis graph demonstrates that there is a difference in the survival time for gestation periods during the initial period. However, as time progresses, there is no significant difference in survival time. The average gestation period for normal delivery is approximately 39 weeks, while for Caesarean delivery, it is 38 weeks. The standard deviation for Caesarean delivery is 2.68, whereas for normal delivery, it is 0.87, indicating a much smaller dispersion compared to Caesarean delivery. The confidence interval for the median gestation period in Caesarean delivery is [38.0000, 39.0000], while for normal delivery, the median is exact. Thus, if the gestation period is 39 weeks, there is a maximum probability of normal delivery.

The p-values obtained from both the Log-rank test and Wilcoxon test are greater than 0.05. Therefore, we fail to reject the null hypothesis and conclude that there is no significant difference in the survival experiences between the groups being compared.

Conclusion:

There is no significant difference observed in the survival experiences between the groups of individuals undergoing Caesarean delivery and those undergoing normal delivery with respect to the gestation period.

PREDICTION OF DELIVERY TYPE USING LOGISTIC REGRESSION

Method:

Logit link is used to build logistic regression model. There are 146 observations out of 30.1% are used for testing and 69.99% are used for training.

Link function	Logit
Rows used	146
Test set fraction	30.1%

Stepwise Selection of Terms:

Stepwise selection method is used to include variable in logistic regression model.

Response Information:

0: Caesarean delivery

1: Normal delivery

Delivery type	Training Count	Test Count
1	28	13
0	74	31
Total	102	44

Regression Equation:

$$\Pi = \exp(Y')/(1 + \exp(Y'))$$

$$Y = -15.8 - 0.0451*\text{Weight\_BD} + 0.0624*\text{Systiolic\_BP} - 0.543*\text{HB} + 0.457*\text{Gestation} - 1.122*\text{Preference\_delivery\_type} - 4.49*\text{Previous\_CD} - 2.77*\text{Breech} + 8.15*\text{Parity} + 1.418*\text{Exercise}$$

The most affecting variable on delivery type are weight before delivery, systolic BP, HB, Gestation Period, Preference for delivery type, previous Caesarean delivery, Breech Pregnancy, Parity and Exercise during pregnancy.

Coefficients:

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	-15.8	10.7	-1.48	0.140	
Weight_BD	-0.0451	0.0311	-1.45	0.147	1.17
Systiolic_BP	0.0624	0.0366	1.71	0.088	1.16
HB	-0.543	0.290	-1.87	0.061	1.25
Gestation	0.457	0.232	1.97	0.049	1.23
Preference_delivery_type	-1.122	0.628	-1.79	0.074	1.37
Previous_CD	-4.49	1.36	-3.31	0.001	14.83
Breech	-2.77	1.33	-2.08	0.038	1.25
Parity	8.15	2.45	3.32	0.001	14.76
Exercise	1.418	0.840	1.69	0.091	1.33

P-value for variable systolic BP, HB, Gestation Period, Preference for delivery type, previous Caesarean delivery, Breech Pregnancy, Parity and Exercise during pregnancy is less than 0.1 and for Weight Before delivery is less than 0.15 thus all variable used in model are significant effect on delivery type at 10% level of significance except Weight Before delivery which is significant for 15% level of significance.

Odds Ratios for Continuous Predictors:

	Odds Ratio	95% CI
Weight_BD	0.9559	(0.8994, 1.0160)
Systiolic_BP	1.0644	(0.9908, 1.1435)
HB	0.5808	(0.3291, 1.0250)
Gestation	1.5798	(1.0028, 2.4889)
Preference_delivery_type	0.3255	(0.0950, 1.1155)
Previous_CD	0.0113	(0.0008, 0.1605)
Breech	0.0629	(0.0046, 0.8538)
Parity	3480.2330	(28.3843, 426715.2193)
Exercise	4.1279	(0.7956, 21.4173)

Model is built by assuming Normal Delivery as Success.

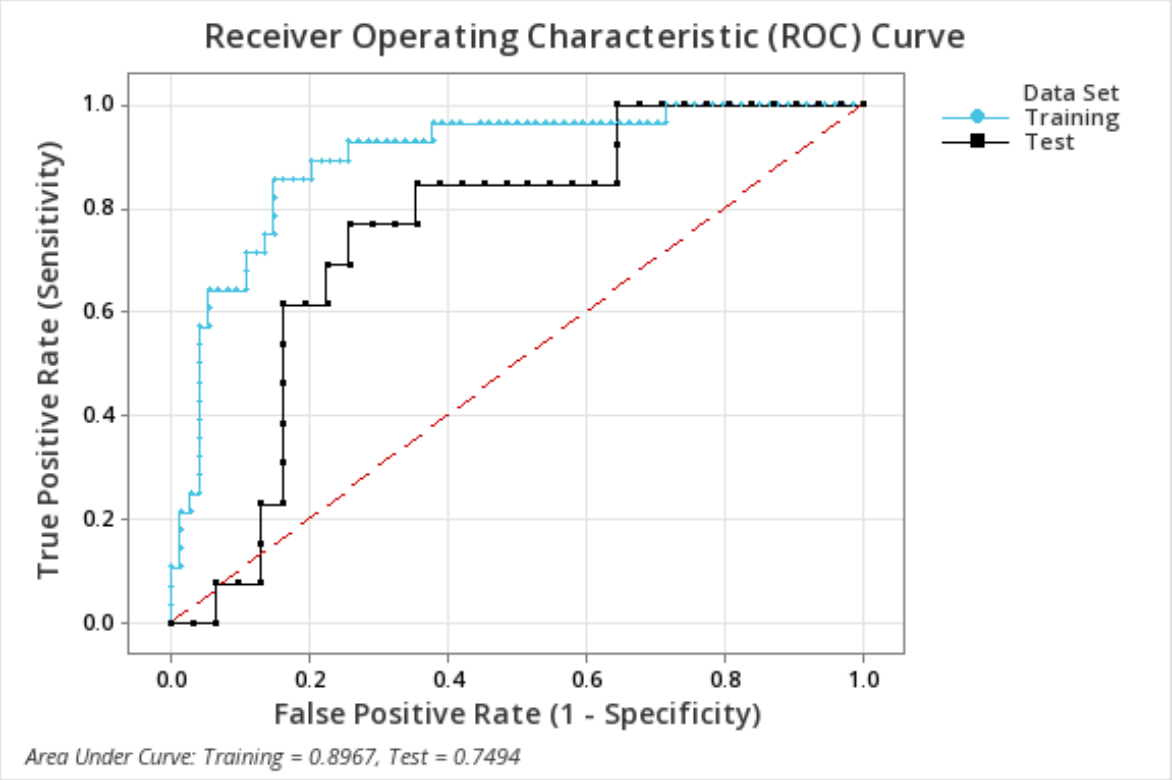
Now,

From odds ratio in above table.

1. If weight before delivery is increased by 1kg, then the chance of normal delivery is decreased by 5%.
2. If gestation period increased by 1 week within range then chance of normal delivery is increased by 57%.
3. Odds ratio for Exercise is 4.1279 thus chance of normal delivery increase by doing exercise during pregnancy 4.1279 times as compared to by taking rest only.

Model Summary:

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve	Test Deviance R-Sq	Test Area Under ROC Curve
38.88%	31.37%	93.28	95.70	119.53	0.8967	0.00%	0.7494



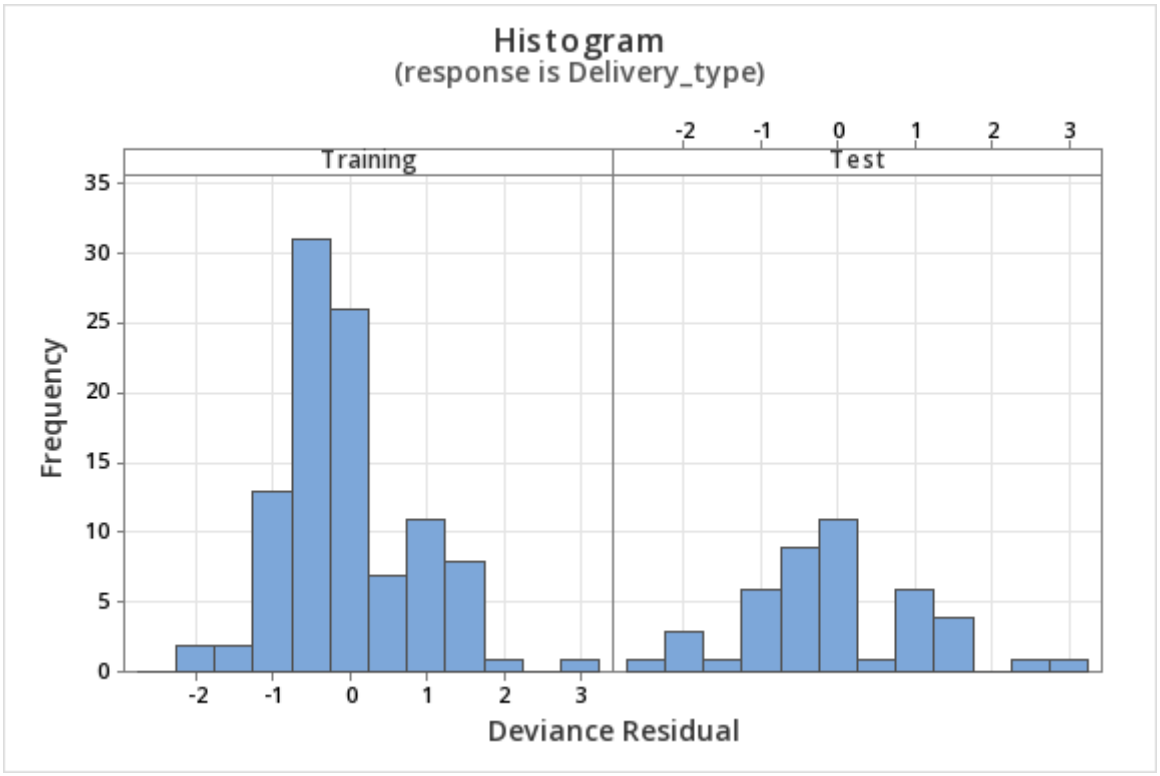
1. Area under receiver operating characteristic curve for training dataset is 0.8967 which is excellent discrimination.
2. Area under receiver operating characteristic curve for testing dataset is 0.7494 which is acceptable discrimination.

