ANALYSIS OF DISCONTINUATION AND NON-ENROLLMENT OF STUDENTS OF VARIOUS SOCIO-ECONOMIC POPULATIONS IN WEST BENGAL AND COMPARISON WITH NATIONAL AVERAGE: STUDY BASED ON NSSO-71ST ROUND EDUCATION SURVEY DATA

PROJECT BY

TANISHA DAS

ADRIJA SENGUPTA

Introduction to NSSO

The National Sample Survey Office (NSSO), led by a Director General, plays a crucial role in gathering large-scale data across various socio-economic areas in India. NSSO conducts nationwide household surveys, runs the Annual Survey of Industries (ASI), tracks rural and urban prices and helps enhance crop statistics through area enumeration and crop estimation surveys.

The National Sample Survey Organization (NSSO) first conducted an all-India survey on social consumption during its 35th round from July 1980 to June 1981 with a goal of gathering data on how public spending on various services, like education, mass immunization, family welfare programs, and healthcare, benefited the people. Following this initial survey, the NSSO continued to explore social consumption in subsequent rounds: The 42nd round (July 1986 to June 1987), the 52th round (July 2007 to June 2008).

The 71st round of the National Sample Survey (NSS) on education took place from January 1 to June 30, 2014. This survey aimed to collect detailed information about the participation of individuals aged 5 to 29 years in various educational activities. With the increasing role of Information Technology in everyday life, this round of the survey incorporated new approaches and methods, reflecting changes in the educational landscape since the previous survey round (the 64th, conducted from July 2007 to June 2008). The main difference between the 71st and 64th rounds of the survey Since Telangana was not a separate state at the start of the survey period, there was no specific code for it in the survey's Schedule of Enquiry.

Objective of the Project

The project aims to understand whether a person's discontinuation/dropping out from education and the unenrollment reasons depend on various socio-economic factors, like the UMPCE with other possible covariates sectors (Urban, Rural), gender (Male, Female), Household type. We carry on this process for both West Bengal and All-India separately and then further move on to comparing them.

For discontinuation/dropping out:

$x_i = 1$, if enrolled

$\quad = 0$, otherwise

$y_i = 1$, if enrolled but discontinued with reason i

$\quad = 0$, otherwise

Let us define the population parameter as R = Population proportion of persons who have enrolled but discontinued for reason i. We estimate R by Rhat as Rhat(enrolled but discontinued for reason i) = Yhat / Xhat, where Xhat = Estimated number of never- enrolled persons and Yhat = Estimated number of enrolled persons who discontinued/dropped out for reason i.

For never-enrollment:

$x_i = 0$, if never-enrolled

$\quad = 1$, otherwise

$y_i = 1$, if never enrolled for reason i

$\quad = 0$, otherwise

We find out the estimates of the ratio $\hat{R}$, their MSE (mean square error), and RSE (relative square error) in percentage for various socio-economic population sub-groups. Then carry out the logistic regression of several well-defined binary response variables with the possible explanatory variables, say, considering the quintile classes of UMPCE and other socio-economic covariates (if any).

Estimation Procedure

Notations:

s = subscript for s-th stratum

t = subscript for t-th sub-stratum

m = subscript for sub-sample (m =1, 2)

Overall Estimate for Aggregates for a Sub-Stratum:

Overall estimate for aggregates for a sub-stratum ($\hat{Y}_{st}$) based on two subsamples in a sub-stratum is obtained as

$$\hat{Y}_{st} = \frac{1}{2} \sum_{m=1}^{2} \hat{Y}_{stm}$$

Overall Estimate for Aggregates for a Stratum:

Overall estimate for a stratum ($\hat{Y}_s$) will be obtained as

$$\hat{Y}_s = \sum_t \hat{Y}_{st}$$

Overall Estimate of Aggregates at State/UT/all-India level:

The overall estimate $\hat{Y}$ at the State/ UT/ all-India level is obtained by summing the stratum estimates $\hat{Y}_s$ over all strata belonging to the State/ UT/ all-India.

Estimates of Ratios:

Let $\hat{Y}$ and $\hat{X}$ be the overall estimates of the aggregates Y and X for two characteristics y and x respectively at the State/ UT/ all-India level. Then the combined ratio estimate ($\hat{R}$) of the ratio ($R = \frac{Y}{X}$) will be obtained as

$$R = \frac{\hat{Y}}{\hat{X}}$$

Estimates of Error:

The estimated variances of the above estimates will be as follows

For aggregate $\hat{Y}$:

$$\widehat{Var}(\hat{Y}) = \sum_s \widehat{Var}(\hat{Y}_s) = \sum_s \sum_t \widehat{Var}(\hat{Y}_{st})$$

For ratio $\widehat{R}$:

$$\widehat{MSE}(\widehat{R}) = \frac{1}{4\widehat{X}^2} \sum_s \sum_t \ (\widehat{Y}_{st\,1} - \widehat{Y}_{st\,2})^2 + \widehat{R}^2 \ (\widehat{X}_{st\,1} - \widehat{X}_{st\,2})^2 - 2\widehat{R}$$

$$(\widehat{Y}_{st\,1} - \widehat{Y}_{st\,2}) \ (\widehat{X}_{st\,1} - \widehat{X}_{st\,2})$$

Estimates of Relative Standard Error (RSE):

- $\widehat{RSE}(\widehat{Y}) = \frac{\sqrt{\widehat{Var}(\widehat{Y})}}{\widehat{Y}} \times 100$

- $\widehat{RSE}(\widehat{R}) = \frac{\sqrt{\widehat{MSE}(\widehat{R})}}{\widehat{R}} \times 100$

Multipliers

Schedule type	Sector	Formula for multipliers	
		hg/ sb 1	hg/ sb 2
25.2	Urban/Rural	$\dfrac{Z_{st}}{n_{stmj}} \times \dfrac{1}{z_{stmi}} \times \dfrac{H_{stmi1j}}{h_{stmi1j}}$	$\dfrac{Z_{st}}{n_{stmj}} \times \dfrac{1}{z_{stmi}} \times D^{*}_{stmi}$ $\times \dfrac{H_{stmi2j}}{h_{stmi2j}}$
		(j=1,2,3)	

Interpretation of Rhat values for West Bengal and India

We see that the number of individuals who unenrolled due to some reasons are low so we merged them together as "others".

We merge 6th ("timings of educational institution not suitable"), 7th ("language/medium of instruction used unfamiliar"), 8th ("inadequate number of teachers"), 10th ("Never enrolled- no tradition in the community"), 12th ("Ever enrolled unfriendly atmosphere at school"), 14th ("Ever enrolled preparation for competitive examination"), 15th ("Girl student – non-availability of female teacher"), 16th ("Girl student – non-availability of girls toilet") never enrolment reasons with 18th ("others") reason and consider them as "others".

On the basis of sector (urban, rural) and gender (male, female) the data can be divided into 9 parts. The 9 divisions are:

1. Urban, Male

2. Urban, Female

3. Urban, Male + Female

4. Rural, Male

5. Rural, Female

6. Rural, Male + Female

7. Urban + Rural, Male

8. Urban + Rural, Female

9. Urban + Rural, Male + Female

West Bengal

- For Urban, Male: The main reason for dropping out after enrollment is "financial constraints" and for unenrollment is "not interested in education".

- For Urban, Female: The main reason for both dropping out after enrollment and unenrollment is "financial constraints".

- For Urban, Male + Female: The main reason for dropping out after enrollment is "financial constraints" and unenrollment is "not interested in education".

