

Canonical Correlation Analysis to Study the Impacts of Different Social Factors on Awareness of Health Hazard of Tobacco Smoking and Smoking Habit



Bhuyan KC*¹ and Urmi AF²

¹Professor, Department of Mathematics, Bangladesh

²Lecturer in Statistics, Department of Mathematics, Bangladesh

Received:  October 24, 2018; **Published:**  November 06, 2018

***Corresponding author:** KC Bhuyan, Professor (retired) Dhaka, Bangladesh

Abstract

The present analysis is done using the data collected from 1012 students of three universities, where students are investigated according to the convenience sampling plan. Most of the students (88.3%) are highly aware of the problem of health hazard of smoking. Still a good number of students (32.9%) are prone to smoking. Smoking habit is prevailed in higher rate among aged students. Awareness of health hazard of smoking and smoking habit is associated, and these two characters are associated with different socioeconomic background of the students. Thus, canonical correlation analysis is performed to study the complex relationship of awareness and smoking habit with other socioeconomic variables. The analysis indicates that important variables for complex relationship of awareness and smoking habit are sex and marital status.

Introduction

Tobacco in smoked form is consumed around the world and due to this a serious health threat is posed throughout the world. In a study Cohen [1] has reported that smoking is increasingly prevalent habit in Bangladesh, particularly among males. According to Global Tobacco Survey [2] 60% tobacco users consume only smokeless tobacco. Though tobacco smoking remains the leading preventable cause of death throughout the world, still global projected tobacco induced death at over 6 million annually [3]. However, by anti-smoking campaigns and programs a considerable success has been achieved to prevent the disease [4]. Tobacco use is a global epidemic among young people. As with adults, it poses a serious health threat to youth and young adults. Most young smokers became adult smokers and 50% of adult smokers die prematurely from tobacco related disease [5]. Thus, the health care provider need ways and means to prevent death among smokers.

The barriers in the implementation of policy related to tobacco control are education and awareness among consumers. Knowledge of health effects of smoking is an important factor in predicting smoking related behavior, including lower likelihood of initiation and greater likelihood of quitting [3,6-9]. Khatun and Bhuyan [10] and Bhuyan et al. [11] observed that among the university students awareness is increasing and highly aware students are less likely

to smoke. Again, awareness and smoking habit are associated with some socioeconomic factors. Thus, we are interested to study the joint relationship of smoking habit and awareness with other socioeconomic characteristics. This type of analysis is possible if there are one dependent set of variables and one independent set of variables. Such analysis is known as canonical correlation analysis [12-14]. In this paper canonical correlation analysis is done to study the complex relationship of smoking habit and awareness of health hazard of smoking with some of the socioeconomic background factors of the respondents.

Methodology

For canonical correlation analysis the criterion set of variables are awareness (y_2) and smoking habit (y_1) [Y-set] and the variables age(x_1), sex(x_2), marital status(x_3), religion(x_4), education of father (x_5), education of mother(x_6), occupation of father(x_7), occupation of mother(x_8), family income (x_9) are used as predictor sets(X-set). All the variables are measured in nominal scale for the analysis purpose. The awareness of health hazard of students [10,14] has been studied on the basis of nominal scale of 20 questions each of which has closed answers like 'True', 'False', 'Don't know'. The alternative answers toward the knowledge of awareness is assigned '3' followed by '2' with less awareness and '1' is answered

to the awareness which is not affirmative to the awareness. The maximum of the sum of the assigned values toward awareness is 60 and the minimum is 20. These values are different for different respondents. According to the sum of the assigned values in favors of awareness, the respondents are classified into 3 classes viz.

- a. Low in awareness (sum of the assigned values <30),
- b. Medium in awareness (sum of the assigned value is 30-40) and
- c. High in awareness (sum 40+).

Let R_{xx} , R_{yy} and R_{xy} be the sample correlation matrices of the variables in X-set, Y-set and in both X-set and Y-set, respectively. According to the objective of the study, it is needed to find $Y^* = \hat{b}Y$ and $X^* = \hat{a}X$, two liner combinations of the variables in Y-set and X-set, respectively, so that the simple correlation coefficient of X^* and Y^* becomes maximum, where a and b are eigen values of the characteristics equations

$$\begin{aligned} (R_{xx} - 1R_{xy} - 1R_{yx} - \lambda I)a &= 0 \\ (R_{yy} - 1R_{yx}R_{xx} - 1R_{xy} - \lambda I)b &= 0 \end{aligned}$$

Here the elements in a and b are the canonical weights, the magnitude of which indicates the importance of the variables in X-set and Y-set, respectively to show the maximum correlation between the variables in both sets. The canonical correlation analysis is fruitful if the variables in X-set and Y-set are significantly correlated. This can be done by the test statistics $X^2 = -(n-1) - 1/2(p=q=1)ln\Delta$ $\Delta = \prod_{j=M'+1}^M (1 - \lambda_j)$, $M' < M$ are the eigen values of characteristic equations given above; p and q are the number of variables in Y-set, (p=2) and X-set (q=9), respectively. The number of λ_j is $M = \min(p, q)$. This X^2 has pq d.f. The rejection of $H_0: \sum_{xy} = 0$ against $H_1: \sum_{xy} \neq 0$ by the above χ^2 -test statistic justifies the fruitful canonical correlation analysis. From the analysis, number of canonical variable of pairs is min (p, q). But all the pairs may not be statistically significant. The significance of j-th canonical variate pair is tested by the statistic.

$$\begin{aligned} x^2 &= -(n-1) - 1/2(p=q=1)ln\Delta^* \text{, where} \\ \Delta^* &= \prod_{j=M'+1}^M (1 - \lambda_j), M' < M \end{aligned}$$

This x^2 has $(p-M')(q-M')$ d.f.

The main objective of the analysis is to study the relationship of any variable in Y-set with any variable in X-set. The amount of relationship can be measured by calculating cross-weights, where the cross-weight is the product of canonical loadings of any variable and canonical correlation coefficient. For j-th canonical variate pair $\sqrt{\lambda_j}$ is the canonical correlation coefficient and

$$R_{xx(1)}^* = R_{xx} \cdot a_j \text{ and } R_{yy(1)}^* = R_{yy} \cdot b_j$$

are the canonical loadings of X-set and Y-set, respectively, corresponding to j-th canonical variate pair. Here a_j and b_j are the vectors of canonical weights for j-th variate pair. Each canonical variate pair explains certain percentage of total variation of Y-set and X-set. This can be measured, respectively by

$$R^2_{(j)y} = 1/p R^*_{yy(i)} \cdot R^*_{yy(j)} \text{ and}$$

$$R^2_{(j)x} = 1/q R^*_{xx(i)} \cdot R^*_{xx(j)}$$

Results and Discussion

Among the investigated units 82.1% are male students and among them 38.1% are smokers (Table 1). More male students are smokers. The differentials in smoking habit among males and females are statistically significant as $p(x^2) \geq 57.822) = .000$ though most of the students 88.3%, (Table 2) are highly aware of health hazard of smoking. More female students are aware (90.1%) of the problem, still a good number (8.8%) of them are smokers. However, the differentials in awareness among males and females are not significant [$\chi^2=0.630, p= 0.427$]. The study indicates that smoking is highly prevailed among males compared to females but both males and females are similarly aware of the health hazard of smoking. On the other hand, it is seen that (Table 3) rate of smokers is less among the students who are highly aware of the problem. Awareness and smoking habit are negatively significantly associated. [$\chi^2= 5.423, p=.02$]. The data show that 70.4% respondents are from urban area and among them 87.4% are highly aware of the problem of health hazard. This latter percentage (Table 4) among rural students is 90.7. Most of the respondents, either from rural area or urban area, are aware of the problem.

Table 1: Distribution of students according to sex and smoking habit.

Sex	Smoking Habit				Total n	%
	Yes		No			
	n	%	n	%		
Male	317	38.1	514	61.9	831	82.1
Female	16	8.8	165	91.2	181	17.9
Total	333	32.9	894	67.1	1012	100

Table 2: Distribution of students according to awareness of health hazard of smoking and sex.

Sex	Awareness				Total n	%
	Medium		High			
	n	%	n	%		
Male	100	12.0	731	88.0	831	82.1
Female	18	9.9	163	90.1	181	17.9
Total	118	11.7	894	88.3	1012	100

Table 3: Distribution of students according to awareness of health hazard of smoking and smoking habit.

Smoking habit	Awareness				Total n	%
	Medium		High			
	n	%	n	%		
Yes	50	42.4	283	31.3	333	32.9
No	68	57.6	611	68.7	679	67.1
Total	118	11.7	894	88.3	1012	100

Table 4: Distribution of students according to residential origin and awareness of health hazard of smoking.

Residential origin	Awareness				Total	
	Medium n	%	High n	%	n	%
Urban	90	12.6	622	87.4	712	70.4
Rural	28	9.3	272	90.7	300	29.6
Total	118	11.7	894	88.3	1012	100

Thus, differentials in origin of residence and awareness are not significantly different. [$\chi^2=2.241, p=0.134$]. This phenomenon has been reported earlier by Bhuyan et al. [10,11]. Insignificant variables (Table 5) [$\chi^2=1.7, p=0.42$]. is also observed in the levels of awareness among the respondents of different ages. The study is also similar to that reported by Bhuyan et al. [10,11]. Insignificance in variations in the levels of education of father [$\chi^2=0.33, p=0.848$], occupation of mother [$\chi^2=3.75, p>0.05$] and education of mother [$\chi^2=1.735, p=0.42$] according to levels of awareness are also observed in analyzing the data. However, father's occupation is significantly associated (Table 6), [$\chi^2=6.32, p<0.05$] with the levels of awareness of their offsprings. The analytical results indicate that some of the socioeconomic variables are associated with the knowledge of health hazard of smoking. Again, smoking is significantly associated with knowledge of awareness (Table 3). Smoking is also associated with some of the socioeconomic characters [10,11]. Important socioeconomic variables of the respondent which are significantly associated with their smoking habit are their sex, (Table 1), age (Table 7) and father's occupation (Table 6).

Table 5: Distribution of students according to age groups and awareness of health hazard of smoking.

Awareness	Age groups						Total	
	<19		19-22		22+			
	n	%	n	%	n	%	n	%
Medium	14	15.1	70	11.7	34	10.5	118	11.7
High	79	84.9	526	88.3	289	89.5	894	88.3
Total	93	9.2	596	58.9	323	31.9	1012	100

Table 6: Distribution of students according to their father's occupation and awareness.

Occupation	Awareness				Total	
	Medium n	%	High n	%	n	%
Service	53	10.8	439	89.2	492	48.6
Business	39	10.4	335	89.6	374	37.0
Others	26	17.8	120	82.2	146	14.4
Total	118	11.3	894	88.7	1012	100

Table 7: Distribution of students according to their smoking habit and age.

Age	Smoking habit				Total	
	Yes		No		n	%
	n	%	n	%		
<19	16	17.2	77	82.8	93	9.2
19-22	200	33.6	396	66.4	596	58.9
22+	117	36.2	206	63.8	323	31.9
Total	333	32.9	679	67.1	1012	100

It is seen that with the increase in ages of the respondents smoking habit is increased significantly [$\chi^2=12.109, p=0.002$]. Prevalence of smoking is more among higher aged students. This is natural as time passes on the students are influenced by their friends and most of them are away of their parents. Social and family restriction on smoking is reduced day by day. Father's education [$\chi^2=3.845, p=0.146$], mother's education [$\chi^2=3.13, p=0.208$], mother's occupation [$\chi^2=2.10, p=0.38$] and family monthly income [$\chi^2=4.716, p=0.194$] are not significantly associated with the smoking behavior of offspring. However, more offspring of servicemen (Table 8) are prone to smoking [$\chi^2=31.47, p=0.000$]. Similar results are reported by Bhuyan et al. [11]. It is seen that some of the socioeconomic variables are associated with awareness of health hazard of smoking and smoking habit. Again, smoking habit is associated with awareness of health hazard. Thus to study the complex relationship of socioeconomic variables with smoking habit and awareness of health hazard of smoking canonical correlation analysis is performed. The analysis is done by transforming the variables in nominal scale.

Table 8: Distribution of students according to their smoking habit and father's occupation.

Father's occupation	Smoking habit				Total	
	Yes		No		n	%
	n	%	n	%		
Business	117	31.3	257	68.7	374	37.0
Service	173	35.2	319	64.8	492	48.6
Others	43	29.5	103	70.5	146	14.4
Total	333	32.9	679	67.1	1012	100

In performing the canonical correlation analysis the following information are observed. Here R_{xx} is the correlation matrix (Table 9) of the predictor variables, R_{xy} (Table 10) is the correlation matrix of the criterion and predictor variables and R_{yy} is the correlation (Table 11) matrix of the criterion variables. The rank of the product matrix $R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$ is $M=\min(p, q)=2$ and hence canonical variate pairs. There will be

at best 2 canonical variate pairs. The variate pairs are related to the eigen values $\lambda^1 = 0.078$ and $\lambda^2 = .017$ and both pairs are found significant (Table 12). The canonical weights are the elements of

eigen vectors corresponding to λ^1 and λ^2 and these weights indicate the importance of the variables to maximize the correlations of two sets. The weights are shown in Table 13.

Table 9: Correlation matrix for the variables in the predictor set (X-set), R_{xx}

Age(x_1)	1								
Sex(x_2)	-0.018	1							
Marital status(x_3)	-0.022	1	1						
Religion(x_4)	0.021	-0.024	-0.067	1					
Father's education(x_5)	-0.14	0.067	0.065	-0.031	1				
Father's occupation(x_6)	0.217	-0.076	-0.046	0.020	-0.063	1			
Mother's education(x_7)	-0.202	0.148	0.005	-0.045	0.554	-0.075	1		
Mother's occupation(x_8)	-0.003	0.008	0.109	-0.033	-0.125	0.004	-0.242	1	
Income(x_9)	-0.032	0.137	-0.040	-0.062	0.117	-0.097	0.126	-0.041	1

Table 10: Correlation matrix for the variables in X-set and Y-set: R_{xy} .

	Smoking habit (y_1)	Awareness (y_2)
Age(x_1)	-.089	.040
Sex(x_2)	.239	.025
Marital status (x_3)	-.040	.105
Religion (x_4)	.002	.013
Father's education (x_5)	-.056	.006
Father's occupation (x_6)	-.021	-.021
Mother's education (x_7)	-.034	.031
Mother's occupation(x_8)	.027	-.002
Income (x_9)	-.030	.005

Table 11: Results related to test of significance of canonical variate pairs.

Smoking habit	1	
Awareness	0.073	1

Table 12: Results related to test of significance of canonical variate pairs: There will be at best 2 canonical variate pairs. The variate pairs are related to the eigen values $\lambda_1 = 0.078$ and $\lambda_2 = .017$ and both pairs are found significant.

Canonical variate pair	Eigen values λ_j	Wilk's Λ	χ^2 -statistics	P-value	Canonical correlation coefficients	% of the variation explained
1	.078	.906	99.801	.0001	.279	82.26
2	.017	.983	17.335	.029	.130	17.74

It is seen that the first canonical variate pair explains 82.26% of variation in the data set and the important variables to explain this variation are sex and smoking habit. These two variables are significantly associated (Table 2). The second canonical variate pair explains 17.74% of variation in the data set and the important variables to explain this variation are marital status and awareness of health hazard of smoking. From the correlation matrix (Table 10) it is seen that the pair sex and smoking habit and marital status and awareness are highly correlated. The canonical correlation may not

provide the real importance of variables if the variables in X-set are collinear. To avoid this problem standardized canonical correlation coefficients are calculated (Table 13). However, from both the analytical results similar conclusion can be drawn (Table 14).

Table 13: Standardized canonical correlation coefficients for X-set and Y-set

Variables	X-set		Y-set	
	λ_1	λ_2	λ_1	λ_2
x_1	-0.404	-0.381		
x_2	0.906	-0.339		
x_3	-0.609	-0.864		
x_4	0.004	-0.169		
x_5	-0.163	0.201		
x_6	0.041	0.176		
x_7	-0.228	-0.345		
x_8	0.012	0.040		
x_9	-0.199	-0.004		
y_1			1.002	-.039
y_2			-.112	-.996

Table 14: Correlation between the variables in X-set and Y-set.

Variables	X-set		Y-set	
	λ_1	λ_2	λ_1	λ_2
x_1	-.334	-.282		
x_2	.847	-.263		
x_3	-.187	-.791		
x_4	.002	-.103		
x_5	.204	-.027		
x_6	-.065	.170		
x_7	-.135	-.227		
x_8	.097	.009		
x_9	-.199	-.033		
y_1			.994	-.111
y_2			-.038	-.999

Conclusion

The present analysis is based on data collected from 1012 students of American International University Bangladesh, Jahangirnagar University and World University. The students are investigated according to convenience sampling under the supervision of teachers of the respective universities. Among the selected students 82.1% are males and 88.0% among them are highly aware of the health hazard of smoking. Still a good number of students (32.9%) are smokers. However, those who are aware of the problem of health hazard of smoking they are less prone (31.3%) to smoking. Lower level of awareness leads the students to be smoker in higher number. From the analysis it is seen that 88.3% respondents are highly aware of the problem. No one is observed, who is unaware of the problem. Awareness is independent of ages of respondents but smoking habit is not independent of ages and awareness, more students of higher ages are prone to smoking.

The study indicates that awareness and smoking habit are highly inter-related. Again, both these aspects are associated with some of the socioeconomic characters of the respondents. Specifically, the offspring of servicemen are more prone to smoking. As some of the socioeconomic characters of the respondents with awareness and smoking habit, are associated, canonical correlation analysis [12-14] has been performed to study the complex relationship of awareness and smoking habit with other socioeconomic variables. The analysis indicates that sex of respondents and their smoking habit and marital status and awareness are significantly inter-related.

References

- Cohen N (1981) Smoking, health and survival; prospects in Bangladesh, Lancet 1 (8229): 1090-1093.
- (2004) IARC, WHO; Tobacco Smoking, IARC Monograph on the evaluation of Carcinogenic Risks to Human, 83.
- (2005) Organization, WH, Waterpipe tobacco smoking; health effects, research needs recommended actions by regulators, WHO, Geneva, Switzerland.
- (2012) Mahato, Pradip' Knowledge of Health Effect of Tobacco Consumption among the Higher Secondary School Students, MPH Thesis, American International, University, Bangladesh.
- Ghani WM, Nabilla Razak IA, Yang WH, Talib NA, Ikeda N, et al. (2012) Factors affecting commencement and cessation of smoking behavior in Malaysian adults. BMC Public Health 12: 207.
- (1986) IARC, WHO, Tobacco Smoking, IARC, Monograph on the evaluation of Carcinogenic Risks to Human 38: 335-394.
- Sansonem CG, Raute LJ, Fong TG, Pedhekar MS, Quah ACK, et al. (2012) Knowledge of health effects and intentions to quit among smokers in India; Findings from the Tobacco Control Policy (TCP). India Pilot Survey. Int J Environ Res, Public Health 9(2): 564-578.
- AKI EA, Gaddam S, Gunukula SK, Honeine R, Jaoude PA, et al. (2010) The effects of waterpipe tobacco smoking on health outcomes: A systematic review. Int Jour Epi 39(3): 834-857.
- Khan N, Siddique M, Poddar AA, Hasmi SAH, Fatima S, et al. (2008) Prevalence, knowledge, attitude and practice of shisha smoking among medical dental students of Karachi, Pakistan Jour. Dow University of Health Science 2(1): 3-10.
- Khatun M, Bhuyan KC (2014) Awareness of health hazard of Tobacco Consumption among Students of American International University-Bangladesh 13(1): 85-92.
- Bhuyan KC, Fardus J, Khatun M (2016) Discriminating the students of universities by their smoking habit 15(1): 143-147.
- Bhuyan KC, Ghffar FA (1999) Canonical correlation analysis to study the influences of socioeconomic factors simultaneously on fertility and child mortality. Jour Stat Studies 19: 7-15.
- Kabir A, Merrill RD, Shamim AA, Klemm RDW, Labrique AB, et al. (2014) Canonical correlation analysis of infant's size at birth and maternal factors: A study in rural north-west Bangladesh. 0094243, Jour Pune.
- Hamid JS, Meaney C, Crowcroft N, Granerod J, Beyene J (2011) Potential risk factors associated with human encephalitis: application of canonical correlation analysis. BMC 11: 120.

ISSN: 2574-1241

DOI: [10.26717/BJSTR.2018.10.002011](https://doi.org/10.26717/BJSTR.2018.10.002011)

KC Bhuyan. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>