

Relation between Drinking and Smoking to Academic Performance

Yunran Chen, Brian Kunding, Rihui Ou

Abstract

We aim to examine the relationship between smoking and drinking behaviors on academic performance of college students. Our data comes from Harvard School of Public Health survey on alcohol and drug use among college students conducted in 2001. We apply a structural equation model to construct latent factors for drinking and smoking behavior, and explore their relationship to GPA with adjusting the potential effects from Greek life and other background such as gender and race. Our result suggests heavy use of alcohol/cigarette may relate to worse academic performance, and use of alcohol and cigarette is positively correlated. In addition, active participation in Greek life may influence the academic performance through drinking and smoking.

1. Introduction

The data we are studying is from a Harvard School of Public Health survey on alcohol and drug use among college students. The study was conducted over the span of a decade, with surveys administered in years 1993, 1997, 1999, and 2001, and it hoped to identify relationships between high risk behaviors and factors such as campus drinking culture, college drinking policies, and prevalence of greek life.

For this paper, we are interested in examining the relationship between smoking and drinking behaviors on academic performance of college students surveyed in 2001. As is the case with many surveys in behavioral and social studies, the smoking and drinking behaviors are measured indirectly through several highly correlated items in the questionnaire. We therefore use a Structural Equation Model (SEM) (Muthén 1984) to summarize a large number of variables into a smaller set of latent variables, and study the interrelationships among observed variables and latent variables.

2. Materials and Methods

2.1 Data cleaning and transformations

The variables have widely varying rates of missingness because respondents are instructed to skip certain questions based on previous answers. Therefore, to prepare our dataset for analysis, we first identified the questions relating to smoking and drinking that were not predicated on previous responses (E3, E4, C5), and retain only the complete cases among these. This left us with 88.5% of the data. The missingness in the other variables in question was imputed based on previous responses. For example, when the survey instructs those who have not had a drink in the last 30 days to skip the C11 (“How many times have you had a drink in the last 30 days?”), we imputed the missing responses as zero.

2.2 Structural Equation Model

A structural equation model (SEM) is built with two components. The first component is a measurement model (equation (1)), or confirmatory factor analysis (CFA), which summarize highly correlated observed variables to latent variables with taking measurement errors into consideration. The second component is a structural model (equation (2)), or structural regression model studying relationships between the latent factors.

$$\mathbf{y} = \boldsymbol{\nu} + \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (1)$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \quad (2)$$

where $\mathbf{\Lambda}$ is a $n \times p$ factor loading matrix, $\boldsymbol{\eta}$ is a $p \times 1$ latent factor vector that is not observable, e.g., drinking behaviour. $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ are intercepts. $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$ are residuals that are normally distributed. Typically, $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$ are assumed to possess a covariance structure. The $p \times p$ matrix \mathbf{B} measures the conditional dependency across latent variables. If $B_{ij} \neq 0$, then factor η_i is a child of η_j , i.e., there exists a directed edge from factor j to i . $\mathbf{I} - \mathbf{B}$ is invertible and identifiability constraints are applied.

The interpretation of y depends on the type of data in the questionnaire. For the continuous questionnaire response, y is the exact response value. For the ordinal response, the continuous \mathbf{y} is assumed to be a normal latent variable that underlies the observable questionnaire response \mathbf{z} , where $z = m$, if $y \in [\tau_{m-1}, \tau_m]$ and $-\infty = \tau_0 < \tau_1 < \dots < \tau_K = +\infty$ are the thresholds.

