# Homework 01

Ihsan Kahveci

2022-04-14

## Contents

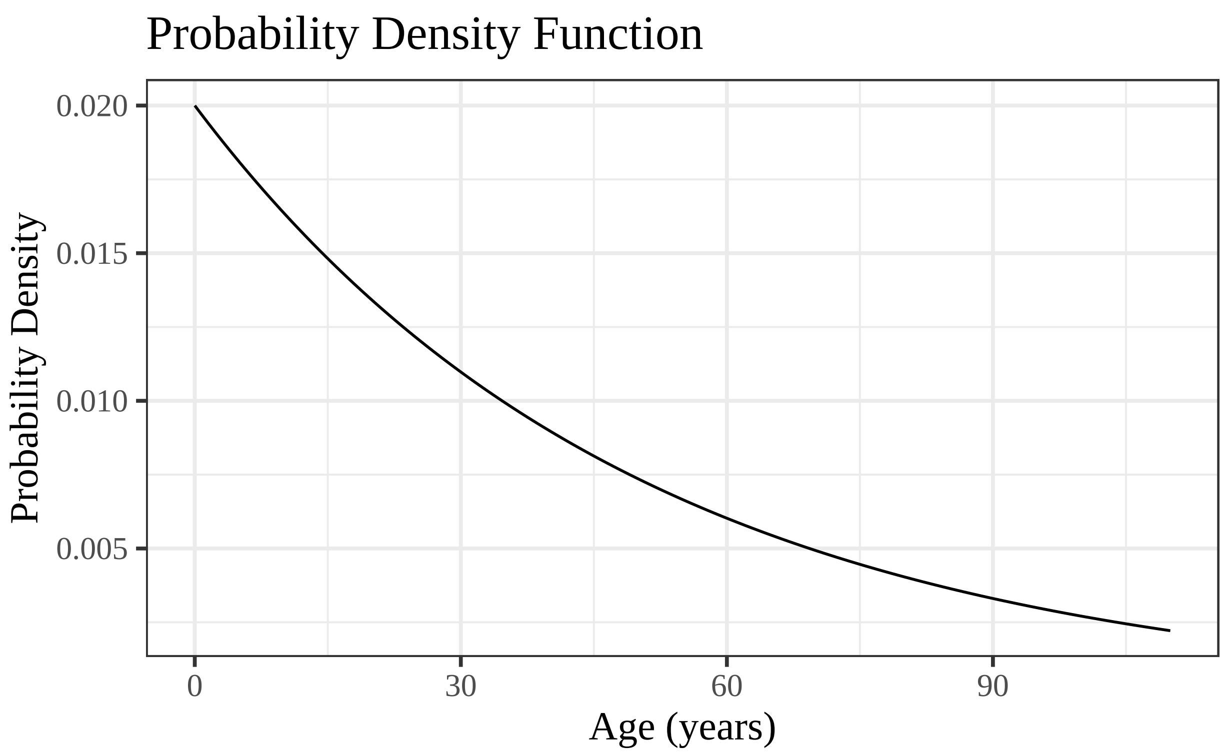
# 1 *Q1*

If the instantaneous mortality rate is constant (0.02) — independent of age, the probability distribution follows an exponential distribution. Let $x$ equals to age at death; hence the relevant probability density function $f(x)$, cumulative distribution function $F(x)$, survival function $S(x)$:

$$
\begin{aligned}
f(x) &= 0.02e^{-0.02x} \\
F(x) &= 1 - e^{0.02x} \\
S(x) &= e^{-0.02x}
\end{aligned}
$$

## 1.1 (a)

Plotting the probability density function:




## 1.2 (b)

Probability that a member of this population is still alive at age 70: $S(70) = 0.246597$

## 1.3 (c)

Probability that a member of this population dies before age 6: $F(6) = 0.1130796$