- For Rural, Male: The main reason for dropping out after enrollment is "financial constraints" and unenrollment is "not interested in education".

- For Rural, Female: The main reason for both dropping out after enrollment and unenrollment is "financial constraints".

- For Rural, Male + Female: The main reason for both dropping out after enrollment and unenrollment is "financial constraints".

- For Urban+ Rural, Male: The main reason for dropping out after enrollment is "financial constraints" and for unenrollment is "not interested in education".

- For Urban + Rural, Female: The main reason for both dropping out after enrollment and unenrollment is "financial constraints".

- For Urban + Rural, Male + Female: The main reason for dropping out after enrollment is "financial constraints" and for unenrollment is "not interested in education".

India

- For Urban, Male: The main reason for dropping out after enrollment is "engaged in economic activities" and for unenrollment is "financial constraints".

- For Urban, Female: The main reason for dropping out after enrollment is "engaged in domestic activities" and for unenrollment is "financial constraints".

- For Urban, Male + Female: The main reason for dropping out after enrollment is "engaged in economic activities" and for unenrollment is "financial constraints".

- For Rural, Male: The main reason for dropping out after enrollment is "engaged in economic activities" and for unenrollment is "not interested in education".

- For Rural, Female: The main reason for dropping out after enrollment is "engaged in domestic activities" and for unenrollment is "Others".

- For Rural, Male + Female: The main reason for both dropping out after enrollment and unenrollment is "not interested in education".

- For Urban+ Rural, Male: The main reason for dropping out after enrollment is "engaged in economic activities" and for unenrollment is "not interested in education".

- For Urban + Rural, Female: The main reason for dropping out after enrollment is "engaged in domestic activities" and for unenrollment is "Others".

- For Urban + Rural, Male + Female: The main reason for both dropping out after enrollment and unenrollment is "not interested in education".

We have seen the main reason for dropping out after enrollment in West Bengal is "financial constraints" and in India it is either "engaged in economic activities" or "engaged in domestic activities".

This data underscores that while a range of factors influence why students remain in school, specific issues are more critical in determining why they drop out or never start their education.

Now we want to know what are the reasons that might lead to this.

Various Quintile Classes of UMPCE

The Usual Monthly Per Capita Consumer Expenditure (UMPCE) is a measure used to understand how much money each person in a household spends on average each month

The UMPCE is divided into five groups, called quintiles, which categorize households based on their spending levels:

1. Quintile 1 represents the lowest spending group.

2. Quintile 2 is the next higher spending group.

3. Quintile 3 is the middle group.

4. Quintile 4 includes households with higher spending.

5. Quintile 5 represents the highest spending group.

The distribution of these quintiles is the same for both households and individuals within those households. For instance, if you look at West Bengal or at the national level, the spending distribution among these quintile groups will be consistent within the rural or urban areas.

In simple terms, these quintiles help us understand and compare how spending varies from the least to the most in different areas.

Table for quintile classes

West Bengal

India

Quintile classes of UMPCE	UMPCE (in Rs)			
	Rural		Urban	
	Lower limit	Upper limit	Lower limit	Upper limit
1	0.00	800.00	0.00	1150.00
2	800.00	1000.00	1150.00	1500.00
3	1000.00	1214.29	1500.00	2000.00
4	1214.29	1500.00	2000.00	3142.86
5	1500.00	-	3142.86	-

Quintile classes of UMPCE	UMPCE (in Rs)			
	Rural		Urban	
	Lower limit	Upper limit	Lower limit	Upper limit
1	0.00	786.00	0.00	1200.00
2	786.00	1000.00	1200.00	1667.00
3	1000.00	1286.29	1667.00	2250.00
4	1287.29	1667.00	2250.00	3333.00
5	1667.00	-	3333.00	-

Chi-Square Presentation of the Explanatory Variables

We will test two hypotheses here :

1. **Null Hypothesis (H0)**: There is no relationship between the explanatory variable and the outcome variable. In other words, any observed association is due to chance.

2. **Alternative Hypothesis (H1)**: There is a relationship between the explanatory variable and the outcome variable. This means that the association observed in the data is significant and not due to random variation.

These hypotheses will guide our analysis, helping us determine whether the explanatory variables have a meaningful impact on the outcome variables.

The explanatory variables are:
Gender
Household type
UMPCE_Quintile

We take the variables y_1, y_2, y_3, y_4, y_5 as:

y_1 = 1 if ever enrolled but discontinued/dropped out

    = 0 otherwise

 y_2 = 1 if ever enrolled and discontinued before completing Primary education (Grade IV)

    = 0 otherwise

y_3 = 1 if ever enrolled and discontinued after completing Primary Education (Grade IV) but

      before completing Middle-Class Education (Grade VIII)

    = 0 otherwise

y_4 = 1 if ever enrolled and discontinued after completing Middle-Class Education (Grade VIII)

      but before completing Secondary Education (Grade X)

    = 0 otherwise

y_5 = 1 if ever enrolled and discontinued after completing Secondary Education (Grade X)

    = 0 otherwise

West Bengal

<u>Gender:</u>

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value > 0.05	p value > 0.05	No association	No association
y_2	p value > 0.05	p value > 0.05	No association	No association
y_3	p value > 0.05	p value > 0.05	No association	No association
y_4	p value > 0.05	p value > 0.05	No association	No association
y_5	p value > 0.05	p value > 0.05	No association	No association

Household type:

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value < 0.05	p value > 0.05	Association	No association
y_2	p value > 0.05	p value > 0.05	No association	No association
y_3	p value > 0.05	p value > 0.05	No association	No association
y_4	p value > 0.05	p value > 0.05	No association	No association
y_5	p value > 0.05	p value < 0.05	No association	Association

UMPCE_Quintile:

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value < 0.05	p value < 0.05	Association	Association
y_2	p value > 0.05	p value < 0.05	No association	Association
y_3	p value > 0.05	p value < 0.05	No association	Association
y_4	p value > 0.05	p value < 0.05	No association	Association
y_5	p value > 0.05	p value < 0.05	No association	Association

India

Gender:

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value > 0.05	p value > 0.05	Not associated	Not associated
y_2	p value > 0.05	p value > 0.05	Not associated	Not associated
y_3	p value > 0.05	p value > 0.05	Not associated	Not associated
y_4	p value > 0.05	p value > 0.05	Not associated	Not associated
y_5	p value > 0.05	p value > 0.05	Not associated	Not associated

Household type:

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value < 0.05	p value > 0.05	Association	No association
y_2	p value > 0.05	p value > 0.05	No association	No association
y_3	p value > 0.05	p value > 0.05	No association	No association
y_4	p value > 0.05	p value > 0.05	No association	No association
y_5	p value > 0.05	p value < 0.05	No association	Association

UMPCE_Quintile:

Study Variable	Relationship between p value and level of significance		Relationship between study variable and explanatory variable	
	Rural	Urban	Rural	Urban
y_1	p value < 0.05	p value < 0.05	Associated	Associated
y_2	p value > 0.05	p value > 0.05	Not associated	Not associated
y_3	p value > 0.05	p value > 0.05	Not associated	Not associated
y_4	p value > 0.05	p value > 0.05	Not associated	Not associated
y_5	p value > 0.05	p value < 0.05	Not associated	Associated

Logistic Regression Analysis

Now we want know how all the explanatory variables together effect the study variable. To understand this we carry out the logistic regression of the study variables.

We take the study variables y_1, y_2, y_3, y_4 and y_5 as indicator variables.

