Relation between Drinking and Smoking to Academic Performance

Yunran Chen, Brian Kundinger, Rihui Ou

Introduction

- Data: 2001 College Alcohol Study, 10904 observations respectively
- Goal: How do drinking and smoking affect academic performance?
- Challenges: High correlation between individual variables (survey data)
- Solution: latent factor based model
- Model:
  - Structural Equation Model (SEM): studying interrelationships among observed and latent variables
  - Build SEM for 2001

Data Preprocessing

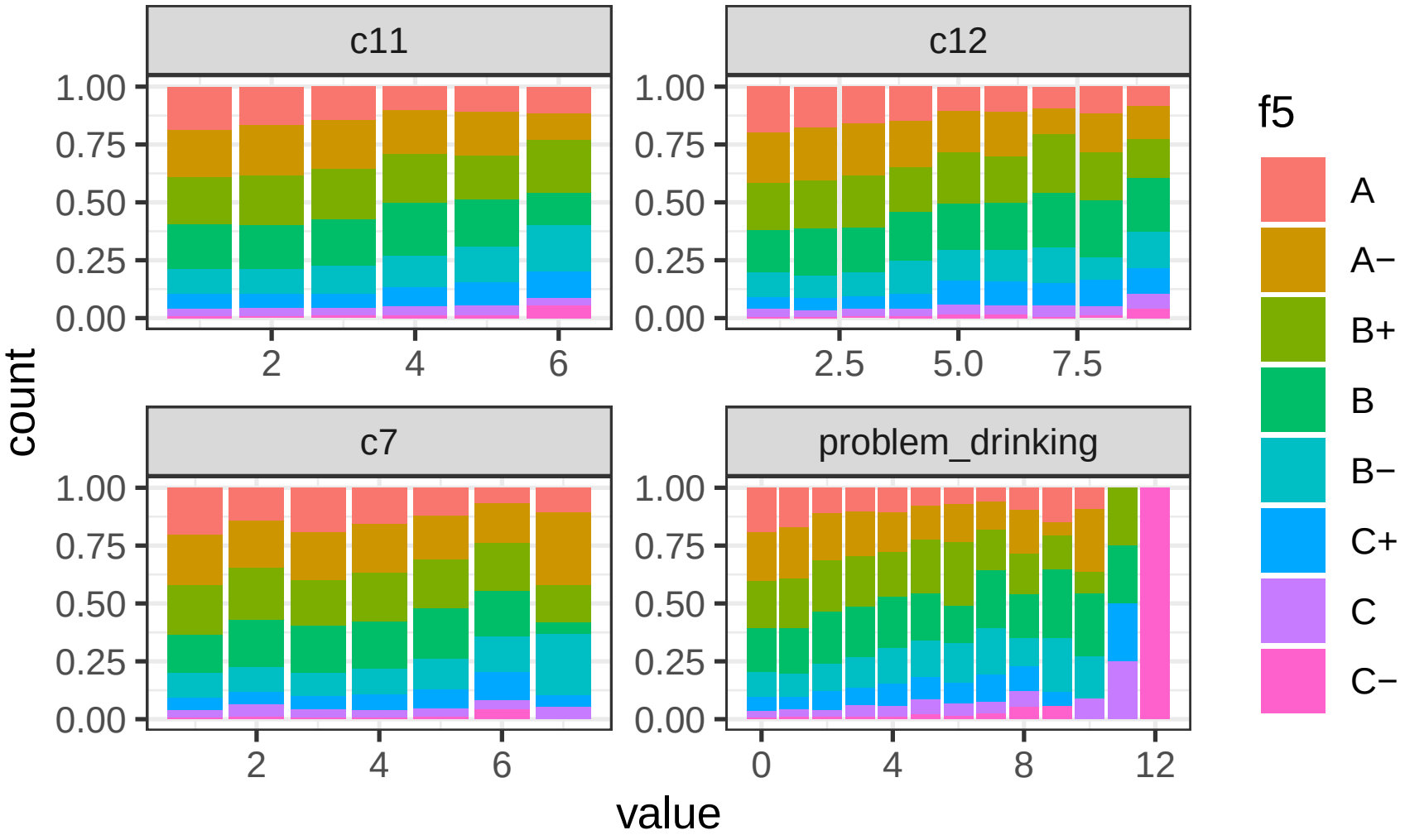
- Obtain the complete cases based on questions of interest with no screening (keep 87.5%)
- Impute the NA's based on the relation to other questions
- Create problem_drinking counting how many times a student cause problems because of drinking

Pick variables

- Academic (ordinal response variable): F5 (GPA)
- Drinking:
  - C7: Self-identified alcohol use (abstainer/former drinker/...)
  - C11: Frequency of drinking
  - C12: Amount of alcohol consumed when drinking
  - problem_drinking (C17): how often has your drinking caused you to ...
  - G15,G16,G17: Alcohol use in family
- Smoking:
  - E3: Frequency of mmoking
  - E4: Average number of cigarettes per day
  - E8: Social smoking (alone vs. with friends)
  - E11-12: How often exposed to smoking
- Background:
  - Fraternity/Sorority: A5,A6,A8d
  - Gender,Race: A2,G3a-e

