SP_DemTech2_Problem Set 4

Sara Peters

2023-10-11

###Soc 756

#Problem Set 4

#Increment Decrement Life Tables

Table 1 (on the class web page) includes the age-specific probabilities of:

- starting smoking (state 1) ***- quitting smoking*** (state 2) ***- dying while being a smoker*** (state 3) ***- dying while being a nonsmoker*** (state 4)

For Italian men born in the 1950's^a. 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 nonsmokers 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

lxi, 1dxij 1Lxij

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.

#set up

```
#Set working directory
setwd("C:/Users/saraa/OneDrive - UW-Madison/SOC 756- Demography Techniques II/Problem Sets/DemTe
ch2/Problem Set 4")

#load libraries
# install.packages("dplyr")
# install.packages("tidyverse")
# install.packages("ggplot2")
#install.packages("readxl")


library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:kableExtra':
##
##     group_rows
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ forcats   1.0.0     ✔ readr     2.1.4
## ✔ ggplot2   3.4.3     ✔ stringr   1.5.0
## ✔ lubridate 1.9.2     ✔ tibble    3.2.1
## ✔ purrr     1.0.1     ✔ tidyr     1.3.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter()     masks stats::filter()
## ✖ dplyr::group_rows() masks kableExtra::group_rows()
## ✖ dplyr::lag()        masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(ggplot2)
#library (readxl)

# Set the 'scipen' option to a large value to prevent scientific notation
options(scipen = 999)

# Set the 'digits' option to control the number of decimal places
options(digits = 6)  # Change the number to the desired decimal places
```

##Load data

```
#qxi values for italian boy's smoking behavior
IT_smoking <- read.csv("ps4_data_2023.csv")


#rename columns to simplify
IT_data <- IT_smoking %>%
  rename(qx_ss = qx_smoke, qx_ns = qx_quit, qx_nsd = qx_NSmortality, qx_sd = qx_Smortality, x =
Age)


#transform columns to numeric
columns_to_process <- c("x", "qx_ss", "qx_ns", "qx_nsd", "qx_sd")

for (col_name in columns_to_process) {
  IT_data [, col_name] <- as.numeric(IT_data [, col_name])
}

str(IT_data)
```

```
## 'data.frame':    41 obs. of  5 variables:
##  $ x     : num  10 11 12 13 14 15 16 17 18 19 ...
##  $ qx_ss : num  0.018 0.018 0.018 0.018 0.018 0.078 0.078 0.078 0.078 0.078 ...
##  $ qx_ns : num  0.006 0.006 0.006 0.006 0.006 0.006 0.006 0.006 0.006 0.006 ...
##  $ qx_nsd: num  0.000363 0.000351 0.000383 0.000457 0.000567 ...
##  $ qx_sd : num  0.000363 0.000351 0.000383 0.000457 0.000567 ...
```

##Create increment decrement table of male Italian smokers

```r
#assign radix values for total population and two subpopulations (s and ns). lx is the number of
individuals in each state i at exact age x

IT_data$lx[1] <- 100000
IT_data$lx_ns[1] <- 98000
IT_data$lx_s[1] <- 2000

#Calculate dx values for lx0, or the number of individuals moving from state i to state j betwee
n ages x and x + 1. Four flows total ns = 1, s = 2, nsd = 3, sd = 4.
#initializecolumn[populate first row] <- round the result of (first row of lx * first row of qx,
no decimals)

#transition from ns(1) to s(2), qx_ss (probabilty of starting smoking)
IT_data$dx_1_2[1] <- round(IT_data$lx_ns[1]* IT_data$qx_ss[1], digits = 0)

#transition from ns(1) to nsd(3), qx_nsd (probability of dying while nonsmoker)
IT_data$dx_1_3[1] <- round(IT_data$lx_ns[1]* IT_data$qx_nsd[1], digits = 0)

#transition from s(2) to ns(1), qx_ns (probability of quitting smoking)
IT_data$dx_2_1[1] <- round(IT_data$lx_s[1]* IT_data$qx_ns[1], digits = 0)

#transition from s(2) to sd(4), qx_sd (probabilty of dying while smoker)
IT_data$dx_2_4[1] <- round(IT_data$lx_s[1]* IT_data$qx_sd[1], digits = 0)


#Use a for loop to calculate the lx and dx values for all other ages
#notation for lx = lxi + 1 = lxi - dxij - dxv + dxji
#notation for dx = dxij = lxi * qxji

for(i in 2:nrow(IT_data)){
  IT_data$lx_ns[i] <- IT_data$lx_ns[i-1]-IT_data$dx_1_2[i-1]-IT_data$dx_1_3[i-1]+IT_data$dx_2_1
[i-1]
  IT_data$lx_s[i] <- IT_data$lx_s[i-1]-IT_data$dx_2_1[i-1]-IT_data$dx_2_4[i-1]+IT_data$dx_1_2[i-
1]
  IT_data$dx_1_2[i] <- round(IT_data$lx_ns[i]*IT_data$qx_ss[i], digits = 0)
  IT_data$dx_2_1[i] <- round(IT_data$lx_s[i]*IT_data$qx_ns[i], digits = 0)
  IT_data$dx_2_4[i] <- round(IT_data$lx_s[i]*IT_data$qx_sd[i], digits = 0)
  IT_data$dx_1_3[i] <- round(IT_data$lx_ns[i]*IT_data$qx_nsd[i], digits = 0)
  }

#Apply for loop to lx column for total population to check numbers

for(i in 2:nrow(IT_data)) {
IT_data$lx[i] <- IT_data$lx_ns[i]+IT_data$lx_s[i]
}

#Calculate LX values, person-years lived in state i between ages x and x + 1, two states lived
(smoker and nonsmoker), accounts for exits to death while in either state.
#notation Lxi = [lxi + lxi + 1]/2

# Initialize the columns with NA values
IT_data$Lx_ns <- rep(NA, nrow(IT_data))
```

```r
IT_data$Lx_s <- rep(NA, nrow(IT_data))
IT_data$Lx <- rep(NA, nrow(IT_data))

#run for loop to populate Lx values
for (i in 1:(nrow(IT_data) - 1)) {
  IT_data$Lx_ns[i] <- (IT_data$lx_ns[i] + IT_data$lx_ns[i + 1]) / 2
  IT_data$Lx_s[i] <- (IT_data$lx_s[i] + IT_data$lx_s[i + 1]) / 2
  IT_data$Lx[i] <- (IT_data$lx[i] + IT_data$lx[i + 1]) / 2
}

print(IT_data)
```

```
##     x  qx_ss  qx_ns      qx_nsd      qx_sd     lx lx_ns  lx_s dx_1_2 dx_1_3
## 1  10 0.0180 0.0060 0.000363480 0.00036348 100000 98000  2000   1764     36
## 2  11 0.0180 0.0060 0.000350770 0.00035077  99963 96212  3751   1732     34
## 3  12 0.0180 0.0060 0.000383240 0.00038324  99928 94469  5459   1700     36
## 4  13 0.0180 0.0060 0.000457050 0.00045705  99890 92766  7124   1670     42
## 5  14 0.0180 0.0060 0.000566960 0.00056696  99845 91097  8748   1640     52
## 6  15 0.0780 0.0060 0.000695670 0.00069567  99788 89457 10331   6978     62
## 7  16 0.0780 0.0060 0.000865080 0.00086508  99719 82479 17240   6433     71
## 8  17 0.0780 0.0060 0.001024450 0.00102445  99633 76078 23555   5934     78
## 9  18 0.0780 0.0060 0.001097010 0.00109701  99531 70207 29324   5476     77
## 10 19 0.0780 0.0060 0.001102460 0.00110246  99422 64830 34592   5057     71
## 11 20 0.0400 0.0100 0.000752794 0.00215084  99313 59910 39403   2396     45
## 12 21 0.0400 0.0100 0.000742980 0.00212280  99183 57863 41320   2315     43
## 13 22 0.0400 0.0100 0.000763238 0.00218068  99052 55918 43134   2237     43
## 14 23 0.0400 0.0100 0.000779933 0.00222838  98915 54069 44846   2163     42
## 15 24 0.0400 0.0100 0.000774312 0.00221232  98773 52312 46461   2092     41
## 16 25 0.0102 0.0128 0.000743092 0.00212312  98629 50644 47985    517     38
## 17 26 0.0102 0.0128 0.000694890 0.00198540  98489 50703 47786    517     35
## 18 27 0.0102 0.0128 0.000661570 0.00189020  98359 50763 47596    518     34
## 19 28 0.0102 0.0128 0.000673008 0.00192288  98235 50820 47415    518     34
## 20 29 0.0102 0.0128 0.000710010 0.00202860  98110 50875 47235    519     36
## 21 30 0.0120 0.0200 0.000778337 0.00333573  97978 50925 47053    611     40
## 22 31 0.0120 0.0200 0.000843234 0.00361386  97781 51215 46566    615     43
## 23 32 0.0120 0.0200 0.000912198 0.00390942  97570 51488 46082    618     47
## 24 33 0.0120 0.0200 0.000957894 0.00410526  97343 51745 45598    621     50
## 25 34 0.0120 0.0200 0.001017562 0.00436098  97106 51986 45120    624     53
## 26 35 0.0108 0.0196 0.001063223 0.00455667  96856 52211 44645    564     56
## 27 36 0.0108 0.0196 0.001116885 0.00478665  96597 52466 44131    567     59
## 28 37 0.0108 0.0196 0.001204854 0.00516366  96327 52705 43622    569     64
## 29 38 0.0108 0.0196 0.001342397 0.00575313  96038 52927 43111    572     71
## 30 39 0.0108 0.0196 0.001515549 0.00649521  95719 53129 42590    574     81
## 31 40 0.0100 0.0220 0.001700384 0.00728736  95361 53309 42052    533     91
## 32 41 0.0100 0.0220 0.001919204 0.00822516  94964 53610 41354    536    103
## 33 42 0.0100 0.0220 0.002191042 0.00939018  94521 53881 40640    539    118
## 34 43 0.0100 0.0220 0.002394728 0.01026312  94021 54118 39903    541    130
## 35 44 0.0100 0.0220 0.002652783 0.01136907  93481 54325 39156    543    144
## 36 45 0.0100 0.0198 0.002990225 0.01281525  92892 54499 38393    545    163
## 37 46 0.0100 0.0198 0.003315557 0.01420953  92237 54551 37686    546    181
## 38 47 0.0100 0.0198 0.003696819 0.01584351  91520 54570 36950    546    202
## 39 48 0.0100 0.0198 0.004169879 0.01787091  90733 54554 36179    546    227
## 40 49 0.0100 0.0198 0.004615513 0.01978077  89859 54497 35362    545    252
## 41 50     NA     NA          NA         NA  88908 54400 34508     NA     NA
##    dx_2_1 dx_2_4   Lx_ns    Lx_s      Lx
## 1      12      1 97106.0  2875.5 99981.5
## 2      23      1 95340.5  4605.0 99945.5
## 3      33      2 93617.5  6291.5 99909.0
## 4      43      3 91931.5  7936.0 99867.5
## 5      52      5 90277.0  9539.5 99816.5
## 6      62      7 85968.0 13785.5 99753.5
## 7     103     15 79278.5 20397.5 99676.0
## 8     141     24 73142.5 26439.5 99582.0
## 9     176     32 67518.5 31958.0 99476.5
```

```
## 10    208     38 62370.0 36997.5 99367.5
## 11    394     85 58886.5 40361.5 99248.0
## 12    413     88 56890.5 42227.0 99117.5
## 13    431     94 54993.5 43990.0 98983.5
## 14    448    100 53190.5 45653.5 98844.0
## 15    465    103 51478.0 47223.0 98701.0
## 16    614    102 50673.5 47885.5 98559.0
## 17    612     95 50733.0 47691.0 98424.0
## 18    609     90 50791.5 47505.5 98297.0
## 19    607     91 50847.5 47325.0 98172.5
## 20    605     96 50900.0 47144.0 98044.0
## 21    941    157 51070.0 46809.5 97879.5
## 22    931    168 51351.5 46324.0 97675.5
## 23    922    180 51616.5 45840.0 97456.5
## 24    912    187 51865.5 45359.0 97224.5
## 25    902    197 52098.5 44882.5 96981.0
## 26    875    203 52338.5 44388.0 96726.5
## 27    865    211 52585.5 43876.5 96462.0
## 28    855    225 52816.0 43366.5 96182.5
## 29    845    248 53028.0 42850.5 95878.5
## 30    835    277 53219.0 42321.0 95540.0
## 31    925    306 53459.5 41703.0 95162.5
## 32    910    340 53745.5 40997.0 94742.5
## 33    894    382 53999.5 40271.5 94271.0
## 34    878    410 54221.5 39529.5 93751.0
## 35    861    445 54412.0 38774.5 93186.5
## 36    760    492 54525.0 38039.5 92564.5
## 37    746    536 54560.5 37318.0 91878.5
## 38    732    585 54562.0 36564.5 91126.5
## 39    716    647 54525.5 35770.5 90296.0
## 40    700    699 54448.5 34935.0 89383.5
## 41     NA     NA      NA      NA      NA
```

#Problem Set 4 Questions

#use the table from Schoen 1988 p95 (in powerpoint) to answer the following questions.

