Applied Analysis 5

Rohan Hore

August 2024

Adolescent smoking This is a longitudinal, natural history study of adolescent smoking. Students included in the longitudinal study were either in grade 8 or 10 at baseline, and self-reported on a screening questionnaire 6 to 8 weeks prior to baseline that they either had never smoked, but indicated a probability of future smoking, or had smoked in the past 90 days. The longitudinal study utilized a multi-method approach to assess adolescents at three time intervals of one week: Baseline, 6 months, and 12 months. Adolescents carried hand-held computers with them during the 7 consecutive day data collection period at each of the three time points and were trained to both respond to random prompts from the computers and to event record (initiate a data collection interview) smoking episodes. Immediately after smoking a cigarette, participants completed a series of questions on the hand held computers. Questions included ones about place, activity, companionship, mood, and other subjective items. The hand held computers dated and time-stamped each entry. For inclusion in the data here, adolescents must have smoked at least two cigarettes during the 7-day baseline data collection period; 96 adolescents met this inclusion criterion.

The 96 adolescents began the study with varying amounts of cigarette smoking experience. Adolescents were divided into three groups based on their lifetime smoking levels: those who had smoked less than 6 cigarettes in their lifetimes ($n = 16$), representing very novice smokers; those who had smoked between 6 and 99 cigarettes in their lifetimes ($n = 46$), representing a group of irregular or experimental smokers; and those who had smoked 100 or more cigarettes during their lifetimes ($n = 34$), representing more regular smokers.

Immediately after smoking the cigarette, subjects turned on their hand-held computer to complete a variety of questions. They responded with an analog ladder-type scale, by moving a stylus to the appropriate point on the ladder scale (the scale is from 1 to 10). They were first asked about their moods and feelings "right now" (after smoking) and then about how they felt just before smoking. For each subject a change score (after - before) was calculated for a number of variables: positive feeling, negative feeling, feeling of tiredness, feeling of frustration and physiological feeling. The data for this experiment are available in the file `mood.csv`

Data description The data are in spreadsheet format with 498 rows, and 10 columns. The columns are:

- *Id* - subject id (from 1 to 96)
- *Study day* - sequential day in study (from 1 to 13)
- *Week day* - day of week (1=Monday, 2=Tuesday, $\cdots$, 7=Sunday)
- *Timebin* - timing of smoking event: 1 = "$3am - 8:59am$"; 2 = "$9am - 1:59pm'$"; 3 = "$2pm - 5:59pm'$"; 4 = "$6pm - 9:59pm$";5 = "$10pm - 2:59am$"
- *SmkLevel* - 1 =smoked less than 6 cigarettes in their lifetimes; 2 =smoked between 6 and 99 cigarettes in their lifetimes; 3 =smoked 100 or more cigarettes during their lifetimes
- *cposit* - change in reported positive affect (after smoking - before smoking)
- *cnegat* - change in reported negative affect (after smoking - before smoking)
- *ctired* - change in reported tiredness/boredom (after smoking - before smoking)

- *cfrust* - change in reported frustration (after smoking- before smoking)
- *cphys* - change in reported physiological sensations (after smoking- before smoking)

The main goal of this data analysis is to study the effect of smoking history on change in physiological sensations. So, unless otherwise specified, all computational questions refer to the data from the last response, *cphys*.

Data Reading

```
#data reading
mood=read.csv("mood.csv")
head(mood)
```

```
##   Id Day Weekday Timebin SmkLevel   cposit cnegat ctired cfrust cphys
## 1  1   1       4       2        3  1.33333   -3.0   -0.5    1.0  -4.0
## 2  1   1       4       3        3  1.33333    0.4    4.0    2.0   0.0
## 3  2   2       5       5        3 -0.66667    0.8    2.0    1.0   1.0
## 4  2   3       6       3        3 -1.66667    0.2   -0.5   -0.5   0.5
## 5  2   3       6       3        3  0.33333    1.4   -1.5    0.0  -1.5
## 6  2   6       2       4        3  1.00000   -1.2   -1.0    3.0   2.5
```

Possible Questions

Problem 1

The experimenters chose to use the trichotomous ordered smoking history variable and they have chose to not report smoking history as a continuous variable. Why do you think they did that? Do you think that this *information loss* might affect our analysis?

Problem 2

Do you think that the positive(*cposit*) and negative feelings(*cnegat*) represent just two opposite extremes in one single emotion-scale? How would you do some elementary analysis to answer this question?

Problem 3

Produce a visual summary of the change in physiological sensations for different smoking levels by looking at box-plots of subject-means. Report any notable observation and if possible, investigate them!

Problem 4

Fit a linear Gaussian model with fixed effects for smoking level and a random effect for subjects. Report the parameter estimates of the variance components and explain the meaning of the fitted model.

Problem 5

As an alternative, fit a model that allows for different between-subject variances in the three smoking-level groups. As before, explain the meaning of the fitted model. How would you assess the significance of the results.

Problem 6

What does the model in problem 5 tell you about the adequacy of the model you fit in problem 4? You can formally answer this using a likelihood ratio test.

Problem 7

What is the impact of non-normality? Describe in detail a model that that you think would be best for analyzing this data.

Problem 8

The histogram of `cphys` reveals a large mass around 0-which probably means that differences near 0 i.e. in the interval $[-.5, .5]$ are not that meaningful. Suppose, we binarize the `cphys` variable- i.e. we consider the difference to be significant if it's more than 0.5. Analyze whether there is any effect of past smoking behaviour on this binarized `cphys` variable? How will you compare this result with previous conclusions?