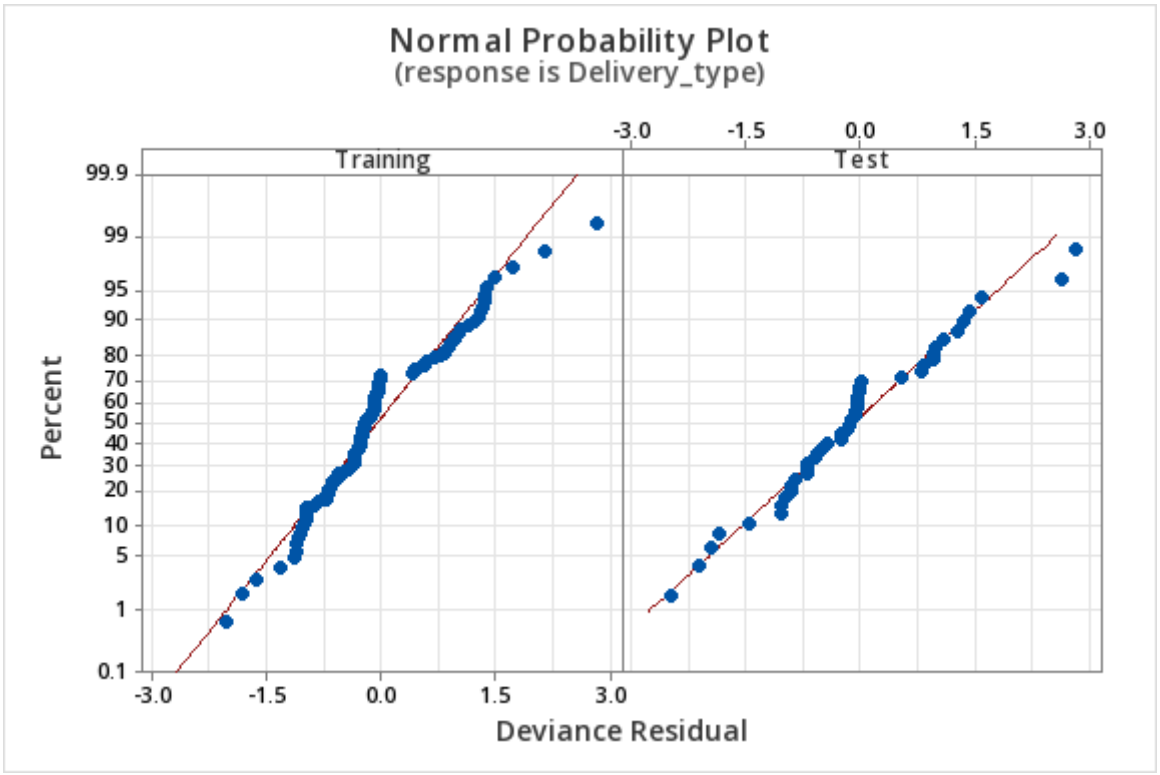
Goodness-of-Fit Tests:

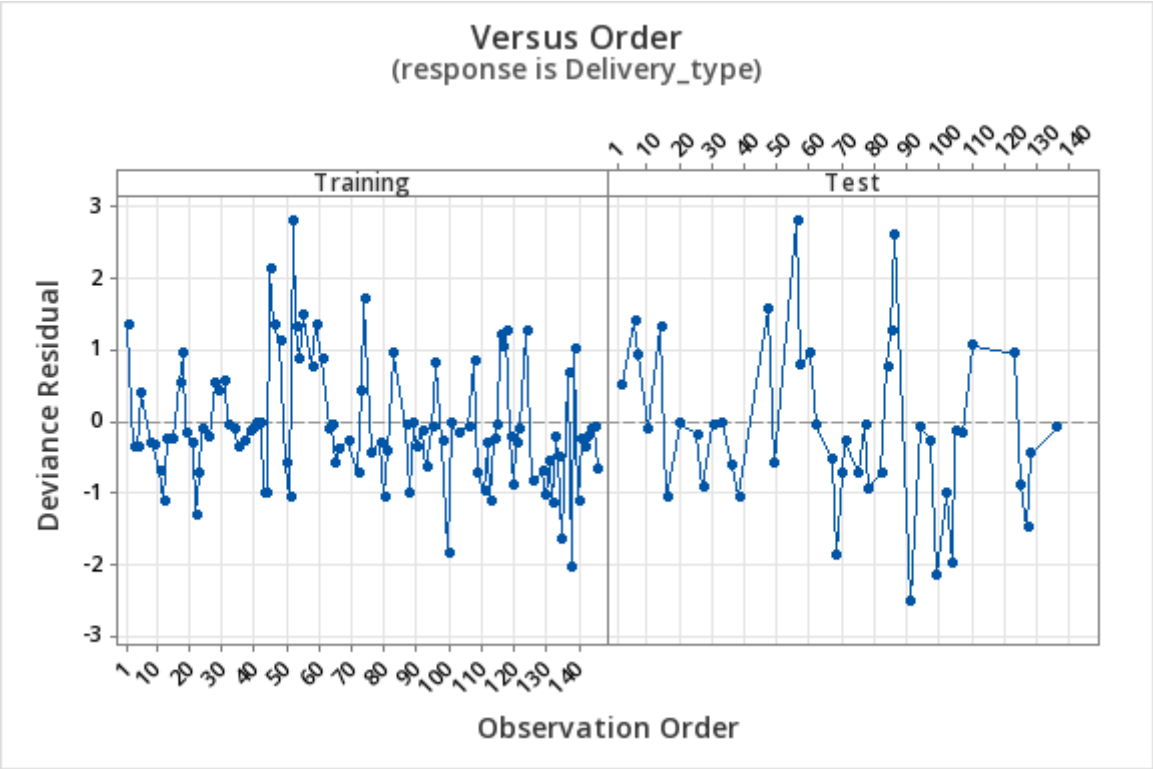
Test	DF	Chi-Square	P-Value
Deviance	92	73.28	0.925
Pearson	92	111.11	0.085
Hosmer-Lemeshow	8	8.04	0.430

From Deviance, Pearson and Hosmer-Lemeshow test for goodness of fit model fit data significantly as p-vaue for this statistic are less than 0.05.

Residual analysis:







1. The percent deviance residuals in logistic regression are normally distributed, it indicates that the model's assumptions are being met, and the model is a good fit for the data. In this case, the model summary would generally show reliable and accurate estimates of the coefficients, as well as valid statistical tests for their significance.
2. A normally distributed percent deviance residual plot suggests that the model is appropriately capturing the relationship between the predictors and the outcome variable.
3. The residuals are randomly scattered around zero, indicating that the predicted probabilities align well with the observed outcomes.

Conclusion:

1. Based on the results of the logistic regression analysis, several variables were found to significantly affect the delivery type. These variables include weight before delivery, systolic blood pressure (BP), hemoglobin (HB) levels, gestation period, preference for delivery type, previous Caesarean delivery, breech pregnancy, parity, and exercise during pregnancy.
2. For every 1 kg increase in weight before delivery, the chance of normal delivery decreases by 5%. This means that higher weight before delivery is associated with a higher likelihood of Caesarean delivery compared to normal delivery.

3. For every 1-week increase in gestation period within the given range, the chance of normal delivery increases by 57%. This suggests that longer gestation periods are strongly associated with a higher probability of having a normal delivery.
4. The odds ratio for exercise during pregnancy is 4.1279. This indicates that the chance of having a normal delivery is approximately 4.1279 times higher for women who engage in regular exercise during pregnancy compared to those who only take rest. This suggests a significant positive association between exercise during pregnancy and the likelihood of normal delivery.
5. The training dataset exhibits excellent discrimination with an area under the receiver operating characteristic (ROC) curve of 0.8967.
6. The testing dataset demonstrates acceptable discrimination with an area under the ROC curve of 0.7494.
7. The normally distributed percent deviance residuals in logistic regression suggest that the model's assumptions are met, indicating a good fit for the data.