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

```
#Probability a person aged x will ever by in state ij
#notation Sum dx nonsmoking to smoking + smokers before age 10/l0, sum dx_1_2+ l0(smokers)/l0

Q1 = (sum(IT_data$dx_1_2[1:40]) + IT_data$lx_s[1])/IT_data$lx[1]

print(Q1)
```

```
## [1] 0.65531
```

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

```
#Proporation of life spent in state i
#Schoen notation Ti(0)/T(0)
#Notation using lx and Lx columns sum of Lx/l0

#a. expect to be a smoker?
#notation sum of Lx_smokers/l0
Q2a = sum(IT_data$Lx_s[1:40])/IT_data$lx[1]



#b. expect to be a non-smoker?
#Notation sum of Lx_nonsmokers/l0
Q2b = sum(IT_data$Lx_ns[1:40])/IT_data$lx[1]


#c. expect to live?
#notation sum of LX/l0
Q2c = sum(IT_data$Lx[1:40])/IT_data$lx[1]

print(Q2a)
```

```
## [1] 14.4775
```

```
print(Q2b)
```

```
## [1] 24.3038
```

```
print(Q2c)
```

```
## [1] 38.7814
```

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

Average nonsmokers are younger.

```
#mean age of persons in state i, person-years lived in state i between 10-50,
#notation Sum(x+.5*n) * Li/ Sum(Li)

#3a- average age of smokers,
#notation (sum(x + 0.5) * Lx_s/sumLx_s)
Q3a = (sum((IT_data$x[1:40] + 0.5)*IT_data$Lx_s[1:40]))/sum(IT_data$Lx_s[1:40])

#3b- average age of nonsmokers,
#notation (sum(x + 0.5) * Lx_ns/sumLx_ns)
Q3b = (sum((IT_data$x[1:40] + 0.5)* IT_data$Lx_ns[1:40]))/sum(IT_data$Lx_ns[1:40])


print(Q3a)
```

```
## [1] 32.4656
```

```
print(Q3b)
```

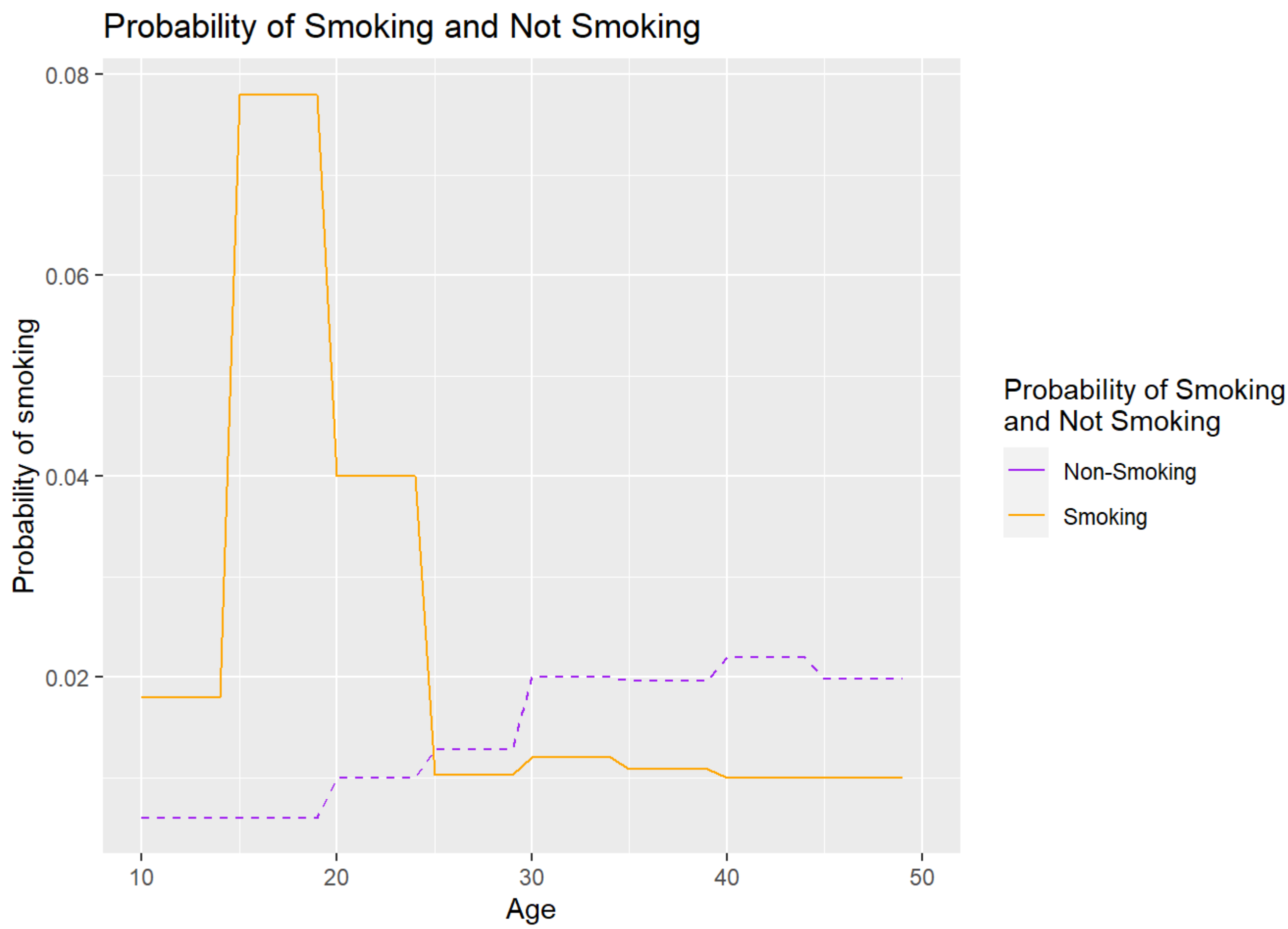
```
## [1] 28.018
```

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?

The graph is not consistent with my answers to question 3, the plot indicates that Italian males tend to be smokers at younger ages and nonsmokers at older ages. This might be because the formula used to calculate question 3 considers the ***average years spent*** in a state without paying attention to the age-specific distribution of those ***probabilities*** (what we consider in question 4). The plot below indicates that the age-specific distribution of the probability of starting smoking is squewed towards younger ages but also quite high compared with the probability of not smoking which is much lower until later in life. This indicates that the age structure of the population should be considered.

```
#create a plot of age specific probability of smoking and not smoking between ages 10 and 50 for
Italian males



IT_data %>%
  ggplot(aes(x = x)) +
  geom_line(aes(y = qx_ss, color = "Smoking"), linetype = "solid") +
  geom_line(aes(y = qx_ns, color = "Non-Smoking"), linetype = "dashed") +
  ylab("Probability of smoking") +
  xlab("Age") +
  scale_linetype_manual(name = "Probability of Smoking\nand Not Smoking", values = c("solid", "d
ashed")) +
  scale_color_manual(name = "Probability of Smoking\nand Not Smoking", values = c("purple", "ora
nge")) +
  labs(color = "Line Type") +
  ggtitle("Probability of Smoking and Not Smoking")
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
## Removed 1 row containing missing values (`geom_line()`).
```



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?

