Relation between Drinking and Smoking to Academic Performance

Yunran Chen, Brian Kundinger, Rihui Ou

Introduction

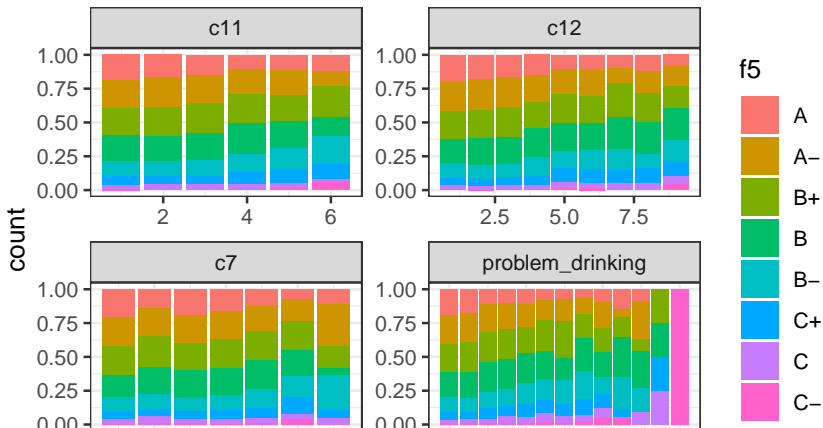- Data: 2001 College Alcohol Study, 10904 observations respectively
- Goal: How do drinking and smoking affect academic performance?
- Challenges:
  - latent construct of drinking behavior and smoking behavior cannot be directly measured by a single observed variable
  - -> latent factor based model
- Model:
  - Structural Equation Model (SEM): studying interrelationships among observed and latent variables
  - Build SEM for 2001

Pick variables

- Academic (response variable): f5 (GPA)
- Drinking:
    - c7: current use of alcohol (abstainer/former drinker/...)
    - c11: on how many occasions have a drink (1-2/3-5/...)
    - c12: how many drinks in these occasions (0/1/2/...)
    - problem_drinking(c17): how often has your drinking caused you to ...
    - family drinking: g15,g16,g17
- Smooking:
    - e3: on how many days smoke
    - e4: how many cigarettes a day
    - e8: smoke with people/alone/...
    - e11-12: exposed to smoking
- Background:
    - fraternity/sorority: a5,a6,a8d
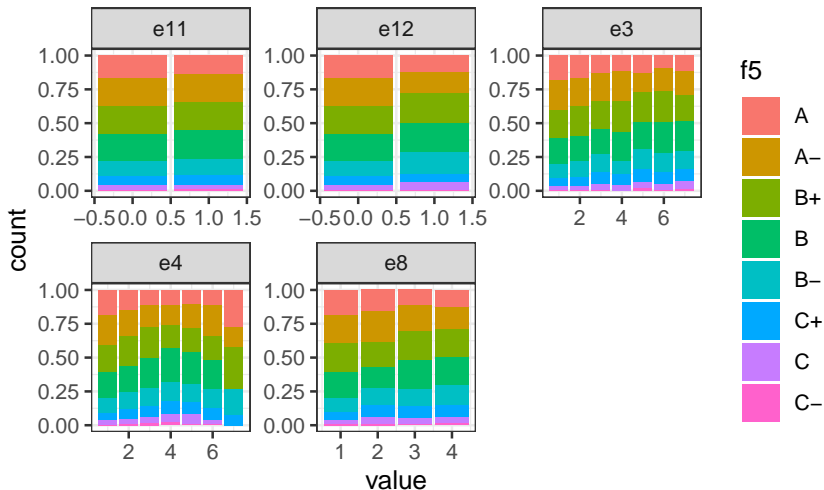   - gender,race: a2,g3a-e

EDA: Relation between drinking behavior and GPA

- ▶ The increasing proportion of higher GPA relate to less alcohol use
- ▶ Students with high GPA tend to "overestimate" the severity (vice versa for students with lower GPA)
- ▶ Students with low GPA tend to be heavy drinker and cause problems