WORD CLOUD FOR CAESAREAN REASON

Were,

MAS: The scientific term for a situation where a baby passes feces (meconium) in the uterus is "meconium-stained amniotic fluid." This condition is also known as meconium aspiration syndrome (MAS). Meconium-stained amniotic fluid can occur before or during labor and may indicate that the baby is experiencing distress. In such cases, a caesarean section may be performed to expedite the delivery and prevent the baby from inhaling or aspirating meconium, which can lead to respiratory problems and other complications.

CPD: stands for "cephalopelvic disproportion." It is a condition that occurs during pregnancy and refers to a mismatch in size between the baby's head and the mother's pelvis. In CPD, the baby's head is larger than the maternal pelvis, making it difficult for the baby to pass through the birth canal during labor. This condition can lead to prolonged labor, difficulty in delivering the baby vaginally, and may increase the risk of complications for both the mother and the baby.

ROM: The loss of water during delivery refers to the rupture of the amniotic sac, which contains the amniotic fluid surrounding the baby in the uterus. This rupture is commonly known as "breaking the water" or "water breaking." The medical term for this event is "rupture of membranes" (**ROM)**.

Cervical dystocia

referring to a condition where the cervix does not dilate sufficiently during labor, it is known as "**cervical dystocia**" or "cervical failure to progress."

This can result in a prolonged labor or difficulty in delivering the baby vaginally.

Macrosomia: refers to a condition where the baby is larger than average for gestational age. This can occur when the baby experiences accelerated growth in the womb, often associated with maternal factors such as gestational diabetes or excessive weight gain during pregnancy.

Transverse lie: The horizontal position of a baby in the uterus is referred to as "**transverse lie**" or "transverse presentation.

Conclusion:

Rupture of the amniotic sac, leading to the loss of water during delivery, is identified as a primary factor contributing to the incidence of caesarean delivery and other are as shown in word cloud above.

MAJOR FINDINGS

1. Weight before delivery, systolic blood pressure, hemoglobin levels, gestation period, preference for delivery type, previous Caesarean delivery, breech pregnancy, parity, and exercise during pregnancy were identified as significant factors affecting the delivery type.
2. Higher weight before delivery was associated with a higher likelihood of Caesarean delivery.
3. Longer gestation periods were strongly associated with a higher probability of normal delivery.
4. Engaging in regular exercise during pregnancy was positively associated with the likelihood of normal delivery.
5. Preferences for Caesarean delivery, previous Caesarean delivery, and breech pregnancy were identified as factors influencing the choice of Caesarean delivery.
6. The project may have highlighted the increasing rates of Caesarean deliveries and the need for strategies to reduce unnecessary interventions and improve maternal and neonatal health outcomes.

LIMITATIONS

1. Sample size: The project may have a relatively small sample size, which could limit the generalizability of the findings. A larger sample size would provide a more representative population and enhance the statistical power of the analysis.
2. Selection bias: There could be potential selection bias in the sample, as the data may have been collected from specific maternity hospitals or specific populations within Kolhapur. This could affect the external validity of the results and limit their applicability to other settings or populations.
3. Recall bias: The reliance on historical data could introduce recall bias or missing information, impacting the reliability of the findings.
4. Self-reporting and social desirability bias: The project utilized questionnaires for data collection, there is a possibility of self-reporting bias, where participants may provide socially desirable responses or inaccurately recall information. This could introduce bias and impact the validity of the results.
5. Confounding variables: The project may not have accounted for all potential confounding variables that could influence the relationship between the variables studied and the delivery type. Uncontrolled confounding variables could affect the accuracy of the results and limit the ability to draw causal conclusions.
6. Temporal limitations: The project have been conducted during a specific time period, and the findings may not reflect changes or trends in maternity health in Kolhapur that have occurred since the data collection. This could limit the project's relevance to the current context.

REFERENCE

1. *Collet, D. (2015). Modeling Survival Data in Medical Research. London: Chapman and Hall.*
2. *Hosmer D.W. and Lemeshow S. (2000): Applied Logistic regression, 2nd ED. Wiley New York.*
3. *Fundamentals of Mathematical Statistics: S.C. Gupta and V.K. Kapoor.*
4. *Statistical Study About The Causes Of Caesarian Section Deliveries In Al-Hindia City: Najat Hamza Hassan: Scientific Journal of Medical Research: Vol. 2, Issue 7, pp 150-155, Summer 2018*