In our setting, we aim to construct latent variables representing “Drinking Behavior”, “Smoking Behavior” and “Academic Performance”. The latent factor for “Greek Life” is also introduced since it may have potential influence on both drinking/smoking behavior and academic performance. We assume that “Smoking Behavior” can be explained by “Greek Life”, “Drinking Behavior” can be explained by both “Greek Life” and “Smoking Behavior”, and “Academic” can be explained by all latent factors together with background variables such as gender and race. The model is shown in Figure 1

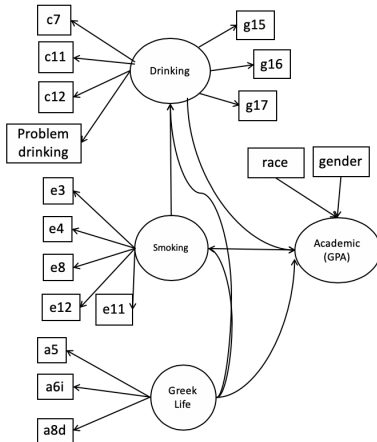


Figure 1: The directed acyclic dependency graph between variables. Observable responses are marked in square and factors are marked in circle.

We choose the variables which were similar enough to construct a coherent latent variable that would be easy to interpret, but diverse enough to capture a broad collection of behaviors. We also prefer ordinal variables and variables with less missing values. For “Drinking Behavior” latent variable, we used variables relating to self-identification of drinking behavior (C7), frequency (C11) and amount (C12) of drinking over the last 30 days (C11), and the use (G15,G17) and altitude (G16) in family. Lastly, we used a “problematic drinking” variables, which was the sum of the number of more extreme signs of alcoholism that responded had exhibited in the last 30 days, reported in C17. For “Smoking Behavior” latent variable, we used frequency of smoking over the last 30 days (E3), amount of cigarettes smoked on average (E4), whether the respondent smoked only on their own or with others (E8), and how often exposed to smoking (E11,E12). For academic performance, we consider the To construct the “Greek Life” latent variable, we used whether or not the respondent was a member of greek life (A4), whether or not they lived in their fraternity or sorority (A5), and whether or not they had indicated that greek life was important to them (A8d).

3. Results

3.1 Exploratory Data Analysis

Data exploration suggests heavier use of alcohol is related to worse academic performance. As shown in Figure 2, as the increase of the level of each item (heavier use of alcohol), the proportion of higher GPA students decrease. Especially, students with extreme high level on **problem drinking** are all students with GPA less than A-. Figure 3 suggests the increasing proportion of higher GPA is related to increasing use of cigarretes and less exposed to smoking environment. Figure 5 and 4 suggest heavier use of alcohol/cigarette relate to active participation in Greek life or heavier use of alcohol/cigarette in family.

3.2 Main Results

We conducted the analysis by apply R package `lavaan`(Rossee 2012). Table 6 presents the result for the measurement model, depicting the factor loading matrix. All the items are significant showing they are clear indicators for latent factors. And the signs align with our expectation. For the survey, the answers are encoded in the way that lower value represent a better lifestyle. For example, the abstainer is encoded as 1 in question c3. Therefore all the signs are reasonable. As a result, lower value of drinking/smoking factor corresponds to light use of alcohol/cigarette. And lower level of greek life factor presents less participation to greek life. Notice that low value of a8d indicates the greek life is very important for students, resulting in the negative sign.

Table 7 presents the relationship between the latent factors. Drinking and smoking are significantly related to the academic performance. Specifically, lower level of smoking/drinking is related to better academic performance. And smoking and drinking are positive correlated. It is worth noting although greek life does not have directly influence on academic performance, but it may indirectly influence academic performance through drinking and smoking. A mediation analysis is needed for further exploration.

In terms of model fit evaluation, the fit statistics of our model is RMSEA 0.03 (<0.08), CFI 1.00 (>0.95) and SRMR 0.02 (<0.06). We have strong evidence that this model achieves a good fit as suggested by (Hu and Bentler 1999).

3.3 Sensitivity Analysis

We conduct a sensitivity analysis on the specification of covariance matrix of residuals in our model. We compare the result based on a structured covariance and unstructured covariance. The result still holds since the p-values and the signs of the estimator stay the same and only slight changes of the esimators. This suggests our model is relatively robust.