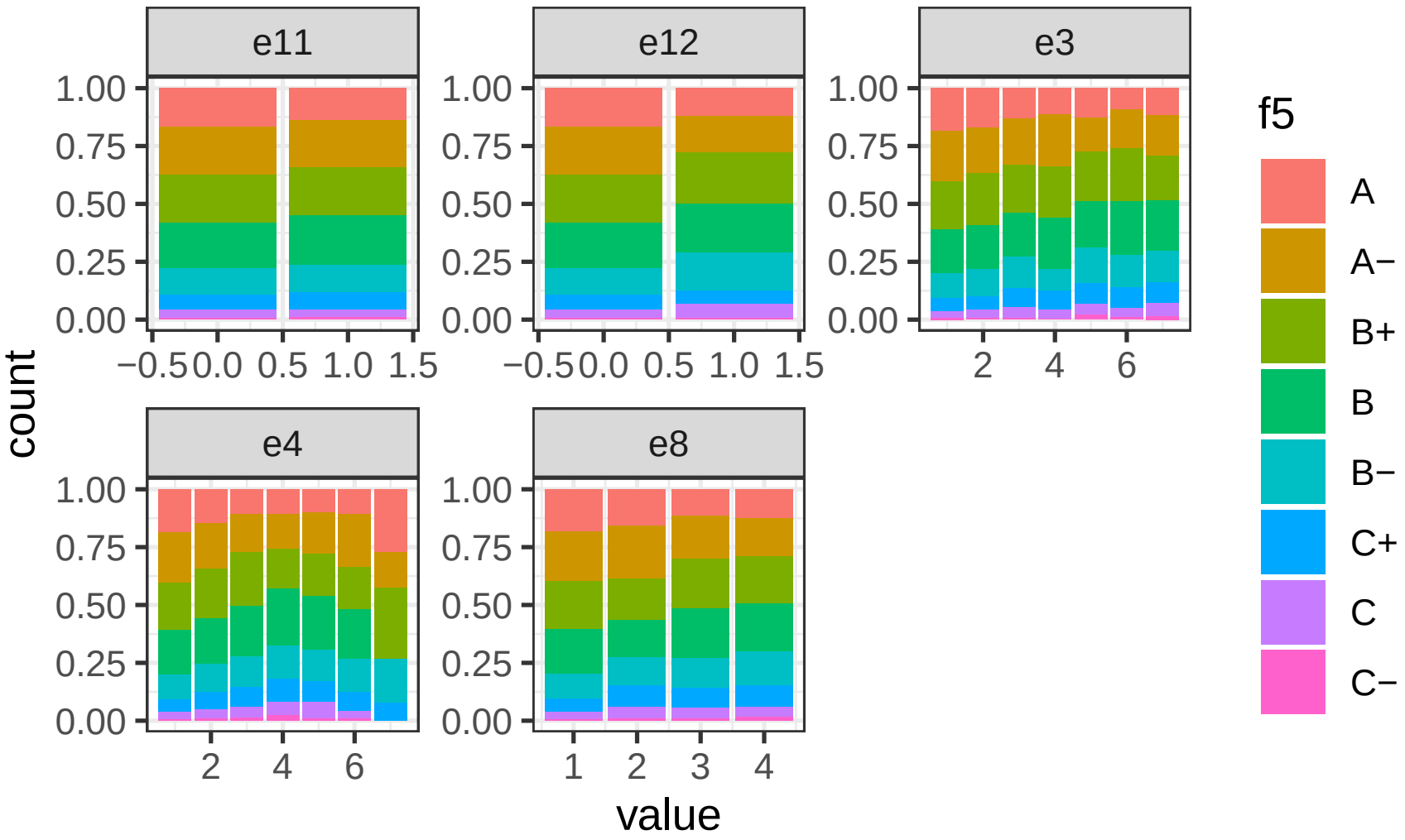
EDA: Relation between drinking behavior and GPA

- Students with higher GPAs (A coded as 1) show lower alcohol use



EDA: Relation between smoking behavior and GPA

Similarly, students with high GPAs show less smoking, but the correlation is less clear (E4 shows many students who smoke more than one pack a day with high GPA)



EDA: Relation between drinking/smoking and background

Heavier use of alcohol and cigarettes is related to

- Member of a fraternity/sorority
- Live in fraternity/sorority
- Think fraternity/sorority life is important
- Mother/father drinks
- Family accepting of heavy drinking

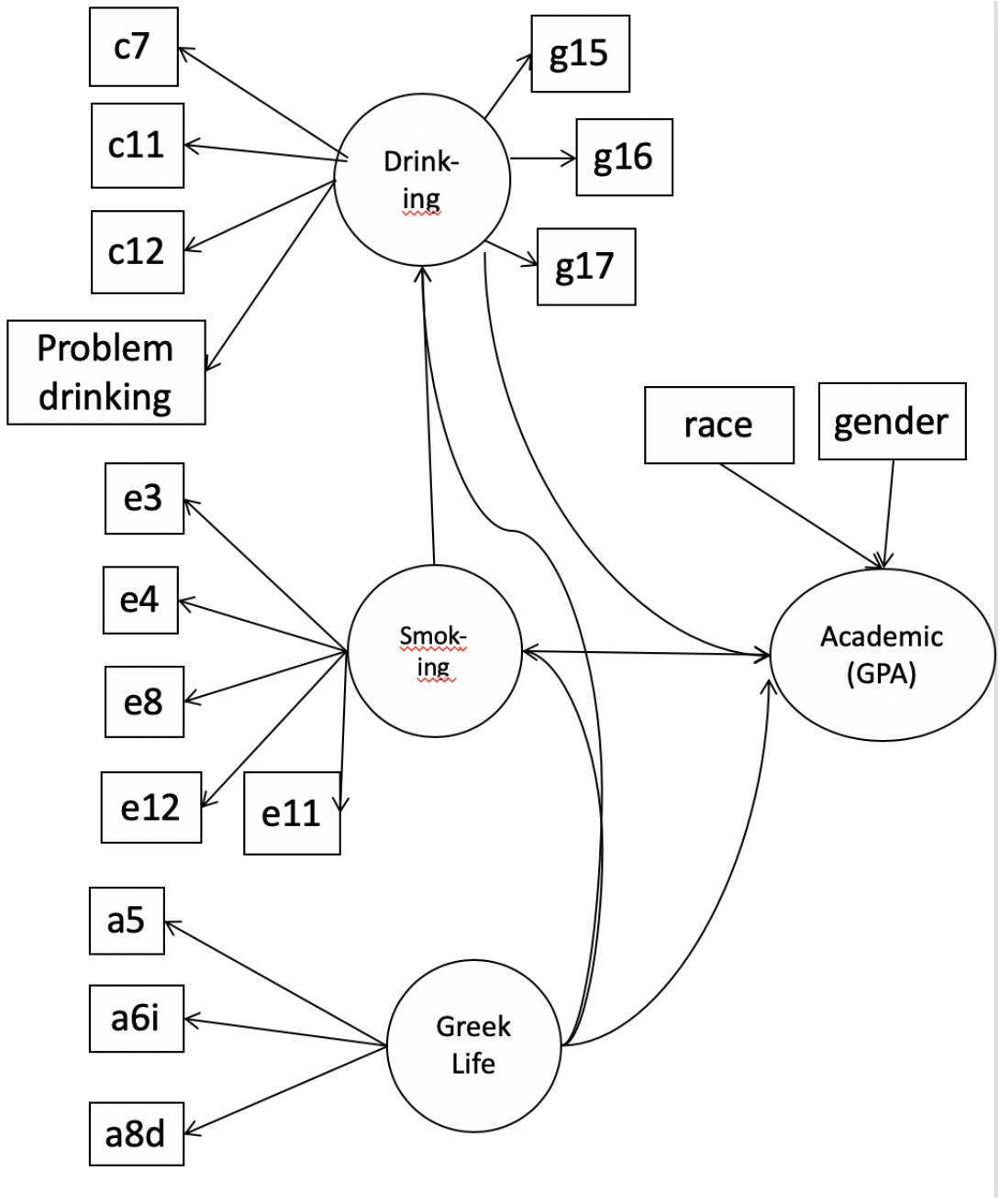
Model

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$$

- $\boldsymbol{\Lambda}$ is the factor loading matrix,
- $\boldsymbol{\eta}$ are the latent factors that are not observable, e.g., drinking tendency,
- $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ are intercepts for the observational and latent variables respectively,
- $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$ are residuals that are normally distrubuted.
- $\boldsymbol{B}$ measures the conditional dependency across latent variables. If $\boldsymbol{B}_{ij} \neq 0$, then $\boldsymbol{\eta}_i$ is a child of $\boldsymbol{\eta}_j$, i.e., there exists a directed edge from factor j to i.

For the ordinal questionnaire data, the continuous $\boldsymbol{y}$ is assumed to be a normal latent variable that underlies the observable questionnaire response $\boldsymbol{z}$, where $z = m$ if $y \in [\tau_{m-1}, \tau_m]$ and τ is the threshold.

Model


c7
c11
c12
Problem drinking
Drink-ing
g15
g16
g17
race
gender
e3
e4
e8
e12
e11
Smok-ing
Academic (GPA)
a5
a6i
a8d
Greek Life

Results

```
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
```

Results

```
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_americn   -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
```

Interpretation

- Latent Variables: All variables have coefficients that align with expectations, and all have significant p-value. We trust these variables constitute latent variables.
- Regressions: Drinking and smoking associated with lower GPA (A is coded as 1)
- Greek life not significant directly on academic performance, but highly correlated with drinking.

Discussion

- Data imputation-probably missing not at random
- Comparison between 1997 to 2001 to study the trend
- Explore effect of greek life more rigorously through Mediation Analysis