No. To calculate the duration of quitting spells you would need more logitundinal data and would likely need probabilities for a interval of less than one year. There is an equation in Schoen (1988) that it seems like you could use (equation 6) but a) you need the T value for this equation and b) The T value would be an unconditional value, it would include those who have never smoked and those who have smoked and quit.

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).

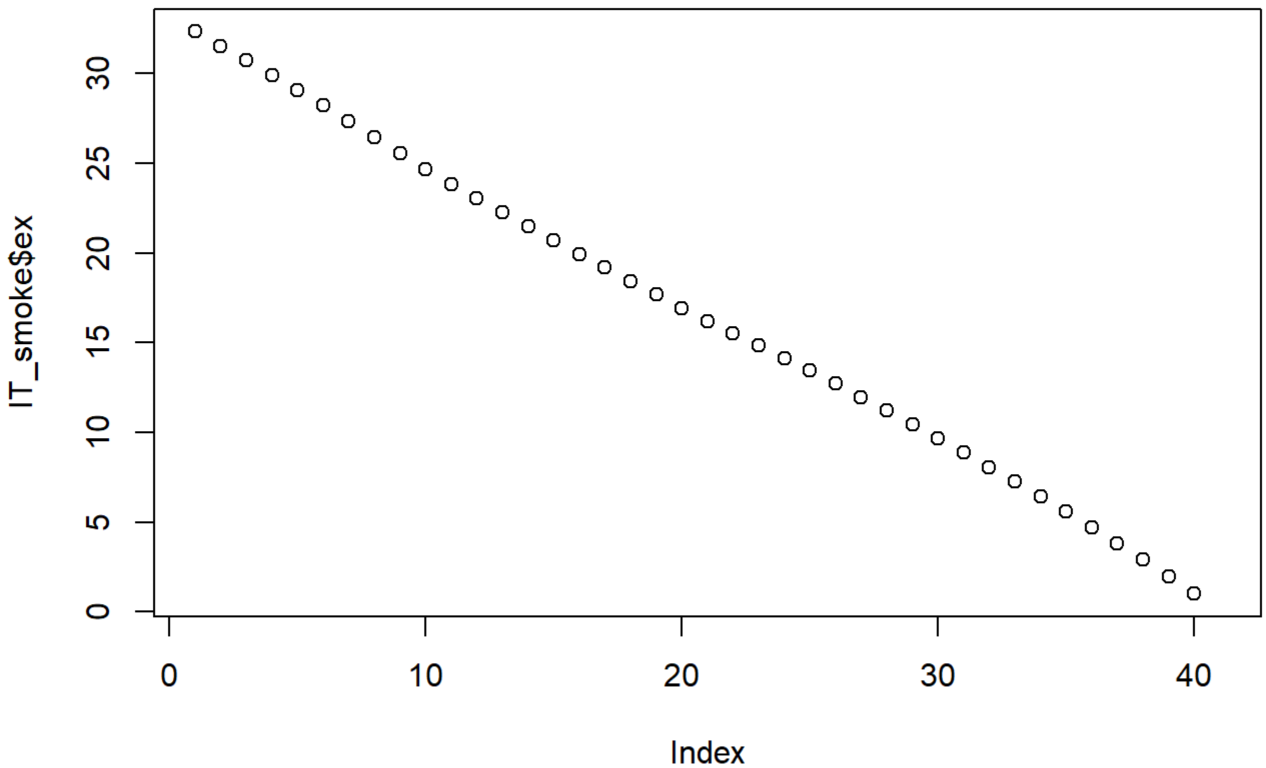
This is problematic when studying smoking because the unconditional probability of smoking doesn't account for a history of smoking in the past which would potentially increase your likelihood of smoking again in the future. A conditional probability for starting smoking for people who have smoked in the past would be a more realistic

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.

##Expected duration of smoking 10-50 years of age

```
#Conditional expectancies (eijx): average # of years spent in state j above age x, for individua
ls in state i at age x.

#expected duration of smoking for children aged 10 is the average  # of years spent smoking abov
e age 10 for individuals smoking at age 10.

#need ex values for probability of exit while smoking.

#create new dataset

IT_smoke <- IT_data %>%
  select(x, qx_ss, qx_ns, qx_nsd, qx_sd)

# Initialize the variables with appropriate initial values
IT_smoke$lx_ns <- numeric(nrow(IT_smoke))
IT_smoke$lx_s <- numeric(nrow(IT_smoke))
IT_smoke$dx_1_2 <- numeric(nrow(IT_smoke))
IT_smoke$dx_2_1 <- numeric(nrow(IT_smoke))
IT_smoke$dx_2_4 <- numeric(nrow(IT_smoke))
IT_smoke$dx_1_3 <- numeric(nrow(IT_smoke))

#assign radix values for total population and two subpopulations (s and ns). lx is the number of
individuals in each state i at exact age x


IT_smoke$lx_ns[1] <- 0
IT_smoke$lx_s[1] <- 2000

#Calculate dx values for lx0, or the number of individuals moving from state i to state j betwee
n ages x and x + 1. Four flows total ns = 1, s = 2, nsd = 3, sd = 4.
#initializecolumn[populate first row] <- round the result of (first row of lx * first row of qx,
5 decimals)

#transition from ns(1) to s(2), qx_ss (probabilty of starting smoking)
IT_smoke$dx_1_2[1] <- round(IT_smoke$lx_ns[1]* IT_smoke$qx_ss[1], digits = 5)

#transition from ns(1) to nsd(3), qx_nsd (probability of dying while nonsmoker)
IT_smoke$dx_1_3[1] <- round(IT_smoke$lx_ns[1]* IT_smoke$qx_nsd[1], digits = 5)

#transition from s(2) to ns(1), qx_ns (probability of quitting smoking)
IT_smoke$dx_2_1[1] <- round(IT_smoke$lx_s[1]* IT_smoke$qx_ns[1], digits = 5)

#transition from s(2) to sd(4), qx_sd (probabilty of dying while smoker)
IT_smoke$dx_2_4[1] <- round(IT_smoke$lx_s[1]* IT_smoke$qx_sd[1], digits = 5)


#Use a for loop to calculate the lx and dx values for all other ages
#notation for lx = lxi + 1 = lxi - dxij - dxv + dxji
#notation for dx = dxij = lxi * qxji

for(i in 2:nrow(IT_smoke)){
  IT_smoke$lx_ns[i] <- IT_smoke$lx_ns[i-1]-IT_smoke$dx_1_2[i-1]-IT_smoke$dx_1_3[i-1]+IT_smoke$dx
```

```
_2_1[i-1]
  IT_smoke$lx_s[i] <- IT_smoke$lx_s[i- 1]-IT_smoke$dx_2_1[i-1]-IT_smoke$dx_2_4[i-1]+IT_smoke$dx_
1_2[i-1]
  IT_smoke$dx_1_2[i] <- round(IT_smoke$lx_ns[i]*IT_smoke$qx_ss[i], digits = 5)
  IT_smoke$dx_2_1[i] <- round(IT_smoke$lx_s[i]*IT_smoke$qx_ns[i], digits = 5)
  IT_smoke$dx_2_4[i] <- round(IT_smoke$lx_s[i]*IT_smoke$qx_sd[i], digits = 5)
  IT_smoke$dx_1_3[i] <- round(IT_smoke$lx_ns[i]*IT_smoke$qx_nsd[i], digits = 5)
}


# Initialize Lx columns with NA values
IT_smoke$Lx_s <- rep(NA, nrow(IT_smoke))

#run for loop to populate Lx values
for (i in 1:(nrow(IT_smoke) - 1)) {
  IT_smoke$Lx_s[i] <- round(((IT_smoke$lx_s[i] + IT_smoke$lx_s[i + 1]) / 2), digits = 5)
}

# Create Tx column from Lx value
IT_smoke$Tx[1] <- sum(IT_smoke$Lx_s[1:40])
for (i in 2:nrow(IT_smoke)) {
  IT_smoke$Tx[i] <- round(IT_smoke$Tx[i-1] - IT_smoke$Lx_s[i-1], digits = 5)
}

IT_smoke$ex_s <- round(IT_smoke$Tx / IT_smoke$Lx_s, digits = 5)

plot(IT_smoke$ex)
```




```
print(IT_smoke)
```

```
##     x  qx_ss  qx_ns      qx_nsd      qx_sd    lx_ns    lx_s  dx_1_2  dx_2_1
## 1  10 0.0180 0.0060 0.000363480 0.00036348   0.0000 2000.00 0.00000 12.0000
## 2  11 0.0180 0.0060 0.000350770 0.00035077  12.0000 1987.27 0.21600 11.9236
## 3  12 0.0180 0.0060 0.000383240 0.00038324  23.7034 1974.87 0.42666 11.8492
## 4  13 0.0180 0.0060 0.000457050 0.00045705  35.1169 1962.69 0.63210 11.7761
## 5  14 0.0180 0.0060 0.000566960 0.00056696  46.2449 1950.65 0.83241 11.7039
## 6  15 0.0780 0.0060 0.000695670 0.00069567  57.0901 1938.67 4.45303 11.6320
## 7  16 0.0780 0.0060 0.000865080 0.00086508  64.2294 1930.14 5.00989 11.5809
## 8  17 0.0780 0.0060 0.001024450 0.00102445  70.7448 1921.90 5.51810 11.5314
## 9  18 0.0780 0.0060 0.001097010 0.00109701  76.6857 1913.92 5.98148 11.4835
## 10 19 0.0780 0.0060 0.001102460 0.00110246  82.1036 1906.32 6.40408 11.4379
## 11 20 0.0400 0.0100 0.000752794 0.00215084  87.0469 1899.18 3.48188 18.9918
## 12 21 0.0400 0.0100 0.000742980 0.00212280 102.4913 1879.59 4.09965 18.7959
## 13 22 0.0400 0.0100 0.000763238 0.00218068 117.1114 1860.90 4.68446 18.6090
## 14 23 0.0400 0.0100 0.000779933 0.00222838 130.9466 1842.92 5.23786 18.4292
## 15 24 0.0400 0.0100 0.000774312 0.00221232 144.0358 1825.62 5.76143 18.2562
## 16 25 0.0102 0.0128 0.000743092 0.00212312 156.4190 1809.09 1.59547 23.1563
## 17 26 0.0102 0.0128 0.000694890 0.00198540 177.8636 1783.69 1.81421 22.8312
## 18 27 0.0102 0.0128 0.000661570 0.00189020 198.7570 1759.13 2.02732 22.5168
## 19 28 0.0102 0.0128 0.000673008 0.00192288 219.1150 1735.31 2.23497 22.2120
## 20 29 0.0102 0.0128 0.000710010 0.00202860 238.9446 1712.00 2.43723 21.9136
## 21 30 0.0120 0.0200 0.000778337 0.00333573 258.2513 1689.05 3.09902 33.7810
## 22 31 0.0120 0.0200 0.000843234 0.00361386 288.7323 1652.73 3.46479 33.0547
## 23 32 0.0120 0.0200 0.000912198 0.00390942 318.0787 1617.17 3.81694 32.3434
## 24 33 0.0120 0.0200 0.000957894 0.00410526 346.3150 1582.32 4.15578 31.6464
## 25 34 0.0120 0.0200 0.001017562 0.00436098 373.4739 1548.34 4.48169 30.9667
## 26 35 0.0108 0.0196 0.001063223 0.00455667 399.5789 1515.10 4.31545 29.6959
## 27 36 0.0108 0.0196 0.001116885 0.00478665 424.5346 1482.81 4.58497 29.0632
## 28 37 0.0108 0.0196 0.001204854 0.00516366 448.5386 1451.24 4.84422 28.4443
## 29 38 0.0108 0.0196 0.001342397 0.00575313 471.5982 1420.14 5.09326 27.8348
## 30 39 0.0108 0.0196 0.001515549 0.00649521 493.7067 1389.23 5.33203 27.2290
## 31 40 0.0100 0.0220 0.001700384 0.00728736 514.8554 1358.31 5.14855 29.8829
## 32 41 0.0100 0.0220 0.001919204 0.00822516 538.7143 1323.68 5.38714 29.1210
## 33 42 0.0100 0.0220 0.002191042 0.00939018 561.4142 1289.06 5.61414 28.3593
## 34 43 0.0100 0.0220 0.002394728 0.01026312 582.9292 1254.21 5.82929 27.5926
## 35 44 0.0100 0.0220 0.002652783 0.01136907 603.2966 1219.57 6.03297 26.8306
## 36 45 0.0100 0.0198 0.002990225 0.01281525 622.4938 1184.91 6.22494 23.4612
## 37 46 0.0100 0.0198 0.003315557 0.01420953 637.8687 1152.49 6.37869 22.8193
## 38 47 0.0100 0.0198 0.003696819 0.01584351 652.1944 1119.67 6.52194 22.1695
## 39 48 0.0100 0.0198 0.004169879 0.01787091 665.4309 1086.28 6.65431 21.5084
## 40 49 0.0100 0.0198 0.004615513 0.01978077 677.5103 1052.02 6.77510 20.8300
## 41 50     NA     NA          NA         NA 688.4381 1017.15      NA      NA
##      dx_2_4  dx_1_3    Lx_s       Tx     ex_s
## 1   0.72696 0.00000 1993.64 64490.78 32.34831
## 2   0.69708 0.00421 1981.07 62497.14 31.54715
## 3   0.75685 0.00908 1968.78 60516.07 30.73787
## 4   0.89705 0.01605 1956.67 58547.29 29.92193
## 5   1.10594 0.02622 1944.66 56590.62 29.10054
## 6   1.34867 0.03972 1934.41 54645.96 28.24947
## 7   1.66973 0.05556 1926.02 52711.56 27.36809
## 8   1.96889 0.07247 1917.91 50785.54 26.47961
## 9   2.09959 0.08412 1910.12 48867.62 25.58355
```

```
## 10  2.10164 0.09052 1902.75 46957.51 24.67875
## 11  4.08484 0.06553 1889.39 45054.76 23.84625
## 12  3.98999 0.07615 1870.24 43165.37 23.08006
## 13  4.05803 0.08938 1851.91 41295.13 22.29866
## 14  4.10672 0.10213 1834.27 39443.21 21.50349
## 15  4.03886 0.11153 1817.35 37608.94 20.69434
## 16  3.84091 0.11623 1796.39 35791.59 19.92421
## 17  3.54133 0.12360 1771.41 33995.20 19.19108
## 18  3.32510 0.13149 1747.22 32223.80 18.44290
## 19  3.33680 0.14747 1723.66 30476.58 17.68136
## 20  3.47296 0.16965 1700.52 28752.92 16.90827
## 21  5.63421 0.20101 1670.89 27052.40 16.19040
## 22  5.97275 0.24347 1634.95 25381.51 15.52431
## 23  6.32220 0.29015 1599.75 23746.55 14.84395
## 24  6.49584 0.33173 1565.33 22146.81 14.14834
## 25  6.75226 0.38003 1531.72 20581.48 13.43687
## 26  6.90380 0.42484 1498.96 19049.76 12.70868
## 27  7.09771 0.47416 1467.03 17550.80 11.96353
## 28  7.49370 0.54042 1435.69 16083.78 11.20281
## 29  8.17028 0.63307 1404.69 14648.09 10.42800
## 30  9.02336 0.74824 1373.77 13243.40  9.64017
## 31  9.89851 0.87545 1341.00 11869.63  8.85135
## 32 10.88748 1.03390 1306.37 10528.63  8.05946
## 33 12.10449 1.23008 1271.63  9222.26  7.25230
## 34 12.87209 1.39596 1236.89  7950.63  6.42791
## 35 13.86541 1.60041 1202.24  6713.74  5.58435
## 36 15.18492 1.86140 1168.70  5511.50  4.71592
## 37 16.37633 2.11489 1136.08  4342.80  3.82261
## 38 17.73953 2.41104 1102.98  3206.72  2.90732
## 39 19.41290 2.77477 1069.15  2103.74  1.96767
## 40 20.80972 3.12706 1034.59  1034.59  1.00000
## 41       NA      NA      NA     0.00       NA
```

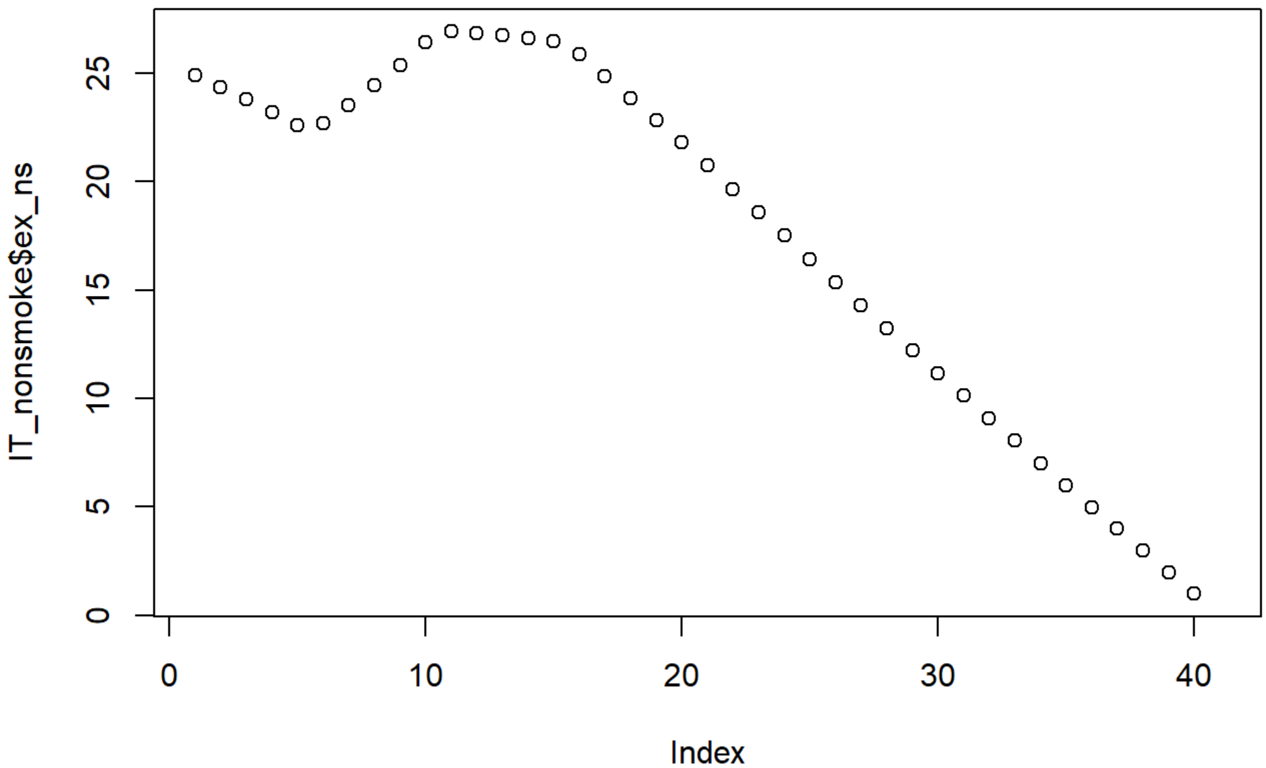
```
print(IT_smoke$ex[1])
```

```
## [1] 32.3483
```

##Expected duration of nonsmoking 10-50 years of age

```
#Conditional expectancies (eijx): average # of years spent in state j above age x, for individua
ls in state i at age x.

#expected duration of nonsmoking for children aged 10 is the average  # of years spent smoking a
bove age 10 for individuals not smoking at age 10.

#need ex values for probability of exit while nonsmoking.

#create new dataset

IT_nonsmoke <- IT_data %>%
  select(x, qx_ss, qx_ns, qx_nsd, qx_sd)

# Initialize the variables with appropriate initial values
IT_nonsmoke$lx_ns <- numeric(nrow(IT_nonsmoke))
IT_nonsmoke$lx_s <- numeric(nrow(IT_nonsmoke))
IT_nonsmoke$dx_1_2 <- numeric(nrow(IT_nonsmoke))
IT_nonsmoke$dx_2_1 <- numeric(nrow(IT_nonsmoke))
IT_nonsmoke$dx_2_4 <- numeric(nrow(IT_nonsmoke))
IT_nonsmoke$dx_1_3 <- numeric(nrow(IT_nonsmoke))

#assign radix values for total population and two subpopulations (s and ns). lx is the number of
individuals in each state i at exact age x


IT_nonsmoke$lx_ns[1] <- 98000
IT_nonsmoke$lx_s[1] <- 0

#Calculate dx values for lx0, or the number of individuals moving from state i to state j betwee
n ages x and x + 1. Four flows total ns = 1, s = 2, nsd = 3, sd = 4.
#initializecolumn[populate first row] <- round the result of (first row of lx * first row of qx,
5 decimals)

#transition from ns(1) to s(2), qx_ss (probabilty of starting smoking)
IT_nonsmoke$dx_1_2[1] <- round(IT_nonsmoke$lx_ns[1]* IT_nonsmoke$qx_ss[1], digits = 5)

#transition from ns(1) to nsd(3), qx_nsd (probability of dying while nonsmoker)
IT_nonsmoke$dx_1_3[1] <- round(IT_nonsmoke$lx_ns[1]* IT_nonsmoke$qx_nsd[1], digits = 5)

#transition from s(2) to ns(1), qx_ns (probability of quitting smoking)
IT_nonsmoke$dx_2_1[1] <- round(IT_nonsmoke$lx_s[1]* IT_nonsmoke$qx_ns[1], digits = 5)

#transition from s(2) to sd(4), qx_sd (probabilty of dying while smoker)
IT_nonsmoke$dx_2_4[1] <- round(IT_nonsmoke$lx_s[1]* IT_nonsmoke$qx_sd[1], digits = 5)


#Use a for loop to calculate the lx and dx values for all other ages
#notation for lx = lxi + 1 = lxi - dxij - dxv + dxji
#notation for dx = dxij = lxi * qxji

for(i in 2:nrow(IT_nonsmoke)){
  IT_nonsmoke$lx_ns[i] <- IT_nonsmoke$lx_ns[i-1]-IT_nonsmoke$dx_1_2[i-1]-IT_nonsmoke$dx_1_3[i-1]
```

```
+IT_nonsmoke$dx_2_1[i-1]
  IT_nonsmoke$lx_s[i] <- IT_nonsmoke$lx_s[i-1]-IT_nonsmoke$dx_2_1[i-1]-IT_nonsmoke$dx_2_4[i-1]+I
T_nonsmoke$dx_1_2[i-1]
  IT_nonsmoke$dx_1_2[i] <- round(IT_nonsmoke$lx_ns[i]*IT_nonsmoke$qx_ss[i], digits = 5)
  IT_nonsmoke$dx_2_1[i] <- round(IT_nonsmoke$lx_s[i]*IT_nonsmoke$qx_ns[i], digits = 5)
  IT_nonsmoke$dx_2_4[i] <- round(IT_nonsmoke$lx_s[i]*IT_nonsmoke$qx_sd[i], digits = 5)
  IT_nonsmoke$dx_1_3[i] <- round(IT_nonsmoke$lx_ns[i]*IT_nonsmoke$qx_nsd[i], digits = 5)
}



# Initialize Lx columns with NA values
IT_nonsmoke$Lx_ns <- rep(NA, nrow(IT_smoke))

#run for loop to populate Lx values
for (i in 1:(nrow(IT_nonsmoke) - 1)) {
  IT_nonsmoke$Lx_ns[i] <- round(((IT_nonsmoke$lx_ns[i] + IT_nonsmoke$lx_ns[i + 1]) / 2), digits
= 5)
}


# Create Tx column from Lx value
IT_nonsmoke$Tx[1] <- sum(IT_nonsmoke$Lx_ns[1:40])
for (i in 2:nrow(IT_nonsmoke)) {
  IT_nonsmoke$Tx[i] <- round(IT_nonsmoke$Tx[i-1] - IT_nonsmoke$Lx_ns[i-1], digits = 5)
}

IT_nonsmoke$ex_ns <- round(IT_nonsmoke$Tx / IT_nonsmoke$Lx_ns, digits = 5)

plot(IT_nonsmoke$ex_ns)
```




```
print(IT_nonsmoke)
```

```
##     x  qx_ss  qx_ns      qx_nsd      qx_sd   lx_ns     lx_s   dx_1_2   dx_2_1
## 1  10 0.0180 0.0060 0.000363480 0.00036348 98000.0     0.00 1764.000   0.0000
## 2  11 0.0180 0.0060 0.000350770 0.00035077 96200.4  1764.00 1731.607  10.5840
## 3  12 0.0180 0.0060 0.000383240 0.00038324 94445.6  3484.40 1700.021  20.9064
## 4  13 0.0180 0.0060 0.000457050 0.00045705 92730.3  5162.18 1669.145  30.9731
## 5  14 0.0180 0.0060 0.000566960 0.00056696 91049.7  6798.00 1638.895  40.7880
## 6  15 0.0780 0.0060 0.000695670 0.00069567 89400.0  8392.25 6973.201  50.3535
## 7  16 0.0780 0.0060 0.000865080 0.00086508 82415.0 15309.26 6428.368  91.8556
## 8  17 0.0780 0.0060 0.001024450 0.00102445 76007.2 21632.53 5928.559 129.7952
## 9  18 0.0780 0.0060 0.001097010 0.00109701 70130.5 27409.13 5470.182 164.4548
## 10 19 0.0780 0.0060 0.001102460 0.00110246 64747.9 32684.79 5050.335 196.1087
## 11 20 0.0400 0.0100 0.000752794 0.00215084 59822.3 37502.98 2392.891 375.0298
## 12 21 0.0400 0.0100 0.000742980 0.00212280 57759.4 39440.18 2310.375 394.4018
## 13 22 0.0400 0.0100 0.000763238 0.00218068 55800.5 41272.43 2232.020 412.7243
## 14 23 0.0400 0.0100 0.000779933 0.00222838 53938.6 43001.72 2157.544 430.0172
## 15 24 0.0400 0.0100 0.000774312 0.00221232 52169.0 44633.43 2086.760 446.3343
## 16 25 0.0102 0.0128 0.000743092 0.00212312 50488.2 46175.11  514.980 591.0414
## 17 26 0.0102 0.0128 0.000694890 0.00198540 50526.7 46001.01  515.373 588.8130
## 18 27 0.0102 0.0128 0.000661570 0.00189020 50565.1 45836.24  515.764 586.7039
## 19 28 0.0102 0.0128 0.000673008 0.00192288 50602.5 45678.66  516.146 584.6869
## 20 29 0.0102 0.0128 0.000710010 0.00202860 50637.0 45522.29  516.498 582.6852
## 21 30 0.0120 0.0200 0.000778337 0.00333573 50667.3 45363.75  608.007 907.2750
## 22 31 0.0120 0.0200 0.000843234 0.00361386 50927.1 44913.16  611.125 898.2632
## 23 32 0.0120 0.0200 0.000912198 0.00390942 51171.3 44463.71  614.056 889.2743
## 24 33 0.0120 0.0200 0.000957894 0.00410526 51399.8 44014.67  616.798 880.2934
## 25 34 0.0120 0.0200 0.001017562 0.00436098 51614.1 43570.48  619.369 871.4096
## 26 35 0.0108 0.0196 0.001063223 0.00455667 51813.6 43128.43  559.587 845.3172
## 27 36 0.0108 0.0196 0.001116885 0.00478665 52044.3 42646.18  562.078 835.8651
## 28 37 0.0108 0.0196 0.001204854 0.00516366 52259.9 42168.26  564.407 826.4979
## 29 38 0.0108 0.0196 0.001342397 0.00575313 52459.0 41688.43  566.558 817.0932
## 30 39 0.0108 0.0196 0.001515549 0.00649521 52639.2 41198.05  568.503 807.4818
## 31 40 0.0100 0.0220 0.001700384 0.00728736 52798.4 40691.48  527.984 895.2126
## 32 41 0.0100 0.0220 0.001919204 0.00822516 53075.8 40027.72  530.758 880.6098
## 33 42 0.0100 0.0220 0.002191042 0.00939018 53323.8 39348.63  533.238 865.6699
## 34 43 0.0100 0.0220 0.002394728 0.01026312 53539.4 38646.71  535.394 850.2276
## 35 44 0.0100 0.0220 0.002652783 0.01136907 53726.0 37935.24  537.260 834.5753
## 36 45 0.0100 0.0198 0.002990225 0.01281525 53880.8 37206.64  538.808 736.6914
## 37 46 0.0100 0.0198 0.003315557 0.01420953 53917.6 36531.94  539.176 723.3325
## 38 47 0.0100 0.0198 0.003696819 0.01584351 53923.0 35828.68  539.230 709.4079
## 39 48 0.0100 0.0198 0.004169879 0.01787091 53893.8 35090.85  538.938 694.7989
## 40 49 0.0100 0.0198 0.004615513 0.01978077 53824.9 34307.89  538.249 679.2962
## 41 50     NA     NA          NA         NA 53717.5 33488.20       NA       NA
##       dx_2_4   dx_1_3   Lx_ns        Tx    ex_ns
## 1    0.00000  35.6210 97100.2 2418193.8 24.90411
## 2    0.61876  33.7442 95323.0 2321093.6 24.34978
## 3    1.33536  36.1953 93588.0 2225770.6 23.78266
## 4    2.35938  42.3824 91890.0 2132182.6 23.20364
## 5    3.85419  51.6216 90224.9 2040292.6 22.61341
## 6    5.83824  62.1929 85907.5 1950067.7 22.69962
## 7   13.24373  71.2956 79211.1 1864160.2 23.53409
## 8   22.16144  77.8655 73068.9 1784949.1 24.42832
## 9   30.06809  76.9339 67439.2 1711880.3 25.38405
```

```
## 10  36.03367  71.3820 62285.1 1644441.1 26.40185
## 11  80.66291  45.0339 58790.8 1582156.0 26.91161
## 12  83.72361  42.9141 56779.9 1523365.2 26.82929
## 13  90.00196  42.5891 54869.5 1466585.2 26.72858
## 14  95.82418  42.0685 53053.8 1411715.7 26.60913
## 15  98.74342  40.3951 51328.6 1358661.9 26.46988
## 16  98.03530  37.5174 50507.5 1307333.3 25.88396
## 17  91.33041  35.1105 50545.9 1256825.8 24.86504
## 18  86.63966  33.4523 50583.8 1206279.9 23.84716
## 19  87.83458  34.0559 50619.8 1155696.1 22.83091
## 20  92.34651  35.9528 50652.2 1105076.3 21.81697
## 21 151.32123  39.4362 50797.2 1054424.2 20.75753
## 22 162.30988  42.9435 51049.2 1003627.0 19.66000
## 23 173.82733  46.6784 51285.6  952577.8 18.57399
## 24 180.69166  49.2356 51507.0  901292.2 17.49845
## 25 190.01000  52.5205 51713.9  849785.3 16.43245
## 26 196.52203  55.0894 51928.9  798071.4 15.36853
## 27 204.13233  58.1275 52152.1  746142.5 14.30705
## 28 217.74255  62.9656 52359.5  693990.4 13.25434
## 29 239.83893  70.4209 52549.1  641630.9 12.21012
## 30 267.59000  79.7772 52718.8  589081.8 11.17405
## 31 296.53348  89.7775 52937.1  536363.1 10.13209
## 32 329.23440 101.8633 53199.8  483426.0  9.08699
## 33 369.49075 116.8347 53431.6  430226.2  8.05191
## 34 396.63583 128.2123 53632.7  376794.6  7.02546
## 35 431.28842 142.5235 53803.4  323161.9  6.00635
## 36 476.81237 161.1157 53899.2  269358.5  4.99745
## 37 519.10173 178.7668 53920.3  215459.3  3.99589
## 38 567.65211 199.3434 53908.4  161539.0  2.99655
## 39 627.10548 224.7306 53859.4  107630.6  1.99836
## 40 678.63642 248.4297 53771.2   53771.2  1.00000
## 41        NA       NA      NA       0.0       NA
```

```
print(IT_nonsmoke$ex_ns[1])
```

```
## [1] 24.9041
```

#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.

```
# Combine the IT_smoke and IT_nonsmoke dataframes

plot_data <- data.frame(
  Status = c("Smoking", "Non-Smoking"),
  Expected_Duration = c(IT_smoke$ex_s[1], IT_nonsmoke$ex_ns[1])
)

# Create a ggplot2 bar plot with light purple and dark purple colors and no legend
ggplot(plot_data, aes(x = Status, y = Expected_Duration, fill = Status)) +
  geom_bar(stat = "identity", width = 0.5, position = position_dodge(width = 0.75)) +
  geom_text(aes(label = Expected_Duration), vjust = -0.5) +
  ylab("Expected Duration") +
  xlab("Status") +
  scale_fill_manual(values = c("Smoking" = "#A6B1E1", "Non-Smoking" = "#635D8D")) +  # Light and dark purple
  ggtitle("Expected Duration by Smoking Status") +
  theme_minimal() +
  theme(legend.position = "none")  # Remove the legend
```