Now in case of logistic regression we took the different covariates under consideration :

- Sector
- Gender
- House hold type
- UMPCE_quintile

West Bengal

For y_1:

The odds ratio of Household type Casual labour in non-agriculture(0.5882404) and Householdtypeothers(0.9845554) is less than 1 therefore their probability decreases.

When we go up with umpce_quintile_factors the probability increases as their odds ratio increases.
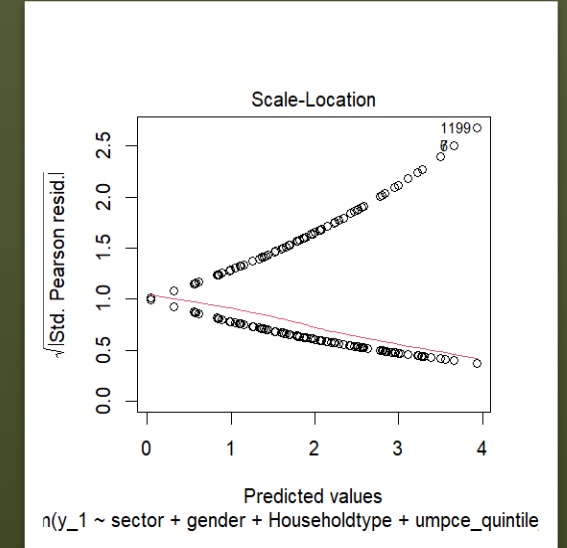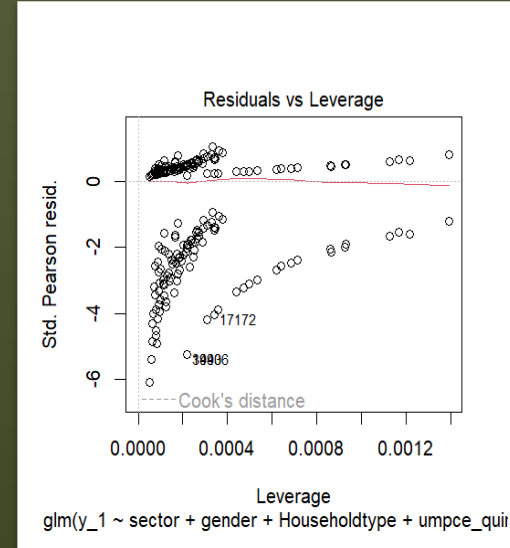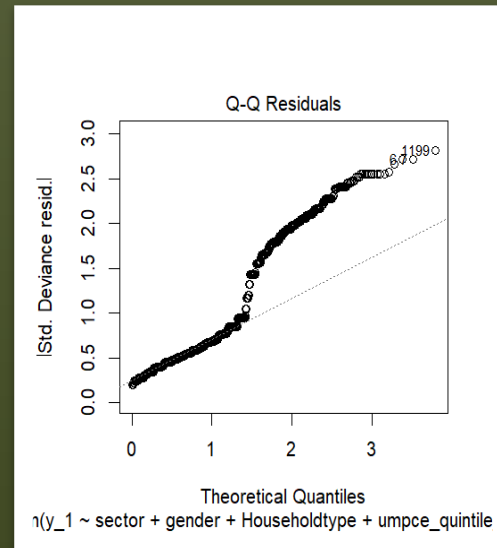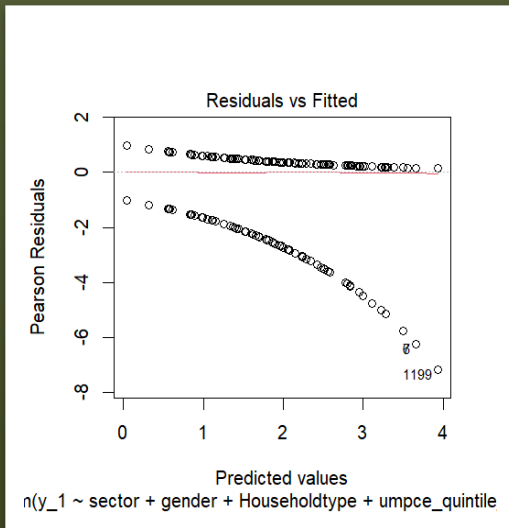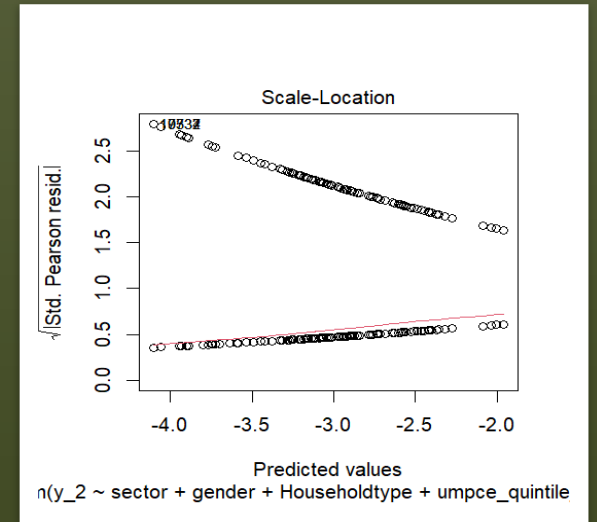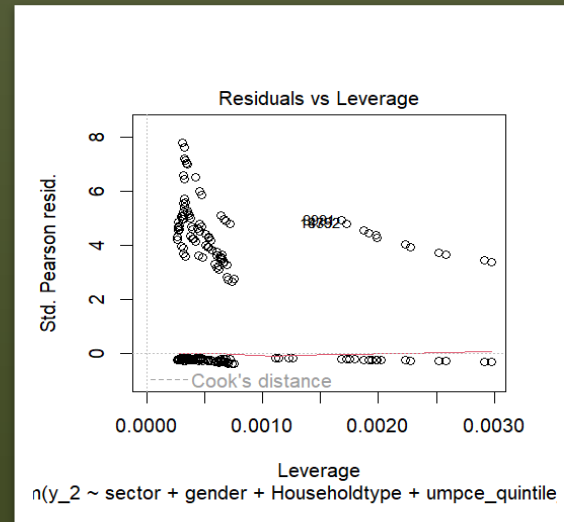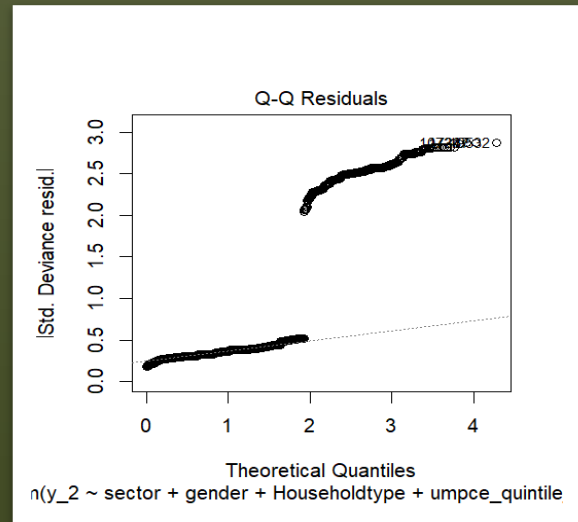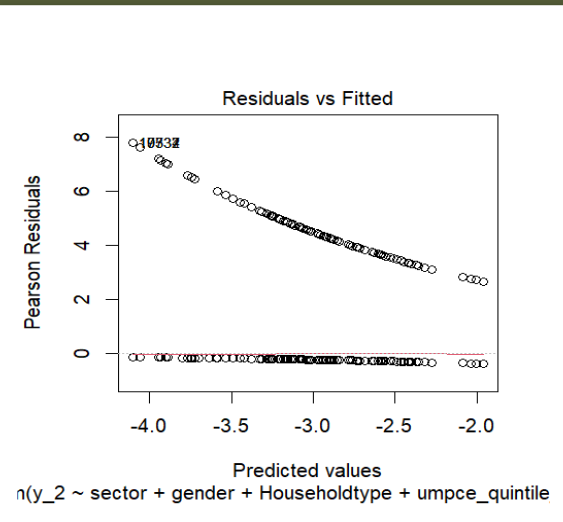
For y_2:

The odds ratio of sectorUrban(0.88083361), HouseholdtypeCasual labour in non-agriculture(0.94291365) and Householdtypeothers(0.28061199), HouseholdtypeSelf-employment in agriculture(0.70493710), HouseholdtypeSelf-employment in non-agriculture(0.59163784) is less than 1 therefore their probability decreases.
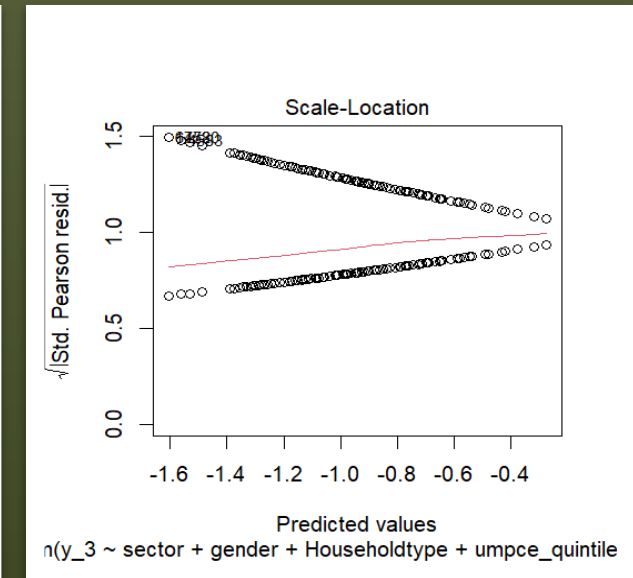
When we go up with umpce_quintile_factors the probability decreases as their odds ratio decreases.
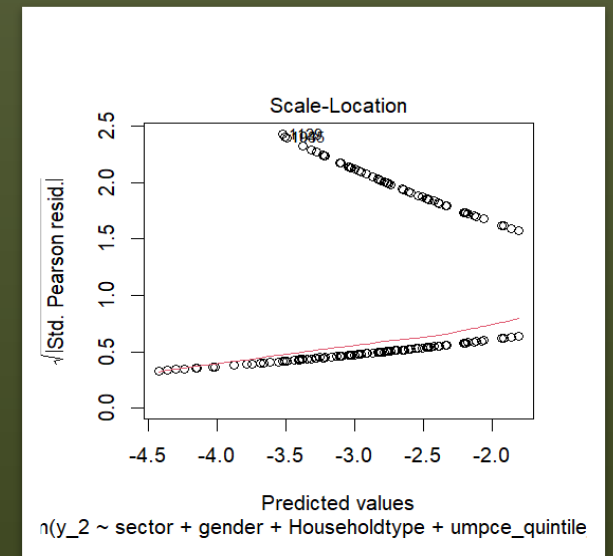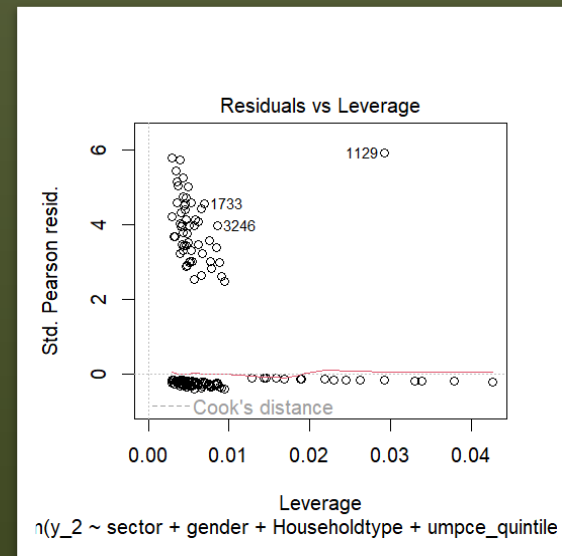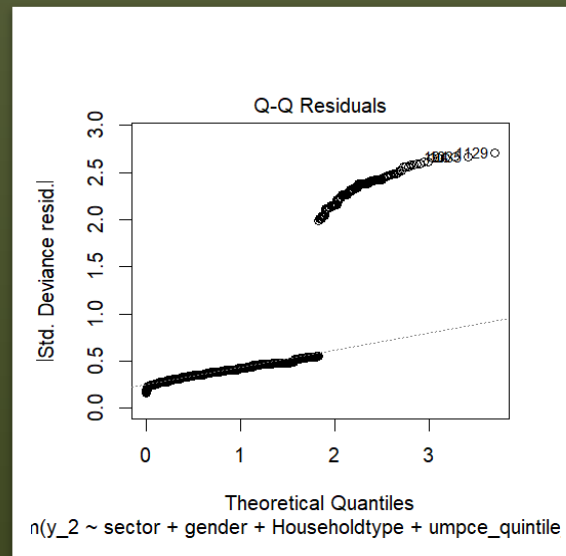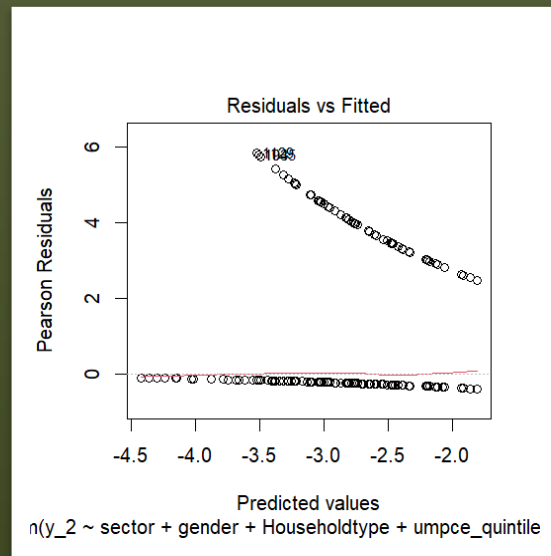
For y_3:

The odds ratio of HouseholdtypeCasual labour in non-agriculture(0.9802636) and Householdtypeothers(0.7452689), HouseholdtypeRegular wage/Salary earning(0.8975728), HouseholdtypeSelf-employment in agriculture(0.6905820), HouseholdtypeSelf-employment in non-agriculture(0.7469325) is less than 1 therefore their probability decreases.
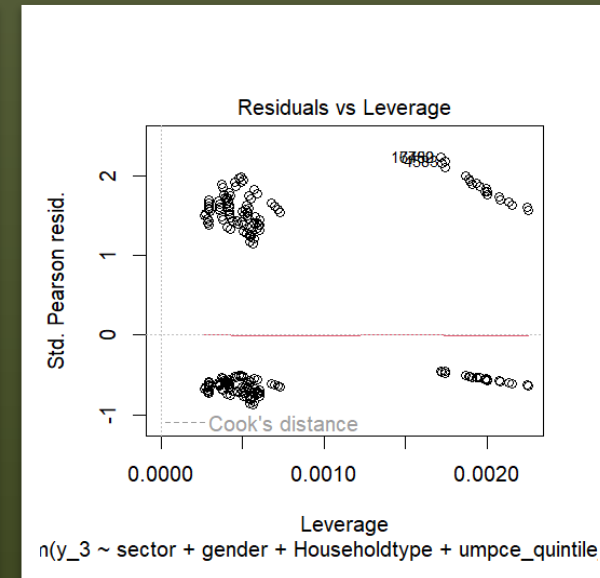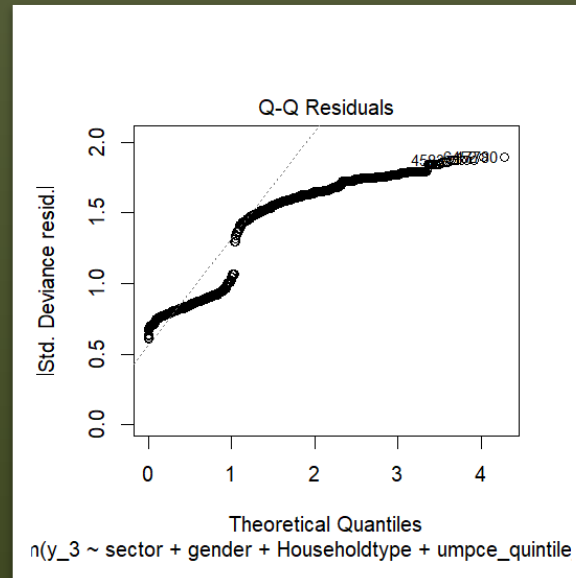
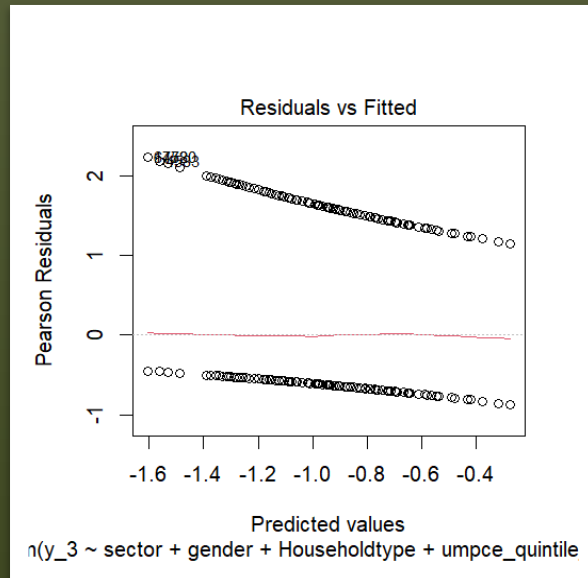When we go up with umpce_quintile_factors the probability decreases as their odds ratio decreases.
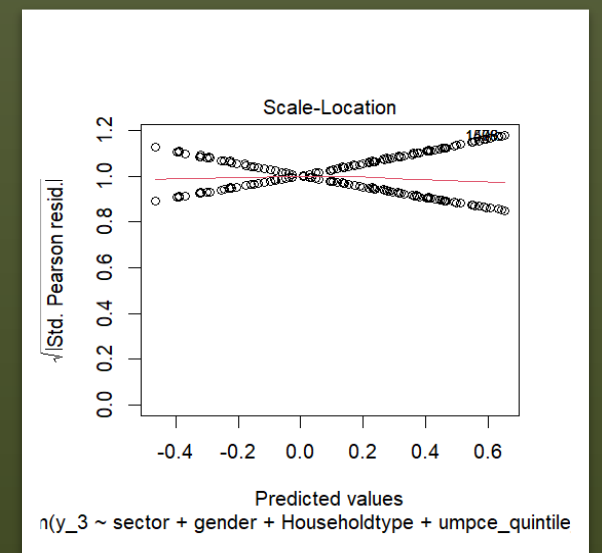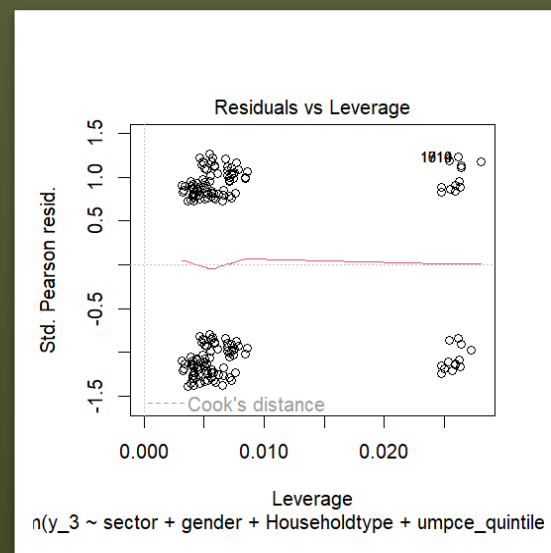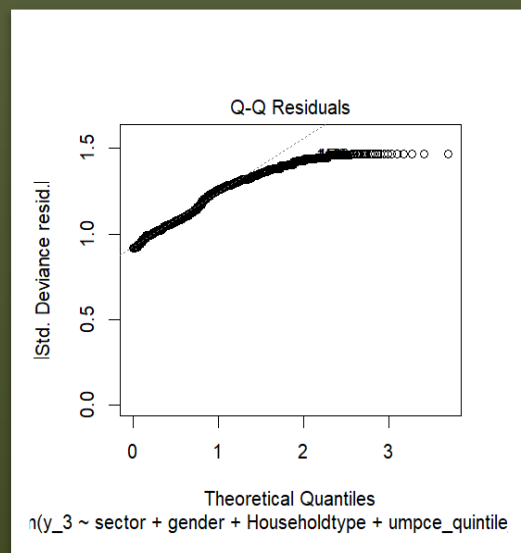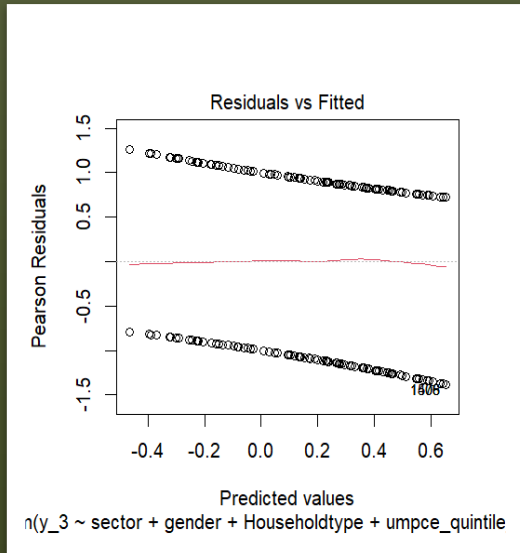
For y_4:

The odds ratio sectorUrban (0.7652154), genderMale (0.8119744) is less than 1 therefore their probability decreases.