EDA: Relation between smoking behavior and GPA

The increasing proportion of higher GPA relate to - less exposed to smoking environment - the degree of addiction decrease - For the students smoke more than a pack per day, students with high GPA take a large proportion.
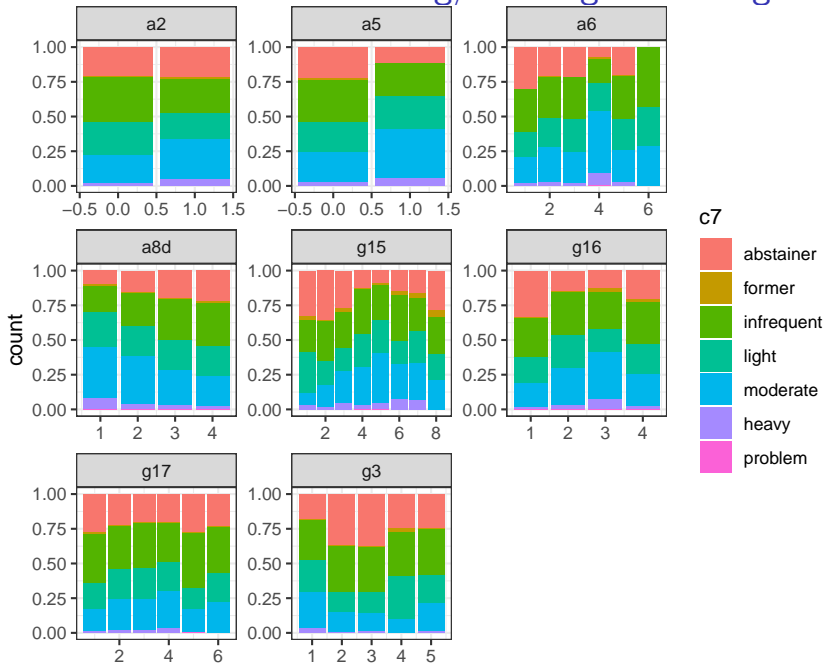
EDA: Relation between drinking/smoking and background

heavier use of alcohol/cigaret relate to

- member of a fraternity/sorority
- live in fraternity/sorority
- think fraternity/sorority life is important
- mother/father is problem drinker
- family accept heavy drinking

heavier use of alcohol relate to

- male

EDA: Relation between drinking/smoking and background

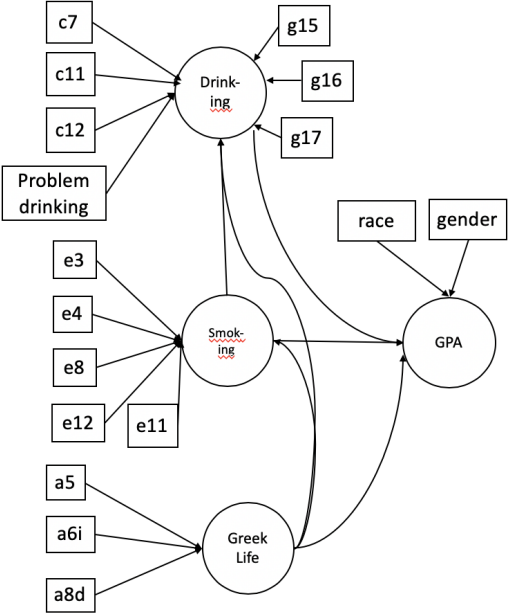# Data Preprocessing

- Impute the NA's based on the relation to other questions
- create indicator a6i presenting whether a student lives in fraternity/sorority house
- create problem_drinking counting how many times a student cause problems because of drinking
- Obtain the complete cases (keep 87.5%)

Model

Add SEM formula here

Model

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

Add here

Discussion

- Data imputation
- Comparison between 1997 to 2001