5. Discussion

We might have improved our analysis through a different construction of our latent variables. For example, when designing the “Problematic Drinking” variable, our choice to include an indicator for respondents who exhibitted a behavior three times or more in the last 30 days was arbitrary. In fact, since having a hangover (C17a) is a much less severe indication of alcohol abuse than requiring medical treatment for alcohol overdose (C17l), we could have used different benchmarks for each variable when determining whether a response was truly a problematic drinking behavior.

In future studies, we would also like to follow up on these results through mediation analysis. As we saw, our greek life latent variable was not significantly correlated with academic performance, but it was highly correlated with heavy drinking, which in tern was significantly correlated with worse academic performance. Since structural equation models do not lend themselves to interaction effects, one approach to investigate this further would be through mediation analysis.

Appendix

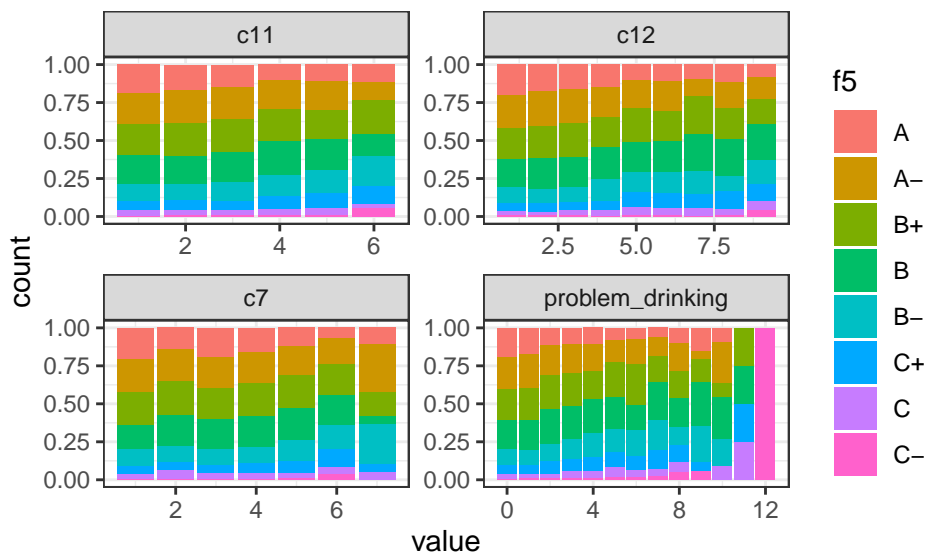


Figure 2: Relation between drinking behavior and GPA

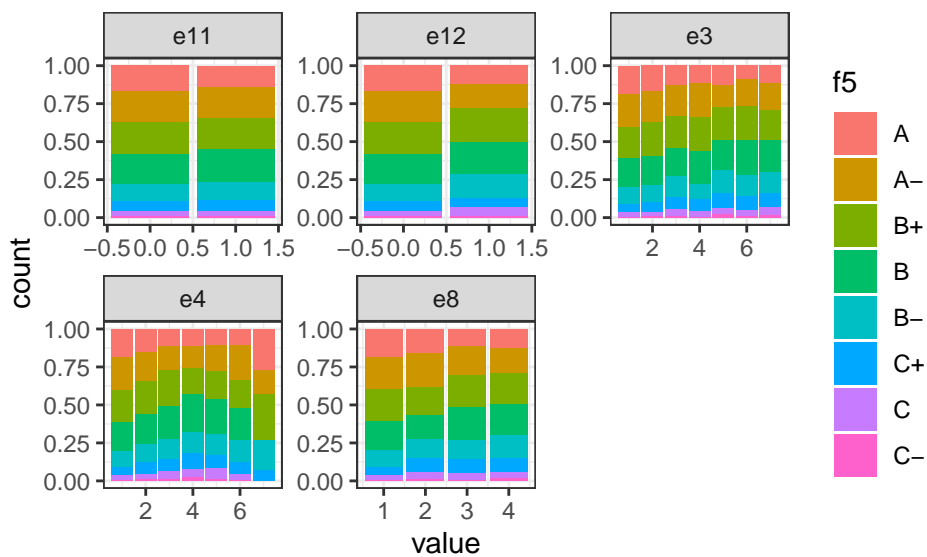


Figure 3: Relation between smoking behavior and GPA

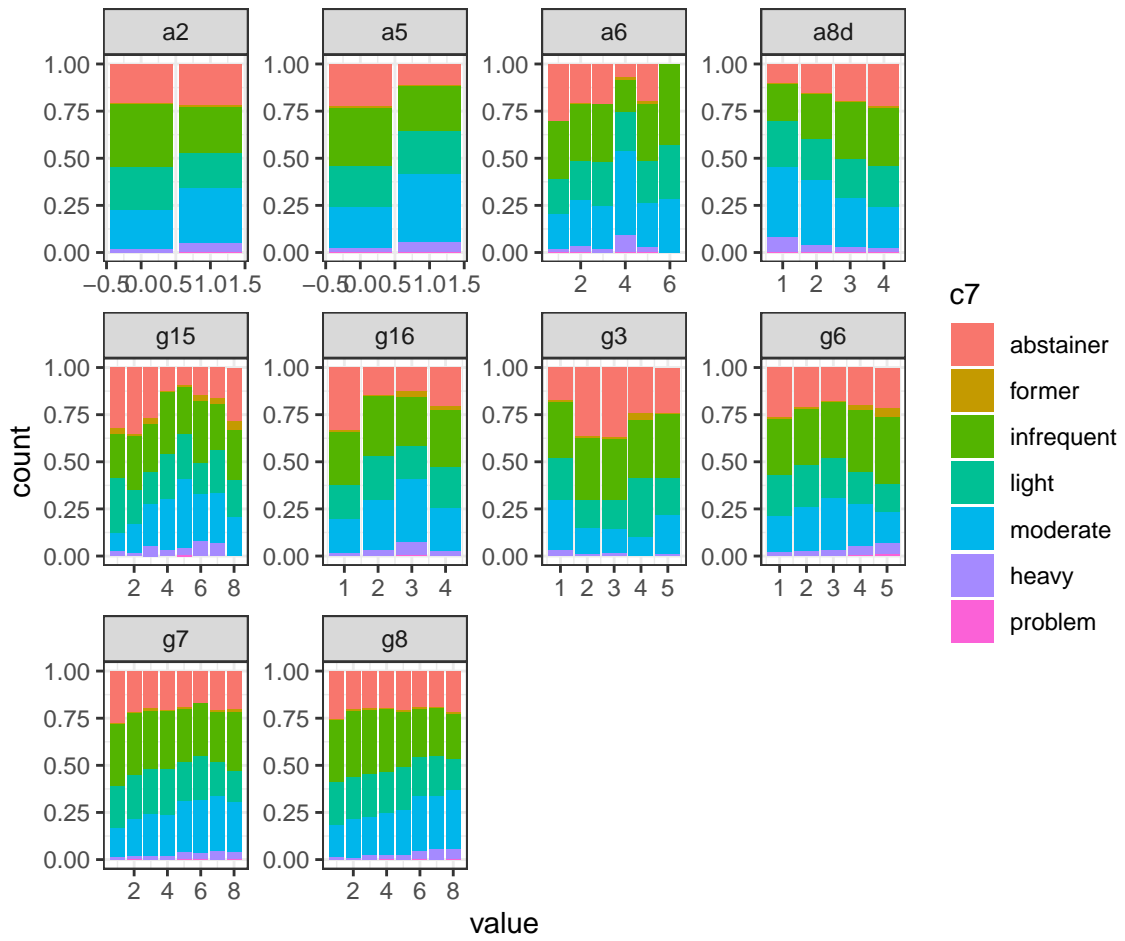


Figure 4: Relation between drinking and background

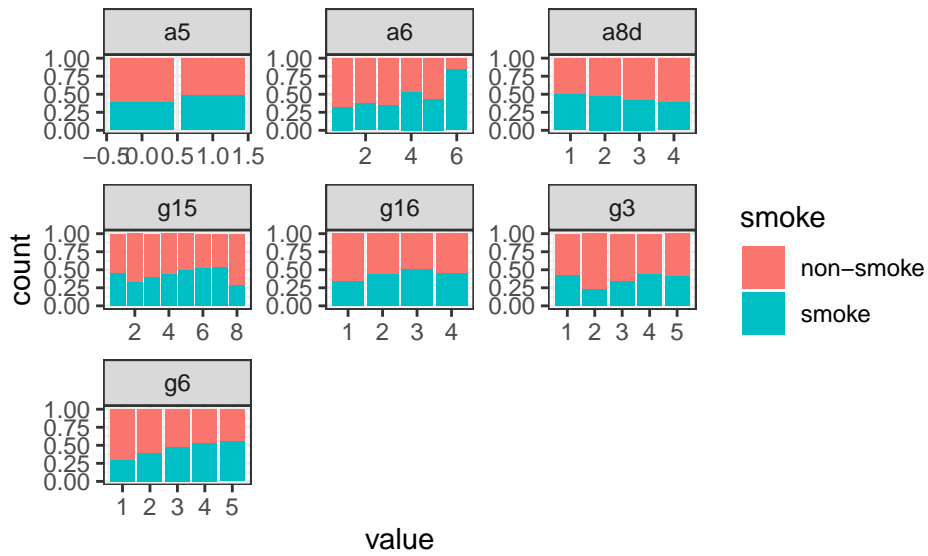


Figure 5: Relation between smoking and background

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)
drink =~				
c7	1.000			
c11	0.936	0.012	77.808	0.000
c12	0.855	0.013	65.463	0.000
problem_drnkng	1.258	0.029	43.175	0.000
g15	0.248	0.014	18.353	0.000
g16	0.189	0.014	13.326	0.000
g17	0.057	0.014	4.183	0.000
smoke =~				
e3	1.000			
e4	0.962	0.003	367.384	0.000
e8	1.016	0.012	86.737	0.000
e11	0.050	0.007	6.962	0.000
e12	0.011	0.002	5.066	0.000
greek =~				
a5	1.000			
a6i	0.067	0.008	8.911	0.000
a8d	-0.880	0.043	-20.553	0.000

Figure 6: Result of Measurement Models

Regressions:				
	Estimate	Std.Err	z-value	P(> z)
academic ~				
drink	0.134	0.017	7.960	0.000
smoke	0.085	0.016	5.190	0.000
greek	-0.002	0.017	-0.096	0.924
male	0.200	0.022	9.047	0.000
white	-0.364	0.041	-8.919	0.000
black	0.235	0.055	4.238	0.000
asian	-0.150	0.056	-2.690	0.007
native_american	-0.008	0.138	-0.058	0.954
drink ~				
greek	0.218	0.037	5.836	0.000
smoke ~				
greek	0.093	0.024	3.943	0.000
drink ~				
smoke	0.514	0.011	45.783	0.000

Figure 7: Result of Structural Models

References

- Hu, Li-tze, and Peter M Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling: A Multi-disciplinary Journal* 6 (1). Taylor & Francis: 1–55.
- Muthén, Bengt. 1984. "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators." *Psychometrika* 49 (1). Springer: 115–32.
- Rosseel, Yves. 2012. "Lavaan: An R Package for Structural Equation Modeling and More. Version 0.5–12 (Beta)." *Journal of Statistical Software* 48 (2): 1–36.