## 1.4 (d)

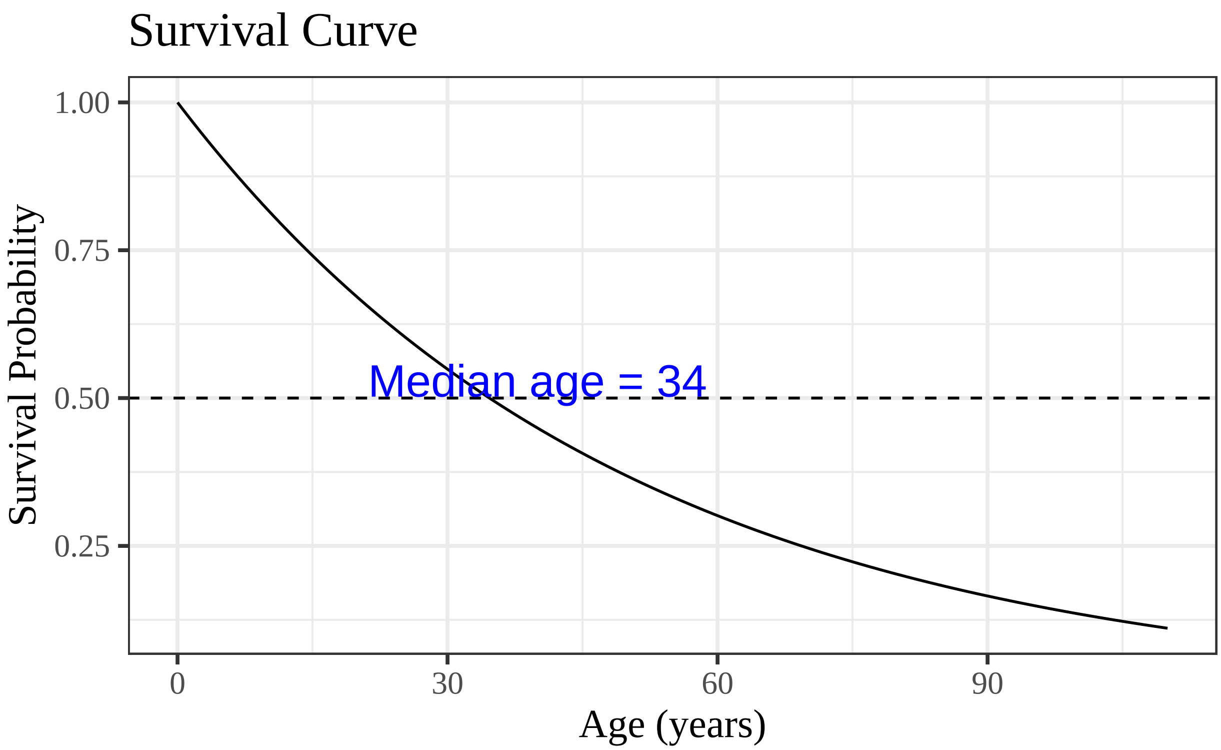
Life expectancy at birth for a member of this population: $e_0 = 50$

### 1.5 (e)

Life expectancy at age 50 for a member of this population: $e_{50} = 50$

### 1.6 (f)

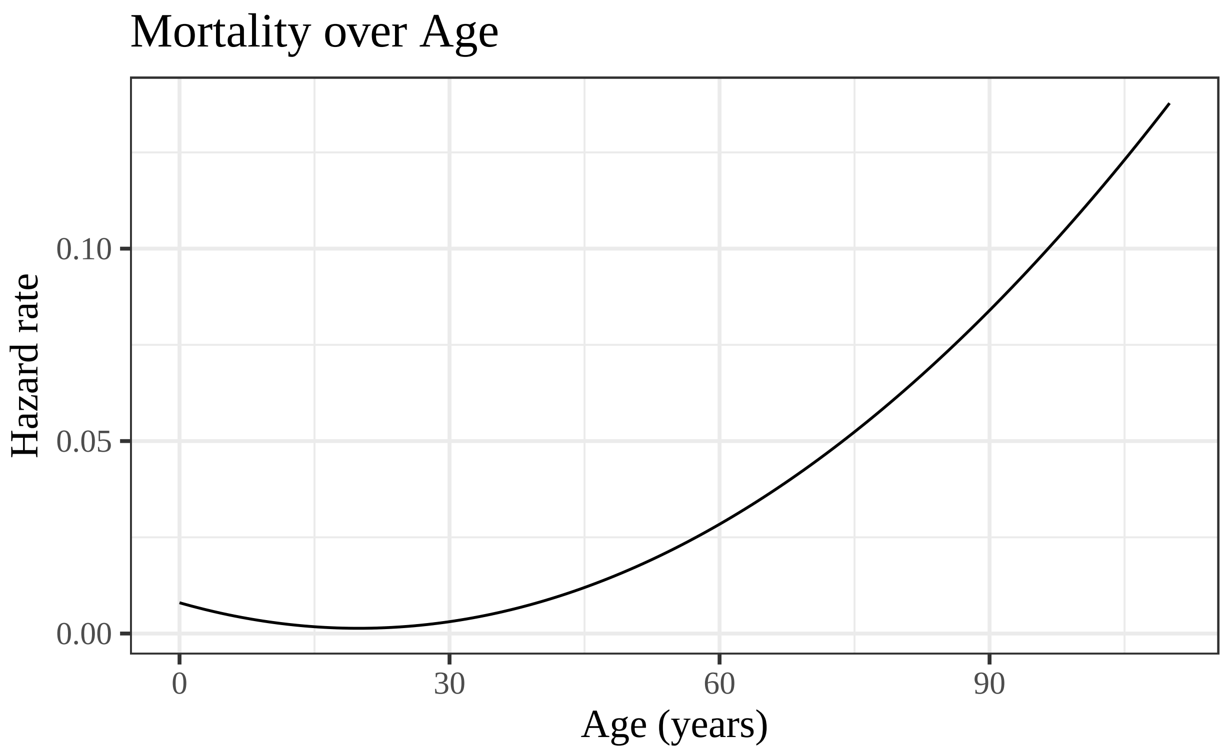
Median age at death for this population: $S(50) = 0.506617$




## 2 *Q2*

### 2.1 (a)

For ages 0 to 110, this mortality rate plot looks like:

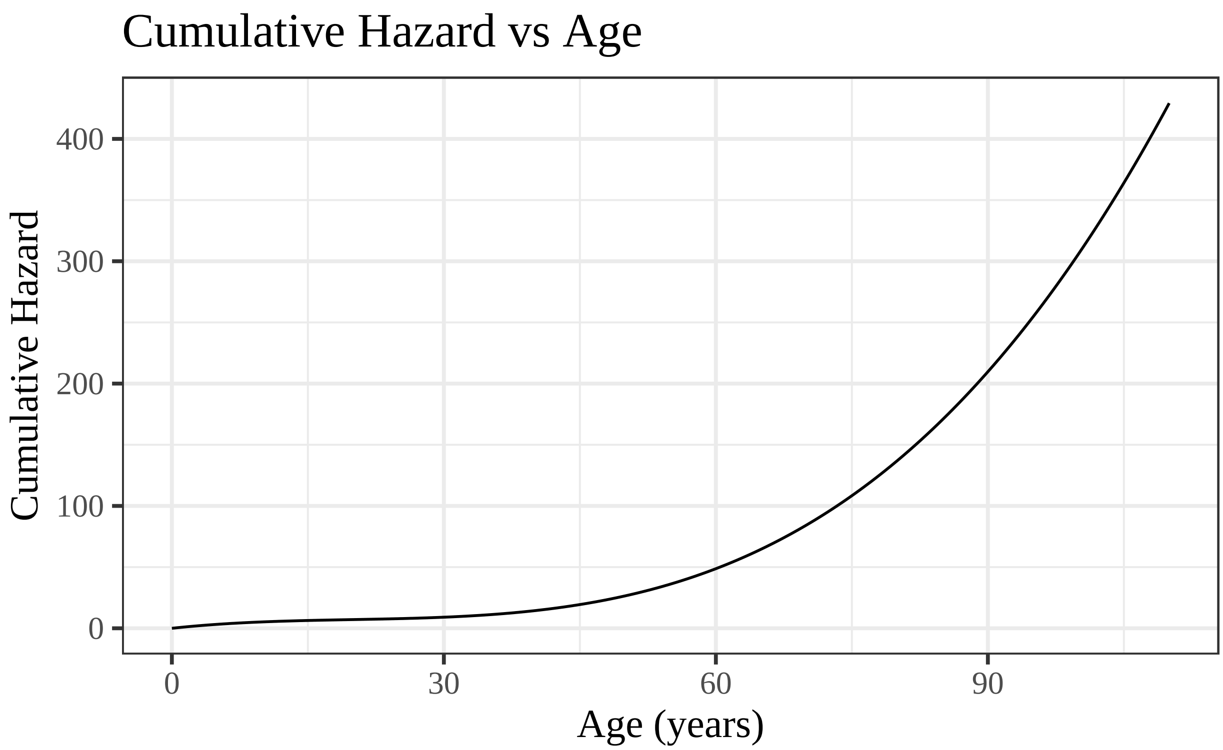



### 2.2 (b)

For a given instantaneous mortality function $\mu(x) = (0.0168x^2 - 0.668x + 8)/1000$, defined as the total area under the curve of $\mu(x)$ bounded on the interval $[0, x]$, or put another way:

$$\begin{aligned}
\Lambda(x) &= \int_0^x \mu(u)du \\
&= \int_0^x \left[(0.0168x^2 - 0.668x + 8)/1000\right] dx \\
&= \int_0^x \frac{21x^2 - 835x + 10000}{1250000} dx \\
&= \frac{21}{125000} \int_0^x x^2 dx - \frac{167}{250000} \int_0^x x dx + \frac{1}{125} \int_0^x 1 dx \\
&= \frac{x \cdot (14x^2 - 835x + 20000)}{25000} + C
\end{aligned}$$

For ages 0 to 110, this cumulative hazard functions looks like:

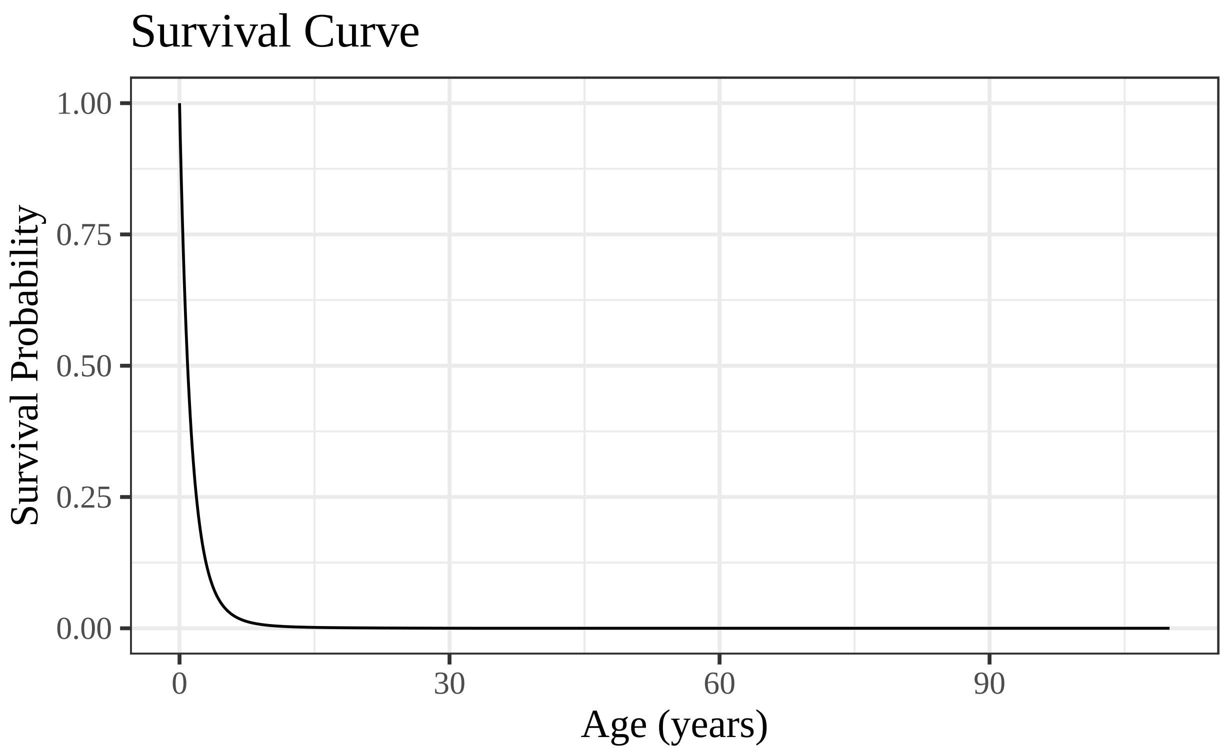



The survival function, $S(x)$, is defined as the exponentiated negative cumulative hazard function, $e^{-\Lambda(x)}$. Using our calculated cumulative hazard function, the survival function is then:

$$S(x) = \exp\left[\frac{-x \cdot (14x^2 - 835x + 20000)}{25000}\right]$$

### 2.3 (c)

For ages 0 to 110, the survival function then looks like:

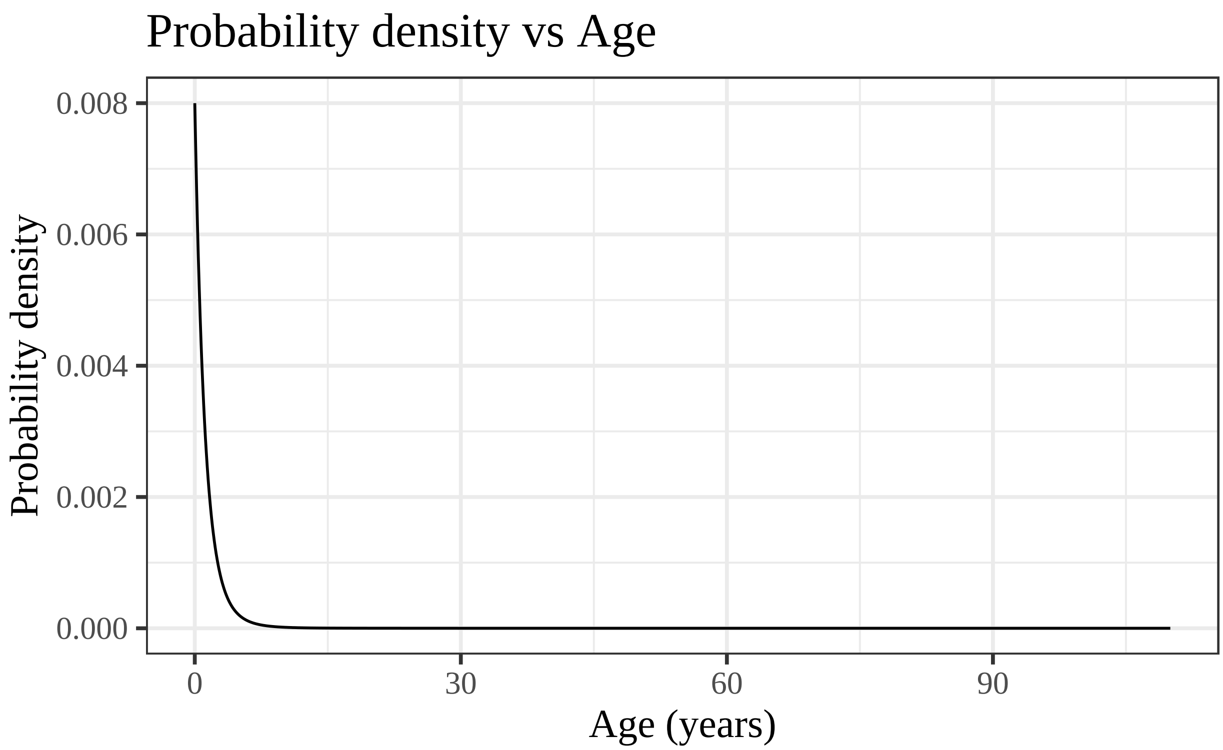



### 2.4 (d)

The probability density function of $X$, $f(x)$, is the negative derivative of the survival function with respect to x, $f(x) = -\frac{dS(x)}{dx}$. Using our calculated survival function, the probability density function of $X$ is then:

$$f(x) = \frac{-d}{dx} = \mu(x)S(x)$$

For ages 0 to 110, the probability density function looks like:




### 2.5 (e)

Life expectancy at age $x$, $e_x$, is defined as:

$$e_x = \frac{\int_x^\infty S(u)du}{S(x)}$$

which simplifies to $\int_0^\infty S(u)du$ for life expectancy at birth, $e_0$. Using numerical integration, the life expectancy at birth for our cohort is calculated to be **1.442**.

### 2.6 (f)

The life expectancy at age 10 ($e_{10}$) for a member of this cohort is numerically calculated to be **4.716**.

### 2.7 (g)

The probability that a person aged $x$ dies within the next $n$ years is defined as:

$${}_nq_x = \frac{S(x) - S(x+n)}{S(x)}$$

The ${}_{45}q_{15}$ value for this cohort is then **1**.

# 3 *Q3*

## 3.1 (a) & (b)

$$_n d_x = l_x - l_{x+n}$$

Table 1: UN Prospects for Peru

| age | 2015 | 2020 | ndx |
|---|---|---|---|
| 70-74 | 275.30 | 352.75 | 31.54 |
| 75-79 | 202.70 | 243.75 | 41.53 |
| 80-84 | 125.20 | 161.17 | 54.06 |
| 85-89 | 52.70 | 71.14 | 30.86 |
| 90-94 | 12.83 | 21.84 | 9.19 |
| 95-99 | 2.18 | 3.65 | 1.79 |

# 4 *Q4*

## 4.1 (a)

$$_n M_x = \frac{_n d_x}{_n L_x}$$

Table 2: UN Life Table for Peru

| age | 2015 | 2020 | ndx | nLx | nMx |
|---|---|---|---|---|---|
| 70-74 | 275.30 | 352.75 | 31.54 | 1570.11 | 0.02 |
| 75-79 | 202.70 | 243.75 | 41.53 | 1116.12 | 0.04 |
| 80-84 | 125.20 | 161.17 | 54.06 | 715.92 | 0.08 |
| 85-89 | 52.70 | 71.14 | 30.86 | 309.60 | 0.10 |
| 90-94 | 12.83 | 21.84 | 9.19 | 86.68 | 0.11 |
| 95-99 | 2.18 | 3.65 | 1.79 | 14.57 | 0.12 |

## 4.2 (b)

$$\text{Solution 1:}$$
$$_n q_x \approx 1 - e^{-_n M_x}$$

$$\text{Solution 2:}$$
$$_n q_x \approx \frac{n_n M_x}{1 + n_n M_x / 2}$$

Table 3: First Method

| age | 2015 | 2020 | ${}_nd_x$ | ${}_nL_x$ | ${}_nM_x$ | ${}_nq_x$ |
|---|---|---|---|---|---|---|
| 70-74 | 275.30 | 352.75 | 31.54 | 1570.11 | 0.02 | 0.10 |
| 75-79 | 202.70 | 243.75 | 41.53 | 1116.12 | 0.04 | 0.17 |
| 80-84 | 125.20 | 161.17 | 54.06 | 715.92 | 0.08 | 0.31 |
| 85-89 | 52.70 | 71.14 | 30.86 | 309.60 | 0.10 | 0.39 |
| 90-94 | 12.83 | 21.84 | 9.19 | 86.68 | 0.11 | 0.41 |
| 95-99 | 2.18 | 3.65 | 1.79 | 14.57 | 0.12 | 0.46 |

Table 4: Second Method

| age | 2015 | 2020 | ${}_nd_x$ | ${}_nL_x$ | ${}_nM_x$ | ${}_nq_x$ |
|---|---|---|---|---|---|---|
| 70-74 | 275.30 | 352.75 | 31.54 | 1570.11 | 0.02 | 0.10 |
| 75-79 | 202.70 | 243.75 | 41.53 | 1116.12 | 0.04 | 0.17 |
| 80-84 | 125.20 | 161.17 | 54.06 | 715.92 | 0.08 | 0.32 |
| 85-89 | 52.70 | 71.14 | 30.86 | 309.60 | 0.10 | 0.40 |
| 90-94 | 12.83 | 21.84 | 9.19 | 86.68 | 0.11 | 0.42 |
| 95-99 | 2.18 | 3.65 | 1.79 | 14.57 | 0.12 | 0.47 |

### 4.3 (c)

$$ {}_nm_x = \frac{{}_nd_x}{{}_nL_x} = \frac{{}_nd_x}{nl_x - 1/2 * n_n d_x} $$

Table 5: Life Table ${}_nm_x$

| age | 2015 | 2020 | ${}_nd_x$ | ${}_nL'_x$ | ${}_nm_x$ |
|---|---|---|---|---|---|
| 70-74 | 275.30 | 352.75 | 31.54 | 1297.62 | 0.02 |
| 75-79 | 202.70 | 243.75 | 41.53 | 909.68 | 0.05 |
| 80-84 | 125.20 | 161.17 | 54.06 | 490.83 | 0.11 |
| 85-89 | 52.70 | 71.14 | 30.86 | 186.36 | 0.17 |
| 90-94 | 12.83 | 21.84 | 9.19 | 41.19 | 0.22 |
| 95-99 | 2.18 | 3.65 | 1.79 | 6.44 | 0.28 |

### 4.4 (d)

Table 6: Comparison

| age | ${}_nq_x^1$ | ${}_nq_x^2$ |
|---|---|---|
| 70-74 | 0.096 | 0.096 |
| 75-79 | 0.170 | 0.170 |
| 80-84 | 0.314 | 0.318 |
| 85-89 | 0.393 | 0.399 |
| 90-94 | 0.411 | 0.419 |
| 95-99 | 0.458 | 0.469 |

As seen from the table above, the estimates from two methods are quite similiar. The difference occurs in later ages because ${}_nM_x$ becomes larger. I think second method is better because it doesn't assume that mortality rate is constant within the age interval. Instead, it is modeled as a linear function of age. As we move from 5-year age groups to single age groups, this became more important because even two-year age difference matters in later ages.

## 4.5 (e)

Table 7: Comparison: ${}_nM_x$ and ${}_nm_x$

| age | 2015 | ${}_nd_x$ | ${}_nL_x$ | ${}_nL'_x$ | ${}_nM_x$ | ${}_nm_x$ |
|---|---|---|---|---|---|---|
| 70-74 | 275.296 | 31.544 | 1570.110 | 1297.620 | 0.020 | 0.024 |
| 75-79 | 202.698 | 41.525 | 1116.125 | 909.678 | 0.037 | 0.046 |
| 80-84 | 125.197 | 54.061 | 715.925 | 490.832 | 0.076 | 0.110 |
| 85-89 | 52.703 | 30.861 | 309.597 | 186.362 | 0.100 | 0.166 |
| 90-94 | 12.832 | 9.186 | 86.685 | 41.195 | 0.106 | 0.223 |
| 95-99 | 2.181 | 1.786 | 14.567 | 6.440 | 0.123 | 0.277 |

${}_nM_x$ uses person-years (${}_nL_x$) assuming mortality rate is constant within 5-year age groups. "${}_nm_x$" uses person-years (${}_nL'_x$) assuming mortality rate is a linear function of age. As a result, former estimates more person-years because it doesn't take into account the deaths occur within the early years of the age group which happens more in older ages. As expected, ${}_nM_x$ underestimates mortality rate in older ages comparing to ${}_nm_x$

# 5 *Q5*

## 5.1 (a)

Table 8: Life Table - First Method

| age | $_nq_x$ | $_np_x$ | $_nd_x$ | $l_x$ | $L_x$ | $T_x$ | $e_x$ |
|---|---|---|---|---|---|---|---|
| 70-74 | 0.10 | 0.90 | 10000.00 | 955.71 | 47610.72 | 163332.68 | 16.33 |
| 75-79 | 0.17 | 0.83 | 9044.29 | 1535.23 | 41383.38 | 115721.96 | 12.80 |
| 80-84 | 0.31 | 0.69 | 7509.06 | 2361.36 | 31641.90 | 74338.59 | 9.90 |
| 85-89 | 0.39 | 0.61 | 5147.70 | 2020.48 | 20687.30 | 42696.68 | 8.29 |
| 90-94 | 0.41 | 0.59 | 3127.22 | 1286.25 | 12420.49 | 22009.38 | 7.04 |
| 95-99 | 0.46 | 0.54 | 1840.97 | 843.68 | 7095.66 | 9588.89 | 5.21 |
| 100+ | 1.00 | 0.00 | 997.29 | 997.29 | 2493.23 | 2493.23 | 2.50 |

Table 9: Life Table - Second Method

| age | $_nq_x$ | $_np_x$ | $_nd_x$ | $l_x$ | $L_x$ | $T_x$ | $e_x$ |
|---|---|---|---|---|---|---|---|
| 70-74 | 0.10 | 0.90 | 10000.00 | 956.48 | 47608.81 | 162423.19 | 16.24 |
| 75-79 | 0.17 | 0.83 | 9043.52 | 1539.15 | 41369.76 | 114814.38 | 12.70 |
| 80-84 | 0.32 | 0.68 | 7504.38 | 2383.42 | 31563.35 | 73444.63 | 9.79 |
| 85-89 | 0.40 | 0.60 | 5120.96 | 2043.16 | 20496.93 | 41881.27 | 8.18 |
| 90-94 | 0.42 | 0.58 | 3077.81 | 1289.23 | 12165.97 | 21384.34 | 6.95 |
| 95-99 | 0.47 | 0.53 | 1788.58 | 839.20 | 6844.91 | 9218.37 | 5.15 |
| 100+ | 1.00 | 0.00 | 949.38 | 949.38 | 2373.46 | 2373.46 | 2.50 |

## 5.2 (b)

Table 10: Life Expectancy Comparison

| age | $e_x^1$ | $e_x^2$ |
|---|---|---|
| 70-74 | 16.33 | 16.24 |
| 75-79 | 12.80 | 12.70 |
| 80-84 | 9.90 | 9.79 |
| 85-89 | 8.29 | 8.18 |
| 90-94 | 7.04 | 6.95 |
| 95-99 | 5.21 | 5.15 |
| 100+ | 2.50 | 2.50 |

The first approximation method generated slightly higher life expectancies for all ages, except 100+. This happens because the first method estimates lower probability of death ($_nq_x$).

## 6 Code Appendix

```
# Prep work ