

STA 223 Project 2: Analysis of Characteristics Influencing Nicotine Usage

Christina De Cesaris



Introduction

- Nicotine usage associated with cancer and heart disease [4]
- Only about 3% of smokers quit nicotine usage successfully each year [1]
- Personality tendencies often linked to drug usage [3]
- Survey data of demographics, personality scores, and drug history from 1885 participants

Project Goal: Investigate the relationship between individual personality and demographic characteristics and nicotine usage status

Data Description and Preprocessing

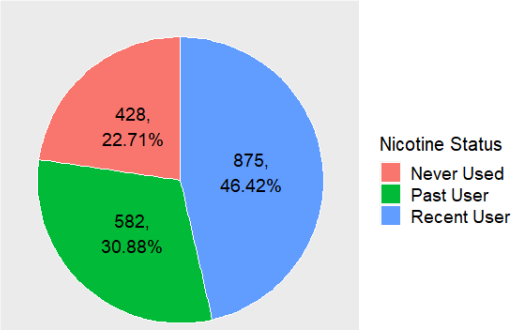
Contents: survey data, 1885 samples, 5 categorical predictors, 7 numeric predictors, 3 response levels

Variable Name	Variable Description
ID	Observation ID
Age	Age Range Group
Gender	Gender
Education	Education Level
Country	Country of Origin
Ethnicity	Ethnicity of Person
Nscore	Neuroticism Score: the long-term tendency to experience negative emotions such as nervousness, tension, anxiety and depression
Escore	Extraversion Score: manifested in outgoing, warm, active, assertive, talkative, cheerful, and in search of stimulation characteristics
Oscore	Openness Score: a general appreciation for art, unusual ideas, and imaginative, creative, unconventional, and wide interests
Ascore	Agreeableness Score: a dimension of interpersonal relations, characterized by altruism, trust, modesty, kindness, compassion and cooperativeness
Cscore	Conscientiousness Score: a tendency to be organized and dependable, strong-willed, persistent, reliable, and efficient
Impulsive	Impulsivity Score: the tendency to act on impulse
SS	Sensation Seeking Score: the tendency to pursue new and different sensations, feelings, and experiences
Nicotine	Nicotine Usage Status

ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Nicotine
1	0.49788	0.48246	-0.05921	0.96082	0.12600	0.31287	-0.57545	-0.58331	-0.91699	-0.00665	-0.21712	-1.18084	CL2
2	-0.07854	-0.48246	1.98437	0.96082	-0.31685	-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575	CL4
3	0.49788	-0.48246	-0.05921	0.96082	-0.31685	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	-1.37983	0.40148	CL0

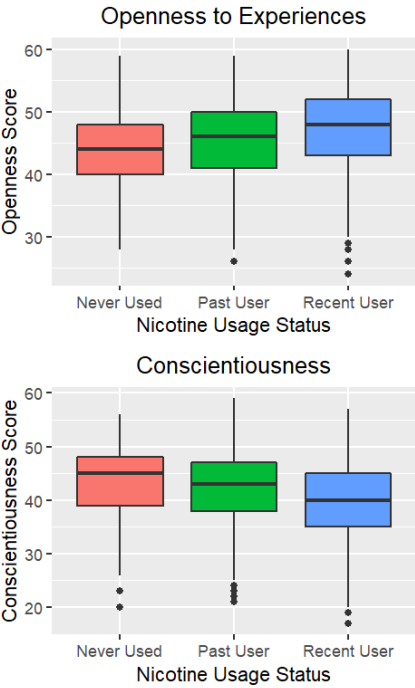
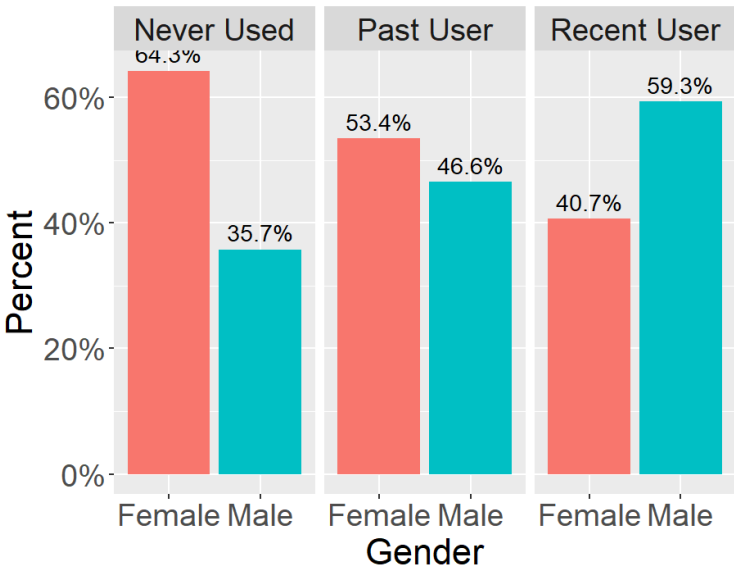
Usage Code	Usage Label	Simplified Labels
CL0	Never Used	Never Used
CL1	Used over a Decade Ago	Past User
CL2	Used in Last Decade	Past User
CL3	Used in Last Year	Past User
CL4	Used in Last Month	Recent User
CL5	Used in Last Week	Recent User
CL6	Used in Last Day	Recent User

Nicotine Status Breakdown



ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Nicotine	NicotineL
1	35-44	Female	Certificate/Trade Degree	UK	Mixed	39	36	42	37	42	355	132	Used in Last Decade	Past User
2	25-34	Male	University	UK	White	29	52	55	48	41	307	223	Used in Last Month	Recent User
3	35-44	Male	Certificate/Trade Degree	UK	White	31	45	40	32	34	276	249	Never Used	Never Used

Gender Breakdown



Methods

- Three-leveled, unordered outcome variable (Never Used = 1, Past User = 2, Recent User = 3) suitable for **baseline odds multinomial regression model** [2]
- Never Used (j=1) defined as baseline

Model:

$Y_i = (y_{i1}, y_{i2}, y_{i3})^T$ follows a multinomial distribution $Y_i \sim Multi(\pi_i, 1)$, with $\pi_i = (\pi_{i1}, \pi_{i2}, \pi_{i3})^T$ response probabilities for the $j = 1 \dots m$ response categories here $m = 3$.

$$\log\left(\frac{\pi_{i2}}{\pi_{i1}}\right) = x_i^T \beta_2, \log\left(\frac{\pi_{i3}}{\pi_{i1}}\right) = x_i^T \beta_3$$

- $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ for p explanatory predictors
- $\beta_j = (\beta_{0j}, \beta_{1j}, \beta_{2j} \dots \beta_{pj})^T, j \neq 1(baseline)$ associated coefficients.

Logit link:

For each sub model $\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \text{logit}(\pi_{ij}) = \eta_j$

Results

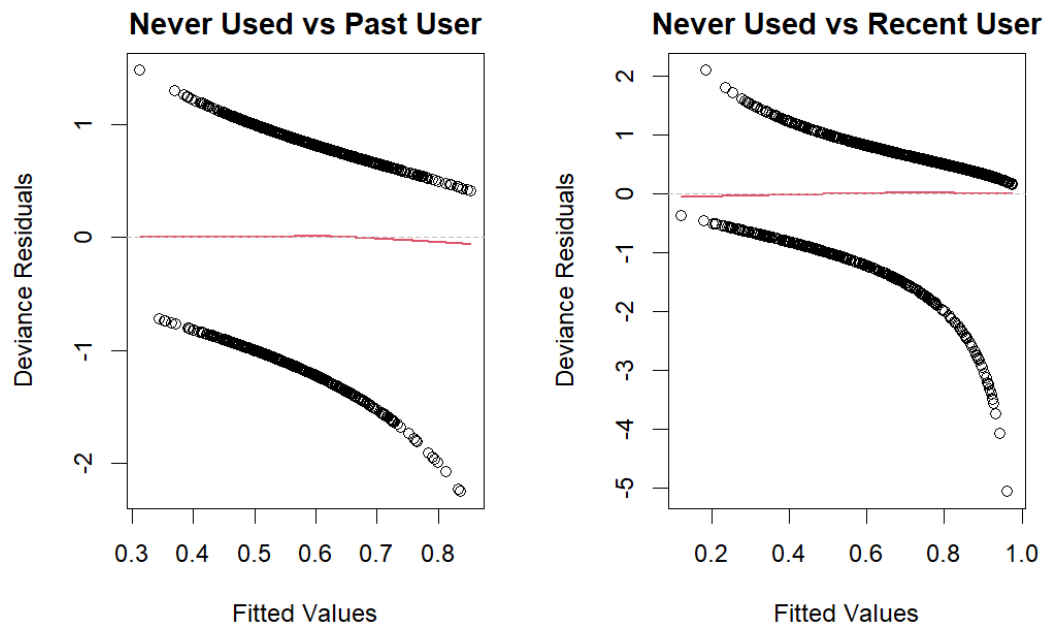
$$\eta_2 = \log\left(\frac{P(PastUser)}{P(NeverUsed)}\right) = \beta_{20} + \beta_{21}gender_{male} + \beta_{22}edu_{highschool} + \beta_{23}edu_{somecollege} + \beta_{24}edu_{somehighschool} + \beta_{25}edu_{university} + \beta_{26}openess + \beta_{27}conscientiousness + \beta_{28}impulsivity$$

$$\eta_3 = \log\left(\frac{P(RecentUser)}{P(NeverUsed)}\right) = \beta_{30} + \beta_{31}gender_{male} + \beta_{32}edu_{highschool} +$$

$$\beta_{33}edu_{somecollege} + \beta_{34}edu_{somehighschool} + \beta_{35}edu_{university} + \beta_{36}openess + \beta_{37}conscientiousness + \beta_{38}impulsivity$$

	Never Used = 0, Past User = 1				Never Used = 0, Recent User = 1			
	Estimate	P Value	CI 5%	CI 95%	Estimate	P Value	CI 5%	CI 95%
Intercept	-0.905	-0.176	-2.004	0.195	-0.007	0.992	-1.064	1.050
Male	0.304	0.025 *	0.080	0.527	0.558	2.142e-05 ***	0.342	0.775
High School	0.925	0.019 *	0.279	1.571	0.861	0.0224 *	0.241	1.481
Some College	0.419	0.073	0.034	0.804	0.511	0.018 *	0.155	0.866
Some High School	0.229	0.442	-0.261	0.718	0.542	0.045 *	0.096	0.987
University	0.084	0.661	-0.231	0.399	-0.499	0.006 **	-0.800	-0.198
Openness Score	0.036	0.0005 ***	0.019	0.0530	0.071	3.858e-12 ***	0.054	0.087
Conscientiousness Score	-0.016	0.1262	-0.034	0.001	-0.047	3.108e-06 ***	-0.064	-0.031
Impulsivity Score	1.499e-06	0.999	-0.002	0.002	-0.003	0.0012 **	-0.004	-0.001
Null deviance: 3980.072		Residual Deviance: 3693.059			AIC: 3729.059			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1								

Diagnostics



Discussion

- Men more likely to belong to Past User and Recent User categories compared to females
- Higher Openness Scores and lower Conscientious scores associated with Past and Recent User status
- Education predictor less significant in Past User model, overall, those without university education had higher probability of being Recent of Past Users
- Additional profiles could be included to improve diversity in data as most participants were white and from the USA or UK

- Possible bias in Education category since most Recent Users between 18-24
- No lack of fit from diagnostic plots

References

- Benowitz N. L. (2010). Nicotine addiction. The New England journal of medicine, 362(24), 2295–2303. <https://doi.org/10.1056/NEJMr0809890>
- Faraway, J. (2016). Extending the linear model with R. Second Edition, Chapman and Hall. ISBN 9781498720960
- Fehrman, Elaine & Muhammad, Awaz & Mirkes, Evgeny & Egan, Vincent & Gorban, Alexander. (2017). The Five Factor Model of Personality and Evaluation of Drug Consumption Risk. 10.1007/978-3-319-55723-6_18
- Saha, S. P., Bhalla, D. K., Wayne, T. F., Jr, & Gairola, C. (2007). Cigarette smoke and adverse health effects: An overview of research trends and future needs. The International journal of angiology: official publication of the International College of Angiology, Inc, 16(3), 77–83. <https://doi.org/10.1055/s-0031-1278254> 10.1109/ICCIT48885.2019.9038605.