IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A STATISTICAL STUDY OF SMOKING PATTERNS IN INDIA

Dr. James Kurian

Associate Professor, Department of Statistics, Maharaja's College Ernakulam, Kerala, India

Abstract: This study try to bring out some of the salient features of the smoking patterns in India with the help of multinomial logistic regression model. The study showed that there is a positive association between smoking and age. Another factor that influences the smoking habit is the average number of years spent in school, as the study showed that the chance of being heavy smoker or light smoker decreases as the average number of years spent in school increases. One of the important implications of this result was that, educating people about dangers of smoking can have significant positive impact on, people convert from smoking to not smoking at all. Study also established few interesting association between smoking habits and the employment status.

Index Terms - ***Multinomial logistic regression, Odds, Relative risk, Significance.***

1. Introduction

Increased use of tobaccos in different forms which accuses of causing significant increase in cancers, respiratory diseases, and many other chronic diseases possess one of the biggest concerns of the global health care system particularly in India. A study of the smoking habits in a country is very important due to health, financial, social aspects as well. As per WHO (2008) report, tobacco usage increases risk factors for a number of chronic diseases, including cancer, lung diseases, and cardiovascular diseases and many others. Worldwide, it is estimated that over 4.9 million people will die prematurely due to tobacco-related diseases (WHO, 2002). In the same report it says that Tobacco use is a major public health problem across the world and the global death toll from tobacco consumption is estimated to reach 10 million per year by the end of 2020. Alberg A. J. et al. (2003), Zalata A et al. (2007), Lash T. et al. (1999), *Vainio H. (June 1987),* Patton G.C. et al. (1998) *etc*, *conducted* studies on the adverse impact of tobacco consumption. Siahpush M (2007) et al., Kishore J.(2014), Levitt SD, et al. (2000) discusses the financial aspects of tobacco usage in great details. In India around 275 million people either smoked or chewed tobacco and are the second largest producer and user of tobacco products after China (Schwartz R. L. et al. (2011)). There are many studies available regarding the influence, effects, patterns and other aspects of tobacco consumption in India. For example, see John, RM (2005), Mistry, R et al. (2017), Mini, GK et al. (2014) etc.

2. GATS survey

Present study is based on the data provided by Global Adult Tobacco Survey (GATS), a component of Global Tobacco Surveillance System (GTSS), a global standard for systematically monitoring adult tobacco use and tracking key tobacco control indicators. GATS is a nationally representative household survey of adults 15 years of age or older using a standard core questionnaire, sample design, and data collection and management procedures that were reviewed and approved by international experts. GATS is intended to enhance the capacity of countries to design, implement and evaluate tobacco control interventions. GATS conducted a household survey of persons 15 years of age or older conducted in all 30 states of India and two Union Territories. The first round of GATS was conducted between June 2009 and January 2010. The second round of GATS survey was conducted, by Tata Institute of Social Sciences, Mumbai, for the Ministry of Health & Family Welfare, Government of India. They used a multi-stage sample design for both rounds of GATS. From each of the sampled household, one household member 15 years of age or older was randomly selected for individual interview. In the first round 69,296 individual interviews were completed with an overall response rate of 91.8%. In the second round, a total of 74,037 individual interviews were completed with an overall response rate of 92.9%.

3. Objective of this study

Objective is to model the tobacco smoking patterns in India using multinomial logistic regression model. Other objectives are; (a) to find out whether work status of an individual is associated with a specific smoking pattern, (b) to find out whether age is associated with a specific smoking pattern and (c) to find out whether education level is associated with a specific smoking pattern.

4. Variables of the study

Y – Tobacco smoking patterns (Y = 0, Never smoked, Y = 1, Less than one pkt. daily, Y = 2, Heavy smokers)

X_1 - Individual's age (in years)

X_2 – Education (Average number of years spent in school, in years)

X_3 –Work Status, categorized as; 1. Government and Non- Government, 2. Self- employed, 3. Student and, 4. Unemployed

5. Logistic and multinomial logistic regression models

Binary Logistic Regression is a special type of regression where *binary response variable* is related to a set of explanatory variables, which can be discrete and/or continuous. Multinomial logistic regression is a simple extension of binary logistic regression that allows for more than two categories of the dependent variable. Multinomial logistic regression is used to predict category membership on a dependent variable based on multiple independent variables. The independent variables can be either dichotomous or continuous. Like binary logistic regression, multinomial logistic regression uses maximum likelihood estimation to evaluate the probability of categorical membership.

Logistic analyses for binary outcomes attempt to model the odds of an event's occurrence and to estimate the effects of independent variables on these odds. The odds for an event is a quotient that compares the probability that an event occurs ("success") to the probability that it does not occur ("failure,"). When the probability of success is greater than the probability of failure, the odds are greater than one; if the two outcomes

are equally likely, the odds are unity; and if the probability of success is less than the probability of failure, the odds are less than one

$$\Pi(X) = \text{Odds} = \frac{P(Y=1|X)}{P(Y=0|X)} = \frac{P(Y=1|X)}{1-P(Y=1|X)}$$

Let us assume that there are k independent (X) variables and Y be dependent variable with more than 2 possible values. Particularly, we consider the case with three outcomes, coded as 0, 1 and 2. In the three outcome category model we need two logit functions. We have to decide which outcome categories to compare. The obvious extension is to use Y = 0 as referent or baseline outcome and to form logits comparing Y = 1 and Y = 2 to it.

We define the two logit functions as follows.

$$g_1(X) = \ln\left[\frac{P(Y=1|X)}{P(Y=0|X)}\right]$$
$$= b_{10} + b_{11}X_1 + b_{12}X_2 + \ldots.. + b_{1k}X_k$$
$$= X'b_1$$
$$g_2(X) = \ln\left[\frac{P(Y=2|X)}{P(Y=0|X)}\right]$$
$$= b_{20} + b_{21}X_1 + b_{22}X_2 + \ldots.. + b_{2k}X_k$$
$$= X'b_2$$

It follows that the conditional probabilities of each outcome category given the covariance vector are,

$$P(Y=0 \mid X) = \frac{1}{1+e^{g_1(X)}+e^{g_2(X)}}$$
$$P(Y=1 \mid X) = \frac{e^{g_1(X)}}{1+e^{g_1(X)}+e^{g_2(X)}}$$
$$P(Y=2 \mid X) = \frac{e^{g_2(X)}}{1+e^{g_1(X)}+e^{g_2(X)}}$$

Let $\pi_j(X) = P(Y=j|X)$; for j = 0,1,2. Each probability is a function of the vector of 2(k+1) parameters $b' = (b_1', b_2')$. A general expression for the conditional probability in the three category model is

$$P(Y=j|X) = \frac{e^{g_j(X)}}{\sum_{k=0}^{2} e^{g_j(X)}}$$, where the vector $b_0 = 0$ and $g_0(X) = 0$.

The maximum likelihood estimator $\hat{b}$ is obtained by setting above equations equal to zero and solving for b. The solution requires the same type of iterative computation.

Data from the GATS 2009 survey was analyzed to determine what factors contributed to a person being a heavy smoker, light smoker or a non-smoker at all depending on their work status, education level and their age. Multinomial logistic regression was used to estimate association between relevant prediction variables work status, number of years in school and age with tobacco smoking patterns. Odds ratio associated with each outcome were reported and interpreted.

6. Model details

Table -1 ***Model Fitting Information***

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2Log Likelihood	Chi-Square	df	Sig.
Intercept Only	16969.968			
Final	10483.332	6486.636	10	0.000

Table -2 ***Pseudo R-Square***

Cox and Snell	.089
Nagelkerke	.136
McFadden	.087

From the model fitting information, log-likelihood value of the model with the intercept only (null model) is 16969.968 while the full model with age, education and work status is 10483.332. This reduction shows that the model is better at predicting a person's smoking status using work status, education and age. The chi-square value 6486.639 is statistically significant at 0.05 which shows that the overall model is predicting someone's smoking status better than a model with intercept only. Therefore work status, education and age contributed significantly to fit of the model. The R-Squared values, McFadden, Cox-Snell/ML and Cragg-Uhler/Nagelkerke are treated as measure of the effect of size, and values of these statistics in Table -2 suggest a moderate effect size for this model.

Table -3: ***Likelihood Ratio Tests***

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	10483.332[a]	0	0	.
AGE	11700	1216.66	2	0
Education	10670.3	187.014	2	0
Job	15327.3	4843.95	6	0

Table -3 shows that the Chi- square value for variable age is 1216.66, which is significant (P value =0) at 5% level of significance, suggesting that the variable age in the model is significant. The Chi- square value for variable education is 187.014, which is significant (P value =0) at 5% level of significance. That is, the variable education is a significant predictor of smoking status. The Chi-square value for variable job is 4843.95, which is also significant (P value =0) at 5% level of significance. Thus addition of variable job is significantly contributes to the model.

Table-4: ***Multinomial logistic regression model results***

B01. Do you *currently* smoke tobacco on a pkg. daily basis, less than daily, or not at all?[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
heavy	Intercept	-3.863	0.05	5994.596	1	0			
	AGE	0.029	0.001	1219.042	1	0	1.029	1.028	1.031
	Education	-0.029	0.002	162.498	1	0	0.971	0.967	0.975
	JOB								
	Govt & NGO	1.479	0.034	1872.088	1	0	4.387	4.103	4.691
	Self-employee	1.719	0.031	3045.13	1	0	5.581	5.25	5.932
	Student	-0.016	0.09	0.033	1	0.856	0.984	0.824	1.175
	Unemployed	0[b]			0				
moderate	Intercept	-4.506	0.087	2683.131	1	0			
	AGE	0.011	0.002	44.131	1	0	1.011	1.008	1.014
	Education	0	0.003	0.015	1	0.904	1	0.993	1.006
	JOB								
	Govt & NGO	1.389	0.061	514.969	1	0	4.01	3.557	4.521
	Self-employee	1.448	0.058	613.399	1	0	4.253	3.793	4.769
	Student	0.519	0.109	22.522	1	0	1.68	1.356	2.082
	Unemployed	0[b]			0				

a. The reference category is: NOT AT ALL
b. This parameter is set to zero because it is redundant.

7. Interpretations of the results

Influence of age on smoking pattern- The model coefficients in table-4 shows that a one unit increase in variable age, is associated with an increase in log odds of being a heavy smoker by the amount 0.029 than being a non-smoker at all and that of being a light smoker by the amount 0.011 than being a non-smoker. This means that as age increases, there is a very high chance of someone becoming a heavy smoker or a light smoker. The relative risk ratio for a one-unit increase in the variable age is 1.029 for being heavy smoker vs. non-smoker at all, while the relative risk ratio for a one unit increase in variable age is 1.011 for being a light smoker vs non-smoker at all. The result for heavy smoker and light smoker are significant at 5% level of significance. But the results shows that there is a higher risk of converting from non-smoker to a heavy smoker than there is in converting from non smoker to light smoker as age increases.

Influence of education on smoking pattern-From Table-4, we can observe that, one unit increase in the variable education is associated with a decrease in log odds of being a heavy smoker by the amount 0.029 than being a non-smoker at all and that of being a light smoker by the amount 0.000 than being non-smoker at all. However,

the p-value (0.000) associated with heavy smoker is highly significant but the p-value (0.904) associated with light smoker isn't significant at 5% level. This means that education is not a good predictor of an individual being a light smoker. But as education increases, there is a very low chance of someone becoming a heavy smoker rather than becoming a light smoker. The relative risk ratio for a one-unit decrease in the variable education is 0.971 for being heavy smoker vs. non-smoker at all, while the relative risk ratio for a one unit decrease in variable education is 1.0 for being a light smoker vs non-smoker at all. The result shows that there is a lower risk of converting from non-smoker to a heavy smoker than converting from non-smoker to light smoker as education increases.

Influence of job status on smoking pattern-The log odds of being a daily smoker vs not being a smoker at all will increase by 1.479 if moving from unemployment to being employed by the government or the NGO's. On the other hand the log odds of being a less than daily smoker vs not being a smoker at all will increase by 1.389 if moving from unemployed to being employed by the government or the NGO. The p-value for heavy smoker and light smoker was 0.000 and 0.000 and hence these findings were significant at 5% level of significance. This shows that when someone is employed by the government or the NGO's they have almost equal chances of either being daily or light smokers, but the higher chance is that they may end up being light smokers. The relative risk ratio switching from, unemployed to being employed by the government or NGO is 4.387 for being a heavy smoker than being a non smoker at all. In other words, the expected risk of being a heavy smoker is high for the people who are employed by the government or NGO. Also, the relative risk ratio for switching from unemployed to being employed by the government or NGO is 4.01 for being a light smoker than being a non smoker at all. In other words, the expected risk of being a non smoker is higher for the people who are employed by the government and NGO. This might be due to the fact that most organization, both private and government, have adopted the anti-smoking policy in their work places, discouraging the smoking habit

Influence of Self-employment job status on smoking pattern-The log odds of being a daily smoker vs not being a smoker at all will increase by 1.719 if moving from unemployment to being self-employed while the log odds of being a less than daily smoker vs not being a smoker at all will increase by 1.448 if moving from unemployment to being self-employed. The p-value associated with heavy smoker and light smoker was 0.000 and 0.000 respectively. Therefore the findings were significant at 5% level of significance. The relative risk ratio switching from unemployed to self-employed is 5.581 for being a heavy smoker than being a non smoker at all. In other words, the expected risk of being a heavy smoker is high for the people who are self-employed. Also, the relative risk ratio switching from unemployed to being self-employed is 4.253 for being a light smoker than being a non smoker at all. In other words, the expected risk of being a light smoker is high for the people who are self-employed. However the higher chance is for this people becoming heavy smokers than light smokers. This habit could be explained by the reason that since most self-employees don't has rules against smoking, a person can smoke at any given time in their work place. Comparing self-employment with formal employment (government and NGO/private) there is a higher chance of heavy smokers in self-employment than in government employment.

Influence of student status on smoking pattern-The log odds of being a daily smoker vs. not being a smoker at all will decrease by 0.016 if moving from unemployment to being student while the log odds of being a less than daily smoker vs. not being a smoker at all will increase by 0.519 if moving from unemployment to being student.

However, the p-value (0.856) associated with heavy smoker isn't significant but the p-value (0.000) associated with light smoker is significant at 0.05. This means that student status is not a good predictor of an individual being a heavy smoker. The relative risk ratio switching from, unemployed to student is 0.984 for being a heavy smoker than being a non smoker at all. In other words, the expected risk of being a heavy smoker is low for the people who are student. Also, the relative risk ratio switching from unemployed to being student is 1.680 for being a light smoker than being a non smoker at all. In other words, the expected risk of being a light smoker is high for the people who are student. However the lower chance is for this people becoming heavy smokers compared to light smokers.

8. Conclusion

The study found that there is a positive association between smoking and age. That is, a person grows old, the more they are likely to be smoker, either heavy a smoker or a light smoker. Study also found that the average number of years spent in school also is a good predictor of smoking patetrn, as the chance of being heavy smoker or light smoker decreases as the average number of years spent in school increases. It is also established that a person is employed or not can also associated with their smoking pattern. Another result found in this study is that, employment status is also associated with a higher chance of following a particular smoking status. Finally from the results, we have more people converting from not being smokers at all to being either heavy or light smokers. Therefore intervention of educating people as they grow, about dangers of smoking, should always be emphasized so that we can convert more people from smoking to not smoking at all status. Awareness programmes regarding deadly effects of smoking should be conducted for people with lower level of education. Anti-smoking campaigns need to be emphasized more on self-employed as they are more likely to shift to heavy smokers category.

REFERENCES

[1] Alberg AJ, Diette GB, Ford J.G. (2003). Attendance and absence as markers of health status—The example of active and passive cigarette smoking. *American Journal of Epidemiology*.157(10).

[2] Frank E. Harrell, Jr.(2015). Regression *Modeling Strategies With Applications to Linear Models,Logistic and Ordinal Regression, and Survival Analysis*. Second Edition Springer Series in Statistics.

[3] GATS (2009). Global Adult Tobacco Survey. https://www.who.int/southeastasia

John, RM (2005). Tobacco consumption patterns and its health implications in India. *Health policy*. 71(2)

[4] Kishore J.(2014). National health programs of India: *National Policies and legislation related to health. 11th Edn. 2014*. New Delhi: Century Publications

[5] Lash T, Aschengrau A(1999). Active and passive cigarette smoking and the occurence of breast cancer. *American Journal of Epidemiology*

[6] Levitt SD, Venkatesh SA (2000). An economic analysis of a drug-selling gang's finances. *Quantiative Journal of Economics* 115:755–789

[7] Mini, GK, Sarma, PS, Thankappan, KR.(2014). Patterns of tobacco use and its co-relates among older adults in India. *Asian Pac J Cancer Prev*. 15

[8] Mistry, R, Dasika, A. (2017). Antenatal tobacco use and secondhand smoke exposure in the home in India. *Nicotine Tob Res*.

[9] Patton GC et al. Depression (1998). anxiety, and smoking initiation: a prospective study over 3 years. *American Journal of Public Health.*

[10] Siahpush M, Spittal M, Singh G.K. (2007) Smoking cessation and financial stress. *Journal Public Health*. ;29(4):338-42

[11] Samwel N. M., Anthony K. W., Anthony G. W. (2015) . Analysis of Tobacco Smoking Patterns in Kenya Using the Multinomial Logit Model. *American Journal of Theoretical and Applied Statistics.* Vol. 4, No. 3.

[12] Schwartz RL, Wipfli HL, Samet JM (2011). World no tobacco day 2011: India's progress in implementing the framework convention on tobacco control. *Indian Journal Medical Res*.

[13] Vainio H (June 1987). *"Is passive smoking increasing cancer risk?". Scandinavian Journal of Work, Environment & Health.* **13** (3).