

SOC 756: Problem Set 4

Mateo Frumholtz

October 8, 2025

The github link for this class and assignment can be found [here](#).

Table 1 (Excel file on class web page and also copied below) includes the age-specific probabilities of starting smoking, quitting smoking, dying while being a smoker, and dying while being a nonsmoker for Italian men born in the 1950's. Note that “starting smoking” does not necessarily mean starting smoking for the first time.

Roughly 1,740,000 Italian boys born in 1955 survived to age 10. Assume that 98% of these boys were non-smokers at age 10, 2% were smokers at age 10, and all of the smokers at age 10 had been smoking continuously since they began smoking. Use the probabilities in Table 1 to calculate the l_{xi} , $1dx_{ij}$, $1Lx_{ij}$ columns of an increment decrement life table for these boys from age 10 to age 50. Note that you do not need matrix algebra or particularly complicated equations to do this. Then, use the table from Schoen 1988: p95 (handed out in class) to answer the following questions.

```
df <- ps5 |>
  mutate(
    nax = 0.5,
    people = case_when(
      age == 10 ~ 1740000,
      TRUE ~ NA
    ),
    lx_smoke = people * 0.02,
    lx_non_smokers = people * 0.98,
    dx_smoke_die = lx_smoke * smoke_dying,
    dx_smoke_nosmoke = lx_smoke * quit_smoke,
    dx_nosmoke_smoke = lx_non_smokers * start_smoke,
    dx_nosmoke_die = lx_non_smokers * non_smoke_dying
  )

for (i in 2:nrow(df)) {
  # lx_smoke = smokers - dying - quitting + starting smoking
```

```

df$lx_smoke[i] <- df$lx_smoke[i - 1] -
  df$dx_smoke_die[i - 1] -
  df$dx_smoke_nosmoke[i - 1] +
  df$dx_nosmoke_smoke[i - 1]

# lx_non_smokers = non_smokers - dying + quitting - starting smoking
df$lx_non_smokers[i] <- df$lx_non_smokers[i - 1] -
  df$dx_nosmoke_die[i - 1] -
  df$dx_nosmoke_smoke[i - 1] +
  df$dx_smoke_nosmoke[i - 1]

# dx_smoke_dead
df$dx_smoke_die[i] <- df$lx_smoke[i] * df$smoke_dying[i]

# dx_smoke_nosmoke
df$dx_smoke_nosmoke[i] <- df$lx_smoke[i] * df$quit_smoke[i]

# dx_nosmoke_smoke
df$dx_nosmoke_smoke[i] <- df$lx_non_smokers[i] * df$start_smoke[i]

# dx_nosmoke_die
df$dx_nosmoke_die[i] <- df$lx_non_smokers[i] * df$non_smoke_dying[i]
}

tr <- df |>
  mutate(
    Lx_smoke = (lx_smoke + lead(lx_smoke)) * 0.5,
    Lx_non_smokers = (lx_non_smokers + lead(lx_non_smokers)) * 0.5,
    mean_age_smokers = (nax + age) * (Lx_smoke / sum(Lx_smoke, na.rm = T)),
    mean_age_non_smokers = (nax + age) *
      (Lx_non_smokers / sum(Lx_non_smokers, na.rm = T))
  )

```

1. What was the probability that a boy alive at age 10 would have ever smoked by age 50?

$$\sum d_{smoke}/l(10)$$

```

# Summing those who transitioned to smokers and those who started
# off smoking at age 10 by total
(sum(tr$dx_nosmoke_smoke, na.rm = T) + tr$lx_smoke[1]) / tr$people[1]

```

```
[1] 0.6552879
```

0.655

2. How many years above age ten could a boy surviving to age 10

a. expect to be a smoker?

```
sum(tr$Lx_smoke, na.rm = T) / df$people[1]
```

```
[1] 14.47706
```

14.48 years

b. expect to be a non-smoker?

```
sum(tr$Lx_non_smokers, na.rm = T) / df$people[1]
```

```
[1] 24.30458
```

24.3 years

c. expect to live?

```
(sum(tr$Lx_smoke, na.rm = T) / df$people[1]) +  
  (sum(tr$Lx_non_smokers, na.rm = T) / df$people[1])
```

```
[1] 38.78165
```

38.8 years

3. Conditioning on persons under age 50 as you are doing, is the average age of smokers or non-smokers younger?

```
sum(tr$mean_age_smokers, na.rm = T)
```

```
[1] 32.46541
```

```
sum(tr$mean_age_non_smokers, na.rm = T)
```

```
[1] 28.01837
```

Nonsmokers (28 vs. 32 years)

4. Graph the age-specific probabilities of transitioning into smoking and out of smoking on the same figure. Is the graph consistent with your answer to question 3? Why or why not?