When we go up with umpce_quintile_factors the probability increases as their odds ratio increases.
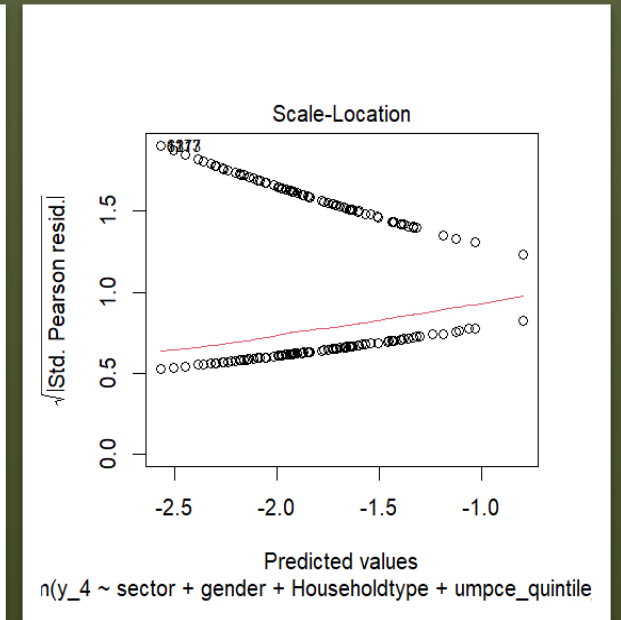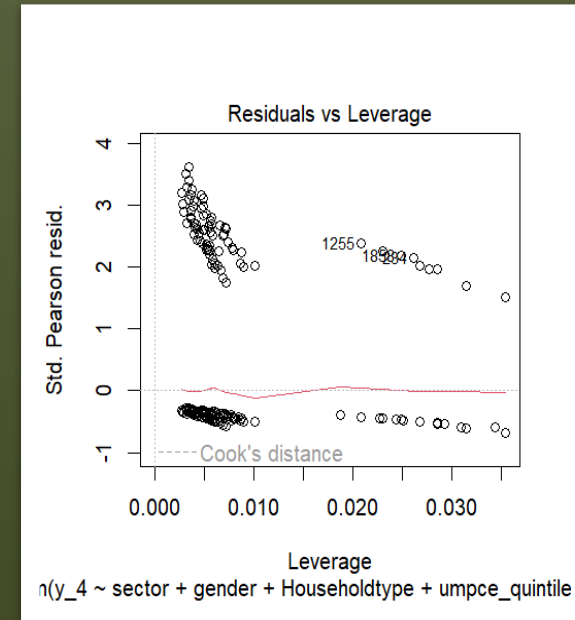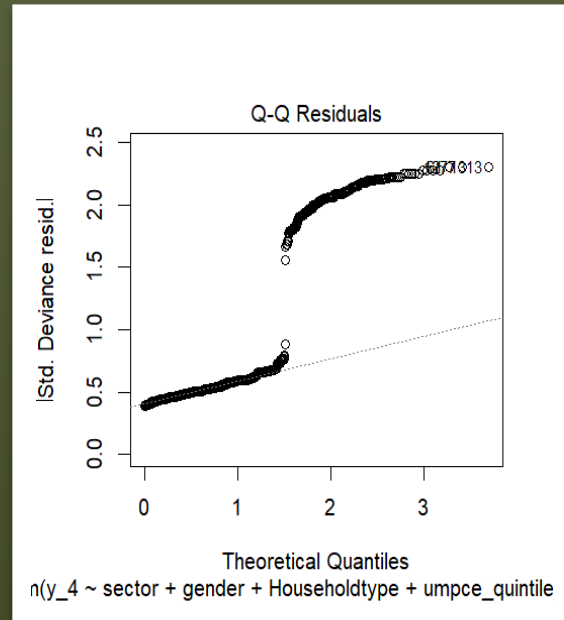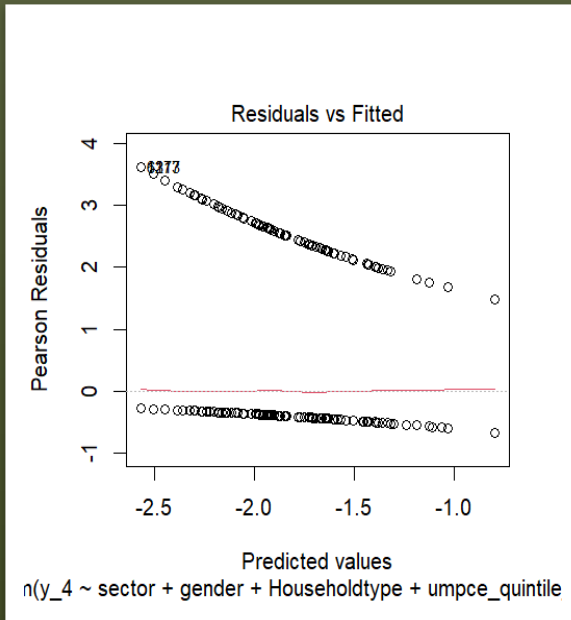
For y_5:

The odds ratio genderMale (0.8134795), HouseholdtypeCasual labour in non-agriculture(0.5771471) and Householdtypeothers(0.6242374), HouseholdtypeRegular wage/Salary earning(0.6837315) is less than 1 therefore their probability decreases.
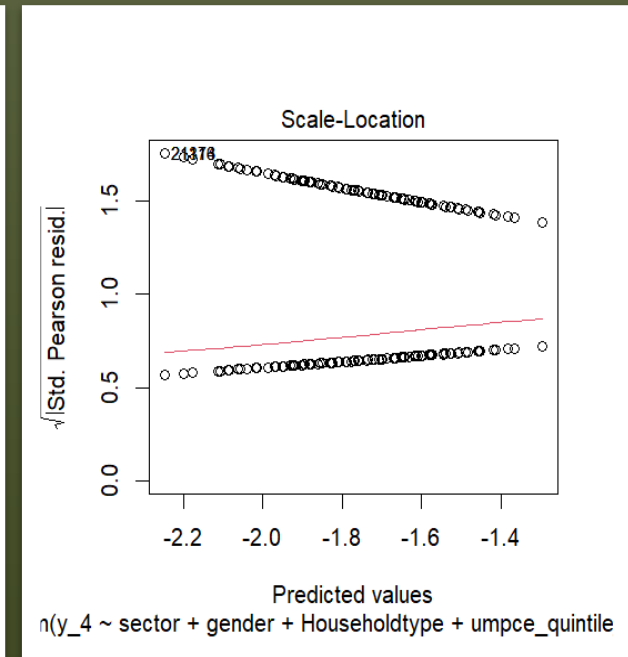
When we go up with umpce_quintile_factors the probability increases as their odds ratio increases and decreases as their odds ratio decreases.
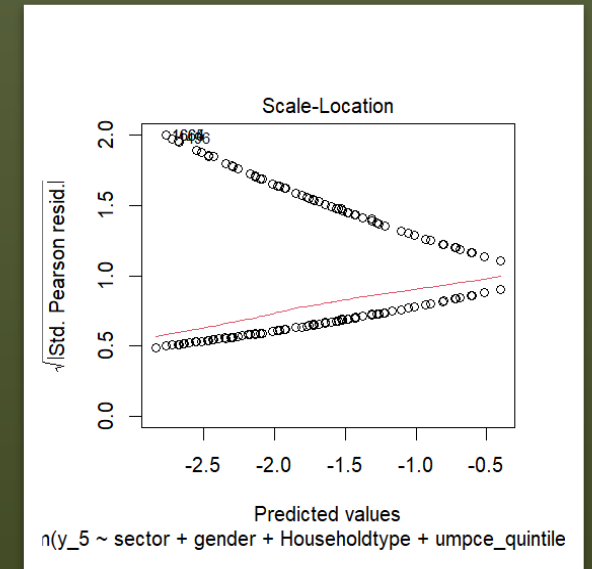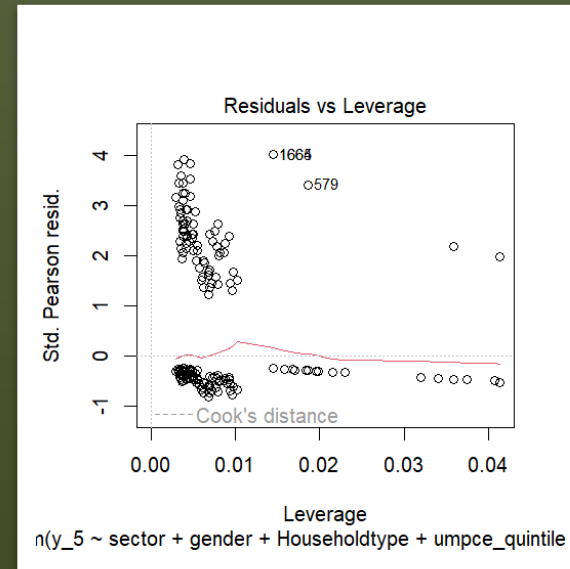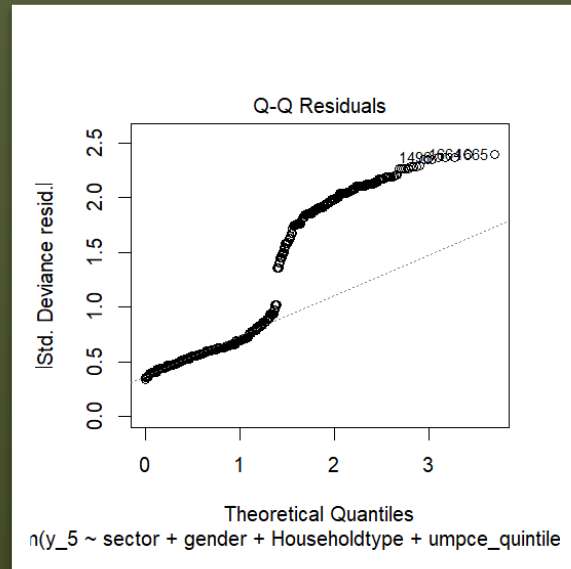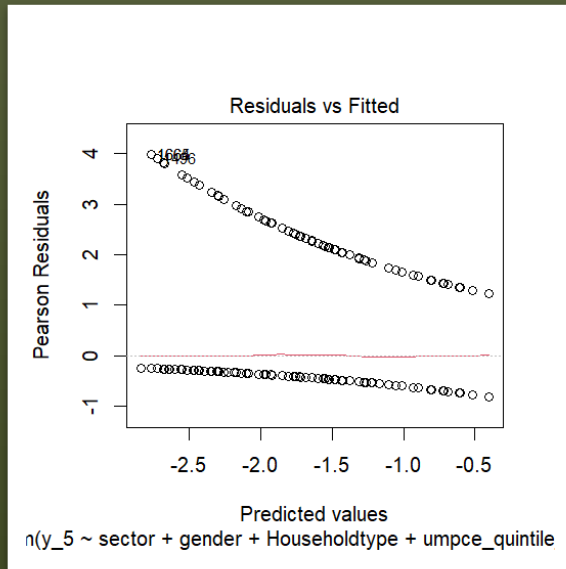
India

For y_1:

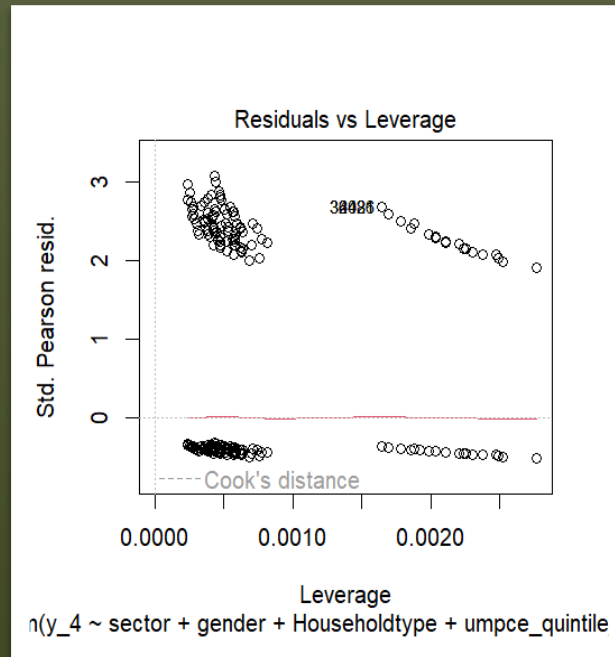The odds ratio of Household type Casual labour in non-agriculture(0.7916175) is less than 1 therefore their probability decreases.

When we go up with umpce_quintile_factors the probability increases as their odds ratio increases.

For y_2:

The odds ratio of sectorUrban, genderMale, Householdtype Casual labour in non-agriculture, Householdtypeothers, HouseholdtypeRegular wage/Salary earning, HouseholdtypeSelf-employment in agriculture, HouseholdtypeSelf-employment in non-agriculture are less than 1 therefore their probability decreases.
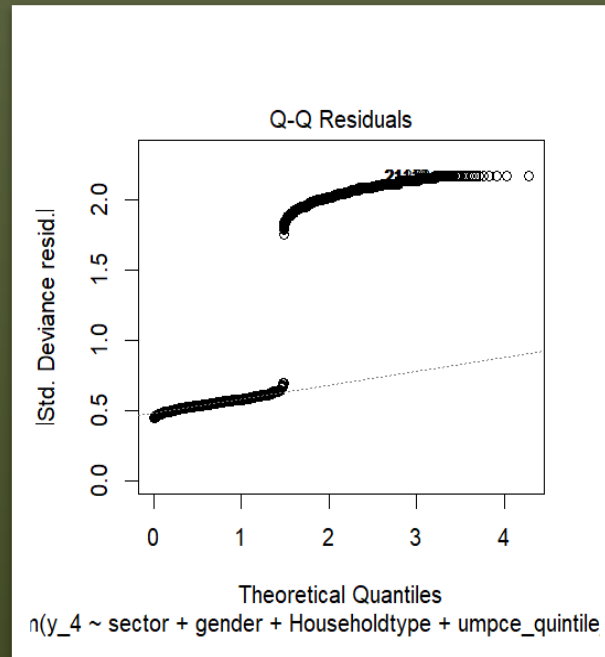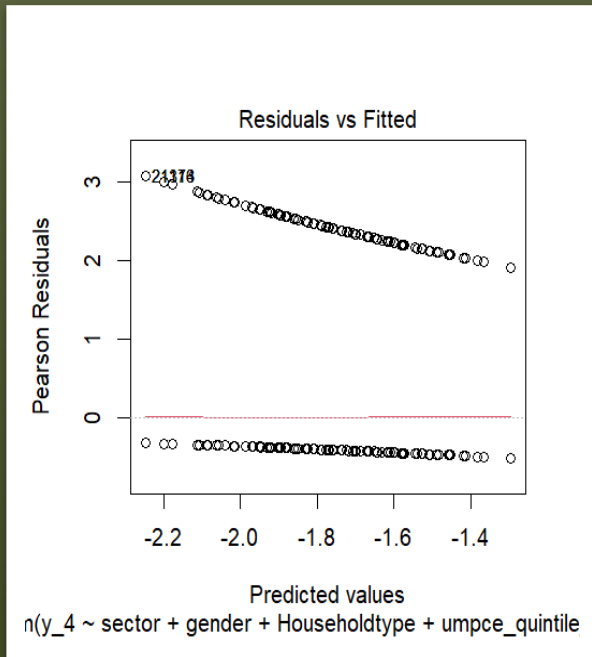
When we go up with umpce_quintile_factors the probability decreases as their odds ratio decreases.

- ## For y_3:

The odds ratio of sectorUrban, genderMale, Householdtype Casual labour in non-agriculture, Householdtypeothers, HouseholdtypeRegular wage/Salary earning, HouseholdtypeSelf-employment in agriculture, HouseholdtypeSelf-employment in non-agriculture are less than 1 therefore their probability decreases.

When we go up with umpce_quintile_factors the probability decreases as their odds ratio decreases.

For y_4:

The odds ratio of Householdtype Casual labour in non-agriculture (0.9533159) is less than 1 therefore their probability decreases.
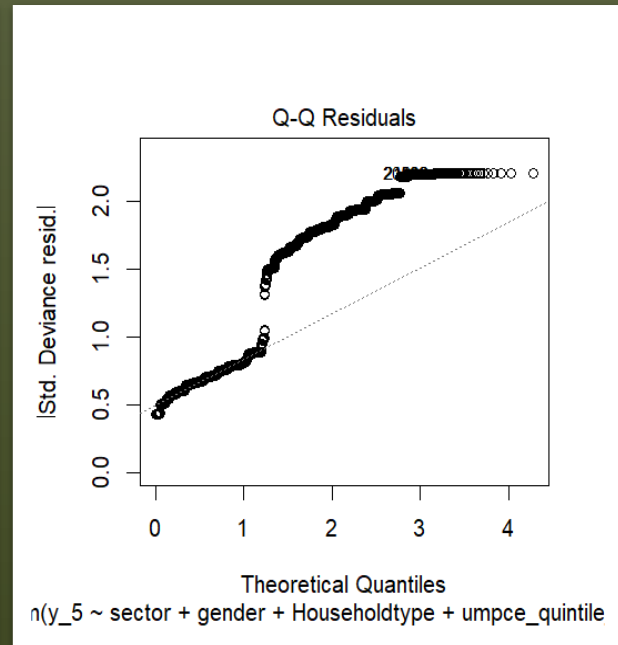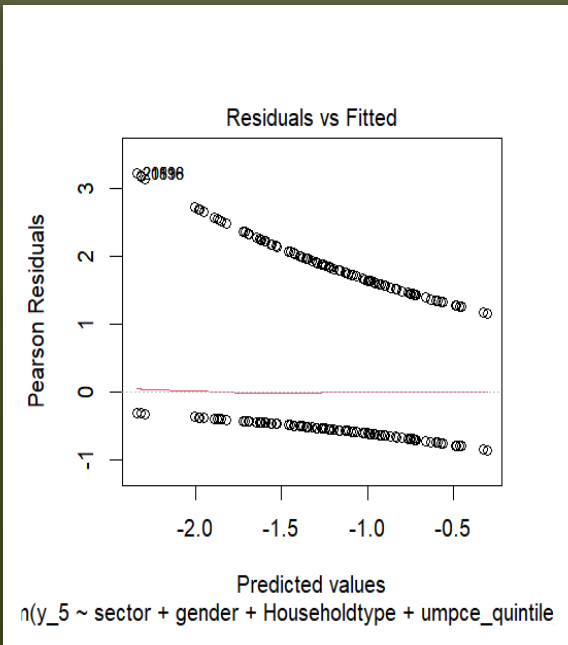
When we go up with umpce_quintile_factors the probability increases as their odds ratio increases.

For y_5:

The odds ratio of genderMale (0.97196655) is less than 1 therefore their probability decreases.

When we go up with umpce_quintile_factors the probability increases as their odds ratio increases.
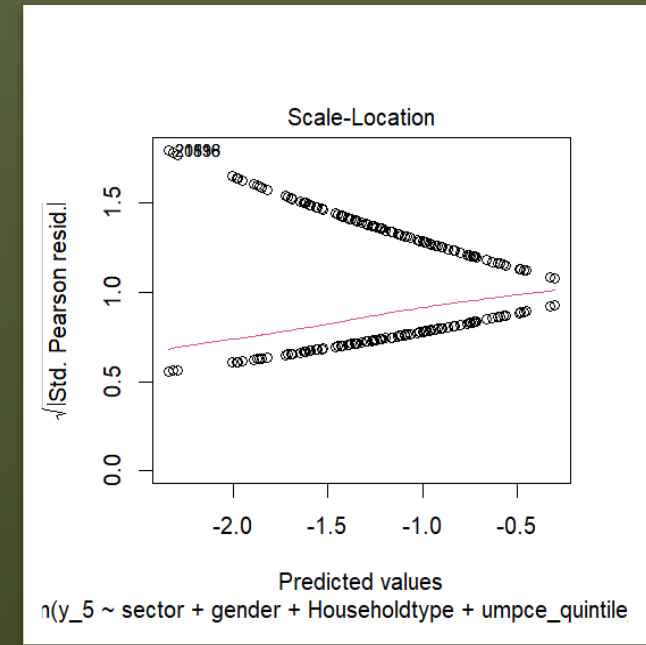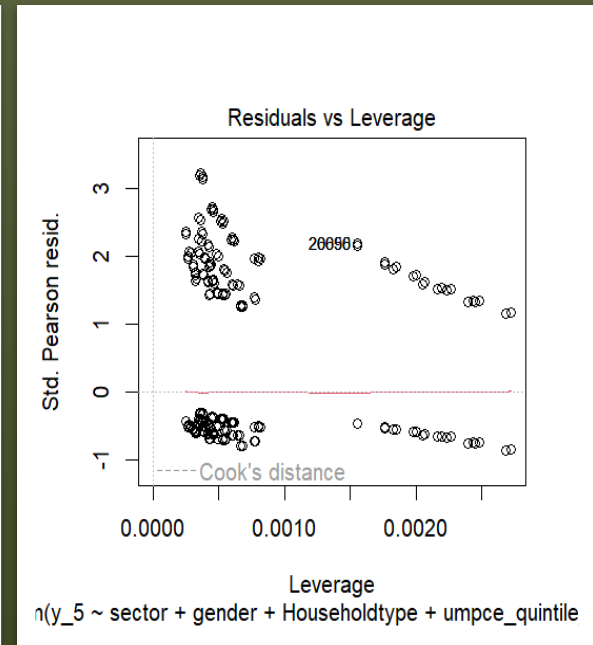
Conclusions

- From outputs of the Rhat values we found that the main reason for dropping out after enrollment in West Bengal is "financial constraints" and in India it is either "engaged in economic activities" or "engaged in domestic activities".

- From the Chi-Square test we found that the two explanatory variables, Gender and Household type have no association with the study variables in both West Bengal and India. The explanatory variable, UMPCE_Quintile is partially associated with study variables in both West Bengal and India.

- From the Logistic regression we see that Urban residents, higher expenditure quintiles, regular wage/salary jobs, and self-employment are associated with increased odds of the outcome, while men have lower odds compared to women.

These findings suggest that addressing financial constraints, increasing interest in education, and understanding the role of wealth and household types in access to resources are crucial for improving educational outcomes across different demographics in India.

References

Theory and Methods of Survey Sampling, Second Edition (2009) : By Parimal Mukherjee

Key Indicators of Social Consumption of India: Education NSS 71st Round (January 2014 – June 2014) National Sample Survey Office.

Note on Sample Design and Estimation Procedure of NSS of 71st Round(2014)

Regression Analysis – Chapter 14 Logistic Regression Models: Shalabh, IIT Kanpur.

ACKNOWLEDGEMENT

THANK YOU