APPLIED MICROECONOMETRICS: COMPUTER ASSIGNMENT -2

Analysis of Time Use Survey-2019 (Gujarat 24)

Group 5

By

Dilipkumaar (EE/2023-25/008)

Harish Chander P (EE/2023-25/011)

Priyadharshini R (EE/2023-25/020)

Vasanth D (EE/2023-25/026)

Madras School of Economics

November 3, 2024

References:

https://github.com/

https://stackoverflow.com/

https://stats.stackexchange.com/

https://economics.stackexchange.com/

https://rpubs.com/

https://cran.r-project.org/

https://www.princeton.edu/~otorres/

https://www.econometrics-with-r.org/

INTRODUCTION

This study examines the time use patterns of youth in the state of Gujarat, aged 15 to 29, utilizing data from the National Sample Survey Office (NSSO). The focus is on understanding how young people allocate their time across various activities, with particular attention to educational engagement, employment, and those not engaged in education, employment, or training (NEET). Key variables include **in_educ**, indicating whether a youth attends an educational institute, and **activity**, which categorizes youth into education, employment, or NEET groups. In addition to these main indicators, the analysis considers a range of demographic and socioeconomic variables, such as **gender**, **marital status**, **type of the day**, **lighting, dwelling type**, **cooking fuel, sector**, **social group**, and **religion**, providing insights into potential influences on time allocation.

Time use is further broken down by specific categories like **time_learn** & **sh_learn_time** reflects the share of time spent on learning relative to total productive time. Additional factors such as **mpce** (monthly per capita consumption expenditure), **household size (hhsize)**, and education level are included to assess how economic conditions and household characteristics shape youth time use patterns. This analysis aims to provide a comprehensive view of how India's young population divides its time among education, work, and other essential tasks, highlighting disparities across demographic and social lines.

Question 0: Preparatory

The summary Statistics of the variables used in the regression models segregated as male and females

Table-1: Summary Statistics

Variable	Category	Obs	Mean	Std.Dev.	Min	Max	t / X2	P Value
in_educ	Female	3134	0.2	0.4	0	1		
	Male	3504	0.27	0.45	0	1		
	Total	6638	0.24	0.43	0	1	45.82	0
age	Female	3134	22.11	4.15	15	29		
	Male	3504	21.8	4.23	15	29		
	Total	6638	21.95	4.19	15	29	-2.94	0
gender	Female	3134	2	0	2	2		
	Male	3504	1	0	1	1		
	Total	6638	1.47	0.5	1	2	6634	0
edu_level	Female	3134	3.07	1.15	1	5		
	Male	3504	3.26	0.98	1	5		
	Total	6638	3.17	1.07	1	5	165	0
marital_status	Female	3134	1.51	0.5	1	2		
	Male	3504	1.31	0.46	1	2		
	Total	6638	1.41	0.49	1	2	288.26	0
type_of_the_day	Female	3134	1.1	0.3	1	2		
	Male	3504	1.12	0.32	1	2		
	Total	6638	1.11	0.31	1	2	5.28	0.02
lnmpce	Female	3134	7.74	0.58	5.87	10.25		
	Male	3504	7.82	0.59	5.52	10.68		
	Total	6638	7.78	0.59	5.52	10.68	5.26	0
hhsize	Female	3134	5.26	2.04	1	15		
	Male	3504	4.81	1.99	1	15		
	Total	6638	5.02	2.03	1	15	-9.15	0
cook_fuel	Female	3134	1.29	0.47	1	3		
	Male	3504	1.31	0.49	1	3		
	Total	6638	1.3	0.48	1	3	3.24	0.19
lighting	Female	3134	1.04	0.19	1	2		
	Male	3504	1.03	0.18	1	2		
	Total	6638	1.04	0.19	1	2	0.46	0.49
dwelling	Female	3134	1.28	0.45	1	2		
	Male	3504	1.28	0.45	1	2		
	Total	6638	1.28	0.45	1	2	0.06	0.79

social_group	Female	3134	2.86	1.02	1	4		
	Male	3504	2.87	1.01	1	4		
	Total	6638	2.86	1.01	1	4	1.16	0.76
religion	Female	3134	1.12	0.4	1	4		
	Male	3504	1.1	0.36	1	4		
	Total	6638	1.11	0.38	1	4	7.02	0.07
Sector	Female	3134	1.46	0.5	1	2		
	Male	3504	1.49	0.5	1	2		
	Total	6638	1.47	0.5	1	2	6.99	0

Test the hypothesis for balancing the variables across males and females.

The hypothesis test for balancing variables across males and females shows that the p-values for type_of_day, cooking fuel, lighting, dwelling, social group, and religion are greater than 0.05, suggesting no significant differences between males and females for these variables. Conversely, the p-values for in_educ, age, gender, education level, marital status, lnmpce, income, and sector are less than 0.05, indicating significant disparity between males and females across these variables.

Question 1: Binary Choice Model

We use logit and Probit model for the Binary Choice Model because the predicted probability of the dependent variable in_educ lie between 0 and 1.

The Logit model uses standardized logistic distribution with mean 0 and variance $\frac{\Pi^2}{3}$.

The probability density function: $\lambda(\varepsilon) = \frac{\exp(\varepsilon)}{[1+\exp(\varepsilon)]^2}$

The cumulative density function: $\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1+\exp(\varepsilon)}$

The Probit model uses standardized normal distribution with mean 0 and variance 1.

The probability density function: $\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\varepsilon^2}{2})$

The cumulative density function: $\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$

Where ε is error term

The econometric equation for probit and logit model:

$$
\begin{aligned}
Pr(\text{in\_educ}) = \mathrm{F}[&\beta_0 + \beta_1(\text{age}) + \beta_2(\text{gender}) + \beta_3(\text{Edu\_level}) + \beta_4(\text{marital\_status}) \\
&+ \beta_5(\text{type\_of\_the\_day}) + \beta_6(\text{lnmpce}) + \beta_7(\text{hhsize}) + \beta_8(\text{cook\_fuel}) \\
&+ \beta_9(\text{lighting}) + \beta_{10}(\text{dwelling}) + \beta_{11}(\text{social\_group}) + \beta_{12}(\text{religion}) \\
&+ \beta_{13}(\text{sector})]
\end{aligned}
$$

Where,

in_educ = 1, if attending educational institute
=0, if not attending educational institute

Age = Age in years (discrete variable)

Gender= 1, if male (as reference group)
=2, if female

Edu_level = 1, if illiterate (as reference group)
= 2, if primary
= 3, if Middle & Secondary
= 4, if Higher Secondary
= 5, if graduation and above

marital_status = 1, if unmarried (as reference group)
= 2, if currently married 2

type_of_the_day = 1, if normal day (as reference group)
= 2, if other day

Lnmpce = natural logarithm of monthly per capita consumption expenditure (continuous variable)

Hhsize = Household size (as a discrete variable)

cook_fuel = 1, if modern method (as reference group)
= 2, if Conventional method
= 3, if none

Lighting = 1, if electricity (as reference group)
= 2, if others

Dwelling	= 1, if Pucca (as reference group)
	= 2, if others
social_group	= 1, if Scheduled Tribe (as reference group)
	= 2, if Scheduled Caste
	= 3, if Other Backward Class
	= 9, if others
religion	=1, if Hindu (as reference group)
	= 2, if Muslim
	= 3, if Christian
	= 9, if Others
Sector	= 1, if rural (as reference group)
	= 2, if urban

<This space is left blank intentionally>

(a) Estimates of probit and logit models

Table-2: Estimates of logit and Probit model

Variables	Logit		Probit		Ratio	
	β	z	β	z	β	z
Intercept	5.66***	6.12	2.82***	5.57	2.01	1.10
age	-0.65***	-27.60	-0.35***	-29.08	1.88	0.95
gender	-0.28***	-3.20	-0.15***	-3.10	1.83	1.03
edu_levelPrimary	0.89**	1.98	0.46**	1.99	1.94	1.00
edu_levelMiddle	2.12***	4.94	1.10***	5.03	1.92	0.98
edu_levelHigher_sec	3.98***	9.02	2.10***	9.27	1.90	0.97
edu_levelGrad	4.52***	9.86	2.34***	9.90	1.93	1.00
marital_status	-1.60***	-8.10	-0.76***	-8.29	2.11	0.98
type_of_the_day	1.04***	8.18	0.61***	8.69	1.70	0.94
lnmpce	0.46***	4.98	0.27***	5.15	1.71	0.97
hhsize	0.02	0.69	0.01	0.95	1.30	0.72
cook_fuel2	-0.42***	-3.62	-0.24***	-3.57	1.79	1.01
cook_fuel3	0.32	0.97	0.19	1.01	1.72	0.97
lighting	0.34	1.43	0.20	1.52	1.68	0.94
dwelling	-0.25**	-2.32	-0.15**	-2.52	1.62	0.92
social_group2	0.11	0.62	0.11	1.14	0.96	0.54
social_group3	-0.28**	-2.09	-0.15**	-2.04	1.81	1.02
social_group9	0.28*	1.94	0.17**	2.05	1.67	0.94
religion2	-0.40**	-2.43	-0.17*	-1.87	2.30	1.30
religion3	-0.17	-0.19	-0.06	-0.12	2.71	1.54
religion9	0.50	1.15	0.26	1.09	1.88	1.05
Sector	0.27**	2.56	0.15**	2.54	1.79	1.01
Log Likelihood	-1821.718	(df=22)	-1853.984	(df=22)		

****, **, * significant 1%, 5%, 10%*

The logit coefficients are expected to be approximately 1.7 times larger than those from the probit model; however, several variables deviate from this pattern. Specifically, age, gender, education level, marital status, household size, social group, and religion do not align with this proportional relationship. Furthermore, age, gender, and education level are statistically significant at the 5% level, highlighting their strong influence on the outcome variable.

(b) **Interpreting the coefficient of gender**

Probit model for discrete difference

"The probability of female in education is 15% (=2.82- 2.67) lesser than the male holding all the other variables at constant."

Logit model using odds ratio

"The odds of being in education is 0.52 (= $e^{-0.28 * 1}$*) times smaller for female than male holding all the variables at constant."*

(c) **Interpreting the coefficient of lnmpce**

Probit model for continuous variable

"For each additional increase in lnmpce, the predicted probabilities of being in education increases by 27%, holding all the other variables at constant."

Logit model using odds ratio

"For each additional unit increase in lnmpce, the odds of being in education increases by 58% (= $e^{0.46 * 1}$ *= 1.58) holding all the variables at constant."*

(d) Graph for predicted probability across age separately for males and females

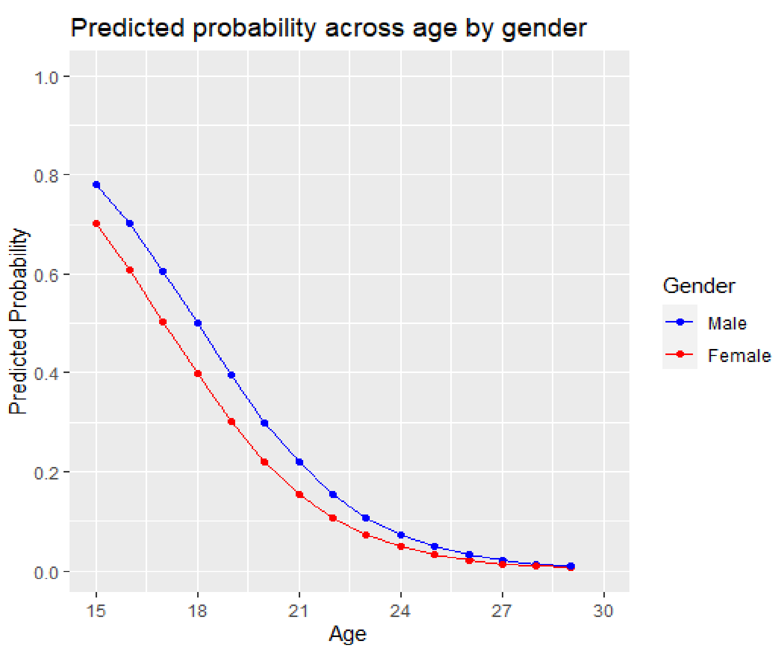

Figure-1: Predicted probability across age by gender

The predicted probability for the individuals who are in education against the age, satisfies our general conception of the fact that as age increases, the probability of being in an education is getting lower for the age group of 15-29. This signifies that the binary probability model with age and gender as the independent variables in decisions for the *in_educ*, that has predicted accurately aligning with the generalised assumption. The slope of the predicted probability shows a smooth downward sloping curve with no kinks or sudden shifts.

The predicted probability curve also states that the motive to pursue higher education is very low such that it shows only about 20% chances of higher studies after the age of 19, which was quite evident from the fact that only 1584 are *in_educ* out of the 6638 individuals.

However, there is disparities within gender, where female individuals have a predicted probability *in_educ* is only 0.7 compared to male with about 0.8 at the age of 15, showing a 10% difference in disparity among female *in_educ*. wherein there are about 954 males educated against 630 females educated in the dataset.

Question 2: Polychotomous Choice Model

The outcome of the dependent variable Activity are Education 1, Employment 2 and NEET 3. We estimate a multinomial logit model where the i^{th} individual chooses from one of the three options with the base category being employment. The probability of each option can be written as follows.

$$\Pr(Y = j) = \frac{e^{\beta_j x_i}}{1+\sum_{i=1}^{3} e^{\beta_k x_i}} \qquad \text{where j=1,2,3}$$

And

$$\Pr(Y = 0) = \frac{1}{1+\sum_{i=1}^{3} e^{\beta_k x_i}}$$

And we estimate the 2 log odds ratio

$$\ln\left[\frac{P_{ij}}{P_{i0}}\right] = \beta_j x_i$$

Motivation for this model based on additive random utility model

The Additive Random Utility Model (ARUM) provides a theoretical foundation for multinomial choice models by assuming that individuals select the option that maximizes their utility from a finite set of alternatives. In this model, utility for each option is composed of an observable component—reflecting known characteristics of the option—and an unobserved, random component that accounts for unknown factors. The multinomial logit model, a common ARUM, assumes that the random components X_i between options follow a logistic distribution, particular option is influenced by the relative values of z_i for each option which leads to probabilities of each choice based on relative utilities. These choice probabilities, often presented as Relative Risk Ratios (RRRs), express the likelihood of choosing one option over a baseline alternative. RRRs provide clear insight into how changes in explanatory variables (like income or cost) shift the probability of selecting one option relative to another, holding other factors constant. This relative comparison of utilities aligns with ARUM's foundation, as individuals are modelled to maximize utility based on observable attributes. Therefore, the use of RRRs in multinomial choice models is both practical and interpretable, allowing for a straightforward assessment of factors influencing choices within an ARUM framework.

$$\mathbf{U_i = z_i + X_i}$$

z_i is the observable component of utility, typically involving known characteristics of the choice

X_i is the random component of utility, representing unobserved factors that might influence the choice

(a) The multinomial logit model

$$\begin{aligned}
\Omega_{Z|Emp}(Activity) &= \mathrm{F}[\beta_{0,\mathrm{Z|Emp}} + \beta_{1,\mathrm{Z|Emp}}(\mathrm{age}) + \beta_{2,\mathrm{Z|Emp}}(\mathrm{gender}) + \beta_{3,\mathrm{Z|Emp}}(\mathrm{Edu_level}) \\
&+ \beta_{4,\mathrm{Z|Emp}}(\mathrm{marital_status}) + \beta_{5,\mathrm{Z|Emp}}(\mathrm{type_of_the_day}) + \beta_{6,\mathrm{Z|Emp}}(\mathrm{lnmpce}) \\
&+ \beta_{7,\mathrm{Z|Emp}}(\mathrm{hhsize}) + \beta_{8,\mathrm{Z|Emp}}(\mathrm{cook_fuel}) + \beta_{9,\mathrm{Z|Emp}}(\mathrm{lighting}) \\
&+ \beta_{10,\mathrm{Z|Emp}}(\mathrm{dwelling}) + \beta_{11,\mathrm{Z|Emp}}(\mathrm{social_group}) + \beta_{12,\mathrm{Z|Emp}}(\mathrm{religion}) \\
&+ \beta_{13,\mathrm{Z|Emp}}(\mathrm{sector})]
\end{aligned}$$

Where Z takes the activity as Education 1 or NEET 3 and Employment 2 activity is the base group.

Table-3: Estimates of Relative Risk Ratio for Education and NEET with Employment as base category

	Education		**NEET**	
	RRR	**SE**	**RRR**	**SE**
age	0.505***	0.024	0.921***	0.013
gender	2.961***	0.101	33.027***	0.09
edu_levelPrimary	2.465**	0.454	1.028	0.152
edu_levelMiddle	8.807***	0.434	1.188	0.147
edu_levelHigher_sec	53.971***	0.448	1.041	0.17
edu_levelGrad	76.252***	0.465	0.652**	0.18
marital_status	0.263***	0.204	1.781***	0.1
type_of_the_day	2.753***	0.136	0.887	0.138
lnmpce	1.581***	0.099	0.937	0.089
hhsize	1.035	0.027	1.040*	0.02
cook_fuel2	0.578***	0.122	0.708***	0.099
cook_fuel3	0.973	0.359	0.263**	0.669
lighting2	1.237	0.246	0.738	0.201
dwelling	0.781**	0.113	1.024	0.093
social_group2	1.384*	0.183	1.682***	0.152
social_group3	0.902	0.138	1.514***	0.113
social_group9	1.634***	0.153	1.667***	0.133
religion2	0.792	0.174	1.451***	0.134
religion3	0.398	0.896	0.207*	0.857
religion9	1.948	0.502	1.728	0.511
Sector	1.395***	0.113	1.189*	0.097
Constant	441.799***	0.975	0.399	0.777
Base category: employment, RRR- Relative Risk Ratio ***, **, * significant 1%, 5%, 10%				

(b) Interpreting the result for Gender

The results indicate that *“Female members are more likely to be involved in Education and NEET activities than in Employment”*.

This finding is supported by the Relative Risk Ratio (RRR) for gender (female), which is greater than 1 and the results are significant at 5%.

Interpreting the result for Edu_level

“Individuals with education levels of Primary, Middle & Secondary, Higher Secondary, and Graduation & above are more likely to be involved in Education activity rather than Employment”.

The Relative Risk Ratio (RRR) is greater than 1 for both Education and NEET activities. However, this variable is not statistically significant for NEET activity.

Interpreting the result for lnmpce

“Individuals with higher levels of lnmpce (Natural Log of Monthly Per Capita Consumption Expenditure) are more likely to engage in Education activities relative to Employment, while showing a comparatively greater involvement in Employment than in NEET activities”.

The Relative Risk Ratio (RRR) is greater than 1 for Education and less than 1 for NEET. Furthermore, lnmpce is statistically significant for Education but not significant for NEET activities at the 5% level.

Interpreting the result for cooking fuel (conventional methods)

“Individuals using conventional cooking methods are more likely to engage in Employment activities than in Education or NEET activities”.

The Relative Risk Ratio (RRR) is less than 1 for both Education and NEET and is statistically significant at the 5% level.

Interpreting the result for Sector

“Individuals in the Urban sector are more likely to engage in Education and NEET activities compared to Employment.”

The Relative Risk Ratio (RRR) is greater than 1 for both Education and NEET activities; however, the Urban sector is statistically significant for Education but not significant for NEET activities at the 5% level.

(c) IIA assumption in the context of this study

The Independence of Irrelevant Alternatives (IIA) assumption implies that the relative odds of a household selecting one activity (e.g., Education) over another (e.g., Employment) remain constant, even if new alternatives (such as NEET or additional categories within contribution types) are added to or removed from the choice set.
The model assumes that each activity choice (Education, Employment, NEET) is an independent and unaffected choice relative to the others.

Testing for IIA

Null Hypothesis (H_0): IIA assumption holds
(Removing NEET does not affect the choice probabilities between Education and Employment)

Alternate Hypothesis (H_1): IIA assumptions does not hold
(Removing NEET affects the choice probabilities between Education and Employment)

Hausman-McFadden Test:

$$\boxed{H_{IIA} = \left(\widehat{\beta_R} - \widehat{\beta_F}\right)^T \left[Var\left(\widehat{\beta_R}\right) - Var\left(\widehat{\beta_F}\right)\right]^{-1} \left(\widehat{\beta_R} - \widehat{\beta_F}\right)}$$

Asymptotically distributed as chi-square (8.4628) with degrees of freedom as 22

P-value: 0.9957

Since P-value > 0.05, Alternate hypothesis is satisfied.

<u>IIA assumptions does not hold.</u>

(d) Testing for the choice on Education activity and NEET activity be combined

Null Hypothesis (H_0): The coefficients for Education and NEET are equal for all independent variables. The choices are indistinguishable with respect to variables in the model.

Alternate Hypothesis (H_1): The coefficients for Education and NEET are not equal for all independent variables. The choices are distinguishable with respect to variables in the model.

Likelihood Ratio test:

The chi-square distribution is 1413.2 with degree of freedom as 22 with **P-value 0.000**.

Therefore, the hypothesis that Education activity and NEET activity are indistinguishable with respect to the variable in the model is rejected at 5% level of significance.

The coefficients for Education and NEET are not equal for all independent variable.

(e) Test that there is no difference in religion categories across activity choices.

Null Hypothesis (H_0): The coefficients for each religion category are equal across all activity
(Religion does not impact the likelihood of choosing Education, Employment or NEET)

Alternate Hypothesis (H_1): The coefficients for each religion category differ across activity
(Religion does impact the likelihood of choosing Education, Employment or NEET)

LR test:

The chi-square distribution is 6382.5 with degree of freedom as (40,2) with **P-value 0.000**

Therefore, the hypothesis that the coefficient for each religion category are equal across all the activity is rejected.

Religion does impact the likelihood of choosing Education, Employment or NEET

Question 3: TOBIT Model

(a) **Histogram of the three dependent variables**

(i) time_learn

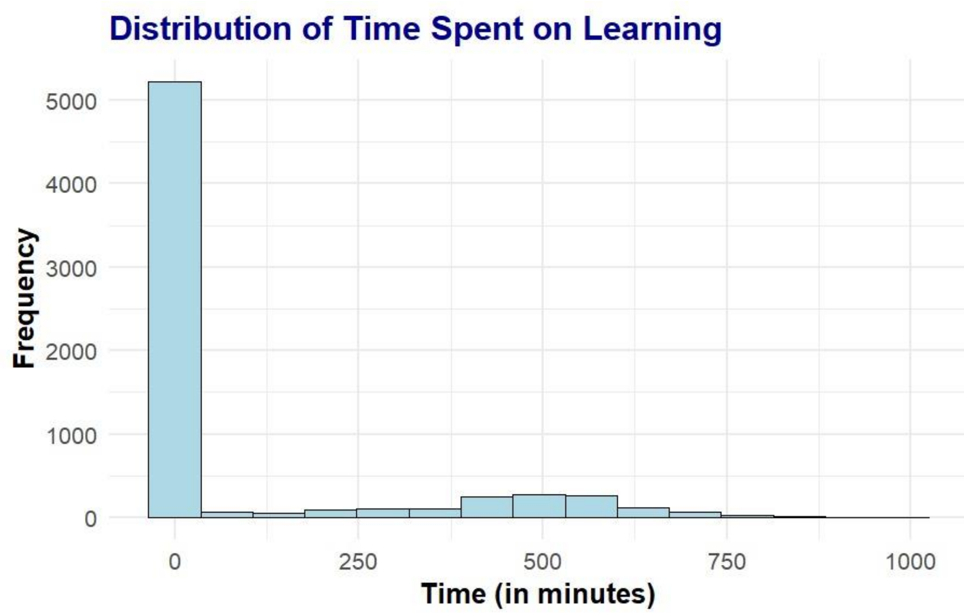

Figure-2: Distribution of Time spending on Learning

(ii) log (time_learn)

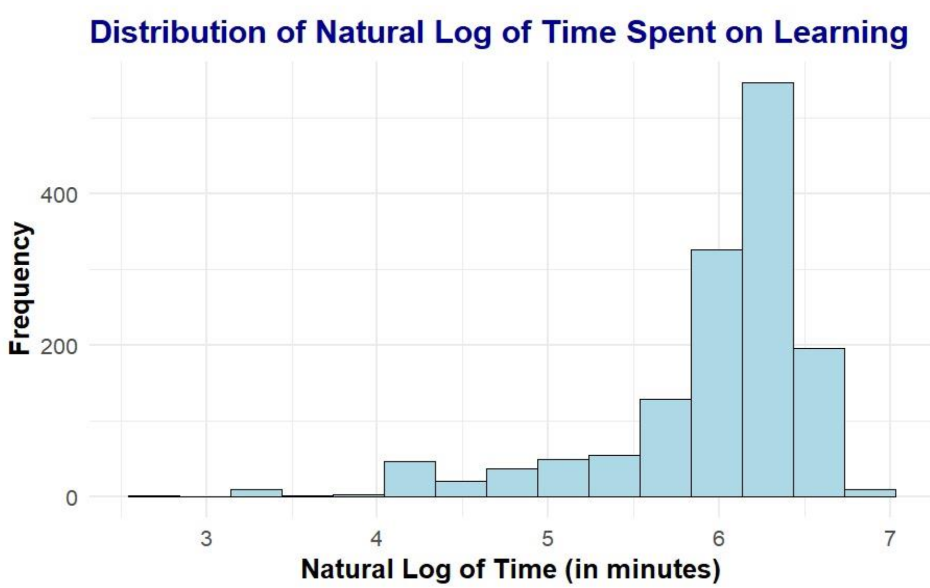

Figure-3: Distribution of natural log of Time spending on Learning

(iii) sh_learn_time

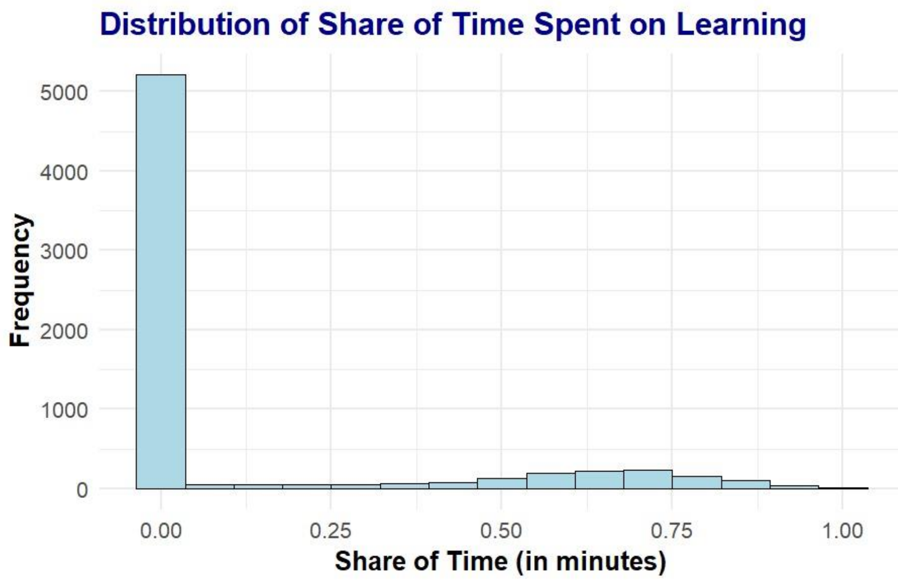

Figure-4: Distribution of Share of Time spending on Learning

The continuous variable of the time spent on learning in minutes shows that the data contains 5214 non-learners out of the 6638, estimating only about 1424 of them spend time on learning which range from 15 minutes to 990 minutes, with the <u>average learning time of 444.4 minutes</u>. The distribution of time spent of learners is not legible due to a <u>lot of non-learners</u> in the dataset, which means there are <u>major outliers</u> to the model in terms of time spent on learning, which arises the need for finding the distribution of learning time of learners that can be addressed by taking the log of the time spent on learning. The <u>natural log of time shows that the time spent on learning of the 1424 learners explains a negatively skewed distribution with the median greater than the mean</u>, reasoning the fact that there can be learners who either spend the major part of the day learning can be those who are in school compared to the less time spent by those who don't go to school and additionally, the learners and non-learners from the data in terms of the proportion of the day spent on learning is very low as there are only 1424 learners in the data.

(b) Results for the three different dependent variables

Table-4: Tobit model with three different dependent variables

Variable	Model-1		Model-2		Model-3	
	time_learn		lntime_learn		sh_learn_time	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
age	-21.405***	2.82	-0.240***	0.03	-0.028***	0.00
edu_levelPrimary	72.56	71.82	0.93	0.81	0.08	0.09
edu_levelMiddle	215.523**	68.26	215.524***	0.77	2.409***	0.09
edu_levelHigher_Sec	295.809***	69.14	3.626***	0.78	0.368***	0.09
edu_levelGrad	370.371***	70.67	4.701***	0.80	0.465***	0.09
marital_status	-95.034***	23.45	-1.320***	0.28	-0.124***	0.03
type_of_the_day	-264.199***	17.04	-2.292***	0.20	-0.321***	0.02
lnmpce	59.794***	12.25	0.517***	0.15	0.069***	0.02
hhsize	-4.70	3.46	-0.03	0.04	0.004	0.00
cook_fuel2	30.933*	17.40	0.29	0.21	0.046**	0.02
cook_fuel3	141.637***	39.68	1.527***	0.49	0.171***	0.05
lighting	47.94	32.20	0.706*	0.38	0.070*	0.04
dwelling	-29.337*	15.98	-0.25	0.19	-0.03	0.02
social_group2	-61.790**	24.69	-0.542*	0.30	-0.074**	0.03
social_group3	-34.050*	18.71	-0.20	0.23	-0.04	0.02
social_group9	-32.954*	19.98	-0.13	0.24	-0.03	0.03
religion2	-8.44	23.25	0.06	0.28	-0.01	0.03
religion3	-72.41	115.34	0.10	1.36	-0.10	0.15
religion9	-81.49	55.88	-1.02	0.68	-0.09	0.07
activityEdu	835.001***	21.92	10.187***	0.26	1.091***	0.03
activityNEET	177.987***	22.98	2.139***	0.27	0.234***	0.03
genSecM:U	23.53	18.27	0.404*	0.22	0.04	0.02
genSecF:R	-40.748**	15.86	-0.376*	0.19	-0.058***	0.02
genSecF:U	-33.18	20.32	-0.26	0.25	-0.047*	0.03
logSigma	5.579***	0.02	1.182***	0.02	-1.069***	0.02
Constant	-707.099***	139.44	-7.486***	1.65	-0.858***	0.18
Observations	6638		6638		6637	
***, **, * significant 1%, 5%, 10%						

Interpreting the coefficients across the three models

(i) **Gender-sector interactions**

The results indicate that females in rural areas allocate less time to learning activities compared to both males and females in urban areas. The female in rural variable is statistically significant in Model 1 and Model 2 at the 5% level, while in Model 3 it is significant at the 10% level.

Similarly, female in urban also spend less time on learning activities relative to urban males. This effect is reflected across all three models; however, the female in urban variable is statistically insignificant in Model 1 and Model 2, with significance at 10% level in Model 3. This suggests that there is minor variation in the proportion of time allocated to learning among urban females.

In addition, male in urban individuals tend to spend more time in learning activities, with this variable reaching statistical significance at the 10% level in Model 2.

(ii) Age

Model-1: For increase in age by a year reduces the learning time by 21.4 minutes holding all the variables at constant.

Model-2: As the age increases a year, the learning time decreases by 0.24% holding all the variables at constant.

Model-3: For the increase in age by a year reduces the share of learning time by 0.028 minutes holding all the variables at constant.

The coefficient of age is statistically significant at 5% level across all the three models.

(iii) Activity

Individual who are involved in Education and NEET activity tend to spend more time in learning than those involved in employment which has been supported by all the three model and statistically significant at 5% level.

Model-1: An individual who is involved in Education activity spends 835 minutes more in learning compared to the individual with employment activity.

Model-2: An individual who is involved in education activity spends 10% time more in learning compared to the individual with employment activity.

Model-3: An individual who is involved in education activity spends 1.09 minutes more of their share time for learning.

Individuals engaged in NEET activity allocate more time to learning compared to those involved in employment. However, the coefficient for NEET is lower than that for the Education activity, suggesting that while NEET individuals dedicate more time to learning than those employed, they still spend less time learning than those engaged in Education activity.

(iv) lnmpce

Model-1: Each unit increase in lnmpce leads to an increase of 59.8 minutes in learning time holding all the variables at constant.

Model-2: Increases in lnmpce are associated with a 0.517% in learning time holding all the variables at constant.

Model-3: The share of learning time increases by 0.069 minutes with each unit increase in lnmpce holding all the variables at constant.

The lnmpce variable is significant across all the three models.

Most appropriate dependent variable for such an analysis

In Model 1, the dependent variable time_learn captures individuals' total learning time, but it is highly influenced by zero values, resulting in a positively skewed distribution. Similarly, in Model 3, the dependent variable sh_learn_time, which represents the share of time dedicated to learning, is also positively skewed due to the prevalence of zeros. In contrast, Model 2 employs the dependent variable lntime_learn (logarithmic transformation of learning time), which is less skewed than the other two models.

Therefore, <u>Model 2 provides a comparatively better fit, making lntime_learn the most suitable dependent variable for this analysis</u>.

(c) Method of estimation that will address the endogeneity on the activity dummy variables

In this model, the activity variable becomes endogenous, as it may be correlated with unobserved factors that influence time spent learning, leading to biased and inconsistent Tobit model estimates. To address this endogeneity, we apply the <u>Instrumental Variable (IV)</u> method within the Tobit model.

First, we estimate the endogenous activity dummy variables using a multinomial logistic regression, where activity is regressed on selected instrumental variables. The predicted values from this regression are then replaced with the endogenous activity variable in the Tobit model, resulting in consistent and reliable estimates.

Total observations:	6638	Left-censored:	5214
Uncensored:	1424	Right-censored:	0

Table-5: Estimate of IV within Tobit model

Variables	Estimate	SE	t value
Intercept	777.99***	183.54	4.24
activity_pred	-259.76***	73.63	-3.53
age	-100.105***	6.59	-15.20
gender	-140.25***	28.57	-4.91
edu_levelPrimary	116.76	91.36	1.28
edu_levelMiddle	415.46***	87.50	4.75
edu_levelHigher_sec	729.399***	92.15	7.92
edu_levelGrad	881.10***	94.92	9.28
marital_status	-293.15***	33.02	-8.88
type_of_the_day	-158.05***	26.30	-6.01
lnmpce	94.49***	18.50	5.11
hhsize	-3.22	5.06	-0.64
cook_fuel2	-18.31	25.13	-0.73
cook_fuel3	190.89***	60.43	3.16
lighting	109.17**	45.77	2.39
dwelling	-57.50**	22.75	-2.53
social_group2	-49.94	35.51	-1.41
social_group3	-70.28***	26.92	-2.61
social_group9	-7.38	29.52	-0.25
religion2	-60.68*	33.64	-1.80
religion3	-41.47	169.93	-0.24
religion9	-26.60	83.28	-0.32
Sector	10.99	21.98	0.50
logSigma	6.057***	0.02	281.17
***, **, * significant 1%, 5%, 10%			

Log-likelihood: -11846.66 on 24 Df

Using IV estimation methods will help ensure that the Tobit model's coefficients for activity are unbiased and consistent, thus providing more reliable results.

End of Assignment