Not really. The graphs suggest that smokers should be younger for two reasons: one, the transition probabilities to become a smoker are higher at younger ages; two, the transition probabilities to become a nonsmoker (quitting) are higher at older ages, meaning we would expect non-smokers to be older.

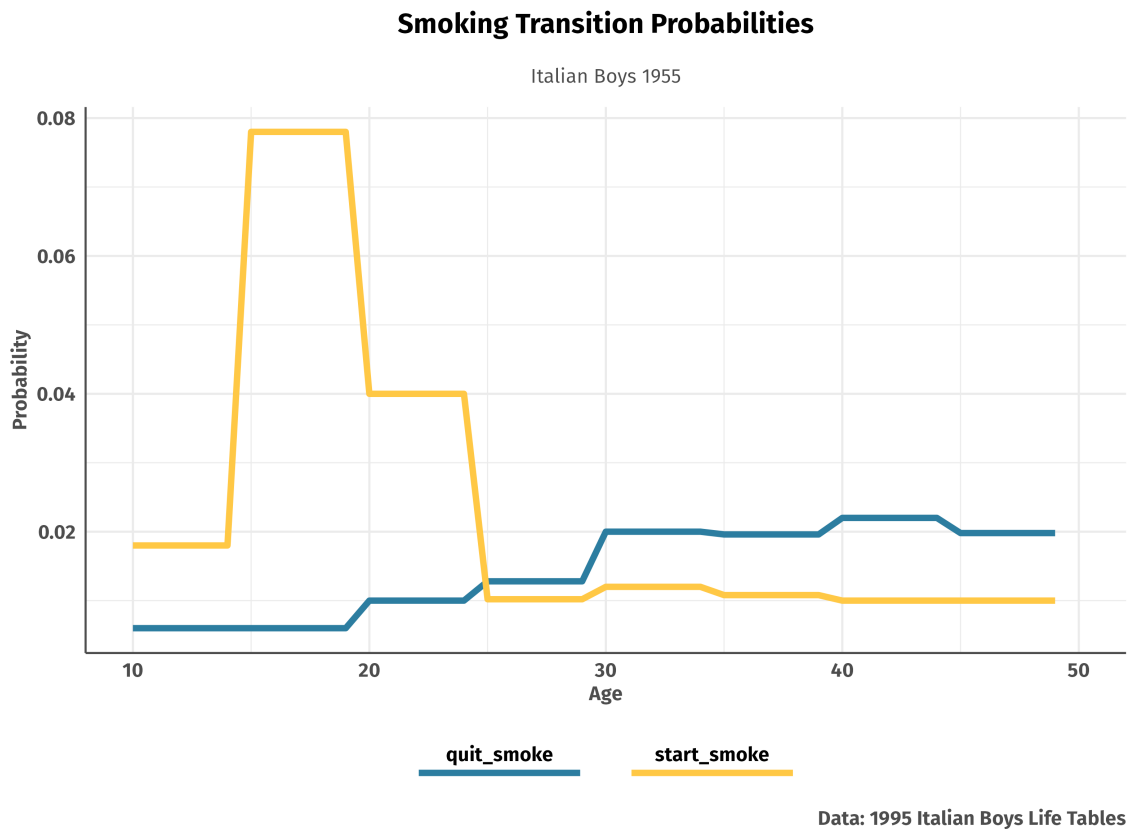


Figure 1: Probabilities of smoking transition.

5. Could you calculate the average duration of quitting spells (periods of time when those who once smoked were not smoking) using the life table you have created? Why or why not?

I don't think so. What we are looking for is to model equation 6 of the Shoen equations - probably something similar to the marriage one. We are missing some granular transition probabilities to be able to estimate it. We can imagine quitting as a state, where people move into it only from smoking, and then can move out to smoking or dying. So we need

to know the probability of smoking to non-smoking (which we know), and the probability of non-smoking to dying but having come from smoking (which we don't know).

6. Related: the increment-decrement life table assumes a homogenous application of transition probabilities to all persons in a given state at a given age. Why might this assumption be problematic when studying smoking - particularly when using the three-state system defined here? If you had better data, how might you improve your ability to model the smoking experiences of this cohort? (If helpful, note that the true cumulative conditional probability of ever smoking for Italian males born in 1955 was 0.53, per Federico et al. 2007 AJPH).

Smoking is a common addictive behavior that people are constantly transitioning in and out of in shorter time spans than a year. It's possible that people transition to quitting and starting more than once in a year, which we fail to capture. I think coming up with a better definition of what it means to be a smoker vs. not could help refine smoking experiences.

7. Generate two additional lifetables that condition on smoking status at age 10. Generate a figure that summarizes the expected duration in each state (smoking and non-smoking) by the smoking status of the children at age 10.

```
# Everyone is a smoker

df_smokers <- ps5 |>
  mutate(
    nax = 0.5,
    people = case_when(
      age == 10 ~ 1740000,
      TRUE ~ NA
    ),
    lx_smoke = people,
    lx_non_smokers = 0,
    dx_smoke_die = lx_smoke * smoke_dying,
    dx_smoke_nosmoke = lx_smoke * quit_smoke,
    dx_nosmoke_smoke = lx_non_smokers * start_smoke,
    dx_nosmoke_die = lx_non_smokers * non_smoke_dying
  )

for (i in 2:nrow(df_smokers)) {
  # lx_smoke = smokers - dying - quitting + starting smoking
  df_smokers$lx_smoke[i] <- df_smokers$lx_smoke[i - 1] -
    df_smokers$dx_smoke_die[i - 1] -
    df_smokers$dx_smoke_nosmoke[i - 1] +
    df_smokers$dx_nosmoke_smoke[i - 1]

  # lx_non_smokers = non_smokers - dying + quitting - starting smoking
```

```

df_smokers$lx_non_smokers[i] <- df_smokers$lx_non_smokers[i - 1] -
  df_smokers$dx_nosmoke_die[i - 1] -
  df_smokers$dx_nosmoke_smoke[i - 1] +
  df_smokers$dx_smoke_nosmoke[i - 1]

# dx_smoke_dead
df_smokers$dx_smoke_die[i] <- df_smokers$lx_smoke[i] *
  df_smokers$smoke_dying[i]

# dx_smoke_nosmoke
df_smokers$dx_smoke_nosmoke[i] <- df_smokers$lx_smoke[i] *
  df_smokers$quit_smoke[i]

# dx_nosmoke_smoke
df_smokers$dx_nosmoke_smoke[i] <- df_smokers$lx_non_smokers[i] *
  df_smokers$start_smoke[i]

# dx_nosmoke_die
df_smokers$dx_nosmoke_die[i] <- df_smokers$lx_non_smokers[i] *
  df_smokers$non_smoke_dying[i]
}

tr_smokers <- df_smokers |>
  mutate(
    Lx_smoke = (lx_smoke + lead(lx_smoke)) * 0.5,
    Lx_non_smokers = (lx_non_smokers + lead(lx_non_smokers)) * 0.5,
    mean_age_smokers = (nax + age) * (Lx_smoke / sum(Lx_smoke, na.rm = T)),
    mean_age_non_smokers = (nax + age) *
      (Lx_non_smokers / sum(Lx_non_smokers, na.rm = T)),
    Lx_smoke = case_when(is.na(Lx_smoke) ~ 0, TRUE ~ Lx_smoke),
    cumsum_smokers = rev(cumsum(rev(Lx_smoke))),
    duration_smoker = cumsum_smokers / people[1]
  )

# Everyone is a non-smoker

df_non_smokers <- ps5 |>
  mutate(
    nax = 0.5,
    people = case_when(
      age == 10 ~ 1740000,
      TRUE ~ NA
    ),
    lx_smoke = 0,

```

```

    lx_non_smokers = people,
    dx_smoke_die = lx_smoke * smoke_dying,
    dx_smoke_nosmoke = lx_smoke * quit_smoke,
    dx_nosmoke_smoke = lx_non_smokers * start_smoke,
    dx_nosmoke_die = lx_non_smokers * non_smoke_dying
  )
}

for (i in 2:nrow(df_non_smokers)) {
  # lx_smoke = smokers - dying - quitting + starting smoking
  df_non_smokers$lx_smoke[i] <- df_non_smokers$lx_smoke[i - 1] -
    df_non_smokers$dx_smoke_die[i - 1] -
    df_non_smokers$dx_smoke_nosmoke[i - 1] +
    df_non_smokers$dx_nosmoke_smoke[i - 1]

  # lx_non_smokers = non_smokers - dying + quitting - starting smoking
  df_non_smokers$lx_non_smokers[i] <- df_non_smokers$lx_non_smokers[i - 1] -
    df_non_smokers$dx_nosmoke_die[i - 1] -
    df_non_smokers$dx_nosmoke_smoke[i - 1] +
    df_non_smokers$dx_smoke_nosmoke[i - 1]

  # dx_smoke_dead
  df_non_smokers$dx_smoke_die[i] <- df_non_smokers$lx_smoke[i] *
    df_non_smokers$smoke_dying[i]

  # dx_smoke_nosmoke
  df_non_smokers$dx_smoke_nosmoke[i] <- df_non_smokers$lx_smoke[i] *
    df_non_smokers$quit_smoke[i]

  # dx_nosmoke_smoke
  df_non_smokers$dx_nosmoke_smoke[i] <- df_non_smokers$lx_non_smokers[i] *
    df_non_smokers$start_smoke[i]

  # dx_nosmoke_die
  df_non_smokers$dx_nosmoke_die[i] <- df_non_smokers$lx_non_smokers[i] *
    df_non_smokers$non_smoke_dying[i]
}

tr_non_smokers <- df_non_smokers |>
  mutate(
    Lx_smoke = (lx_smoke + lead(lx_smoke)) * 0.5,
    Lx_non_smokers = (lx_non_smokers + lead(lx_non_smokers)) * 0.5,
    mean_age_smokers = (nax + age) * (Lx_smoke / sum(Lx_smoke, na.rm = T)),
    mean_age_non_smokers = (nax + age) *
      (Lx_non_smokers / sum(Lx_non_smokers, na.rm = T)),

```

```

Lx_non_smokers = case_when(
  is.na(Lx_non_smokers) ~ 0,
  TRUE ~ Lx_non_smokers
),
cumsum_non_smokers = rev(cumsum(rev(Lx_non_smokers))),
duration_non_smoker = cumsum_non_smokers / people[1]
)

```

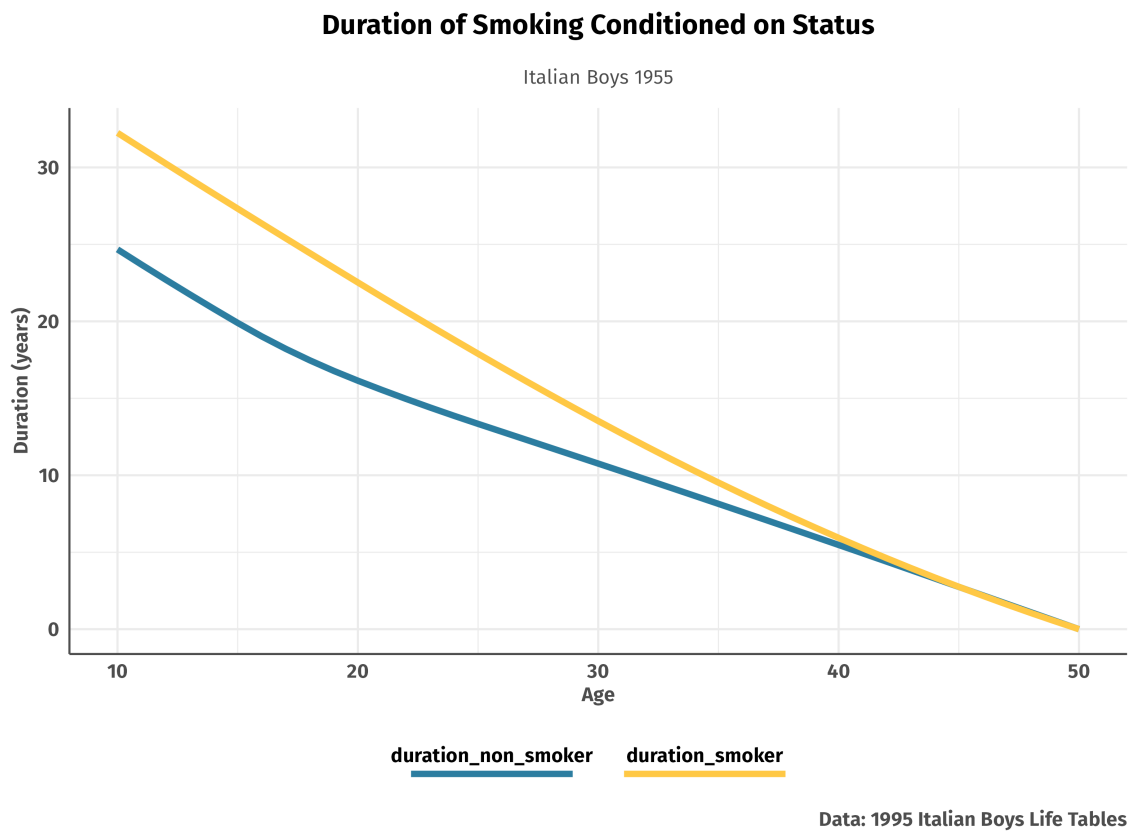


Figure 2: Duration of smoking status, conditioned on smoking status.