**Case Study 2: Boys Town Study of Youth Development (Version 2)**

**Overview**: The Boys Town Study of Youth Development surveyed 3,065 students in junior high and high schools in the Midwestern United States (predominantly in Nebraska and Iowa) in the mid-1970s. The study focused on adolescent substance use and deviant behavior, school aspirations, and parental and friendship relationships. Additional topics included opinions toward, influences for or against, and legal ramifications of substance use.

**Driving Challenges**:

* Are there relations between social status and delinquency covariates?
* Can we predict delinquent activities using alcohol abuse amount and parent closeness?
* Are there relations between parent closeness and attitudes on drinking alcohol?

**Meta-data**

**ID:** Case subject identifier

**Demographics:**

* **Sex:** dichotomous variable (1=male, 2=female)
* **GPA:** Interval-level variable with range of 1-4 (1-“A” average, 2- “B” average, 3- “C” average, 4- below “C” average)

***Alcohol abuse:***

* **Alcoholuse**- Interval level variable with 6 categories (1- everyday, 2- once or twice/wk, 3- once or twice/month, 4- less than once or twice per month, 5- once or twice, 6- never)
* **Attitudes on drinking:**
* **Alcatt-** Interval level variable (1- Approve generally, 2- Approve/disapprove depending on context, 3- Indifferent, 4- Disapprove generally)

**Social status-**

* DadJob: dichotomous variable with 1-yes and 2- no,
* MomJob: dichotomous variable with 1-yes and 2-no

**Parent closeness** (example: In your opinion, does your mother make you feel close to her?)

* Dadclose: Interval level variable (1-usually, 2- sometimes, 3-seldom, 4-never)
* Momclose: interval level variable (1-usually, 2- sometimes, 3-seldom, 4-never)

**Delinquency:**

* larceny (how many times have you taken things >$50?): Interval level data (1-never, 2- once or twice, 3- several times, 4- many times),
* vandalism: Interval level data(1-never, 2- once or twice, 3- several times, 4- many times)

**Provenance**: This case study only uses simulated data. The complete R-script generating the data is included below. The entire case study is CC-BY licensed and can be used, updated, refactored and expanded by the entire community.

**R-Code**:

library(truncnorm)

set.seed(1234)

NumSubj <- 3065

id <- c(1:NumSubj)

sex <- ifelse(runif(NumSubj)<.4,1,2)

gpa <- ceiling(rtruncnorm(NumSubj, a=0, b=4, mean=2, sd=1))

alcatt <- ceiling(rtruncnorm(NumSubj, a=0, b=4, mean=2, sd=1))

dadjob <- ifelse(runif(NumSubj)<.9,1,2)

momjob <- ifelse(runif(NumSubj)<.3,1,2)

dadclose <- ceiling(rtruncnorm(NumSubj, a=0, b=4, mean=2, sd=1))

momclose <- ceiling(rtruncnorm(NumSubj, a=0, b=4, mean=2, sd=1))

## create Alcoholuse variable using linear regression

## treat dadjob and momjob as binary variables "0" and "1"

Alcoholuse <- 3-0.3\*dadclose-0.5\*momclose+0.9\*alcatt-0.01\*gpa-0.02\*(dadjob-1)-0.04\*(momjob-1)

Alcoholuse <- Alcoholuse+rnorm(NumSubj, mean=0, sd=0.1)

Alcoholuse <- ceiling(Alcoholuse)

## create delinquent activities using linear regression

larceny <- 2-0.3\*Alcoholuse+0.1\*dadclose+0.3\*momclose-0.01\*gpa+0.02\*(dadjob-1)+0.05\*(momjob-1)

larceny <- larceny+rnorm(NumSubj, mean=0, sd=0.1)

larceny <- ceiling(larceny)

vandalism <- 2-0.4\*Alcoholuse+0.15\*dadclose+0.4\*momclose-0.04\*gpa+0.1\*(dadjob-1)+0.2\*(momjob-1)

vandalism <- vandalism+rnorm(NumSubj,mean=0,sd=0.1)

vandalism <- ceiling(vandalism)

boystown2 <- cbind(id,sex,gpa,Alcoholuse,alcatt,dadjob,momjob,dadclose,momclose,larceny,vandalism)

write.csv(boystown2, file="boystown2.csv", row.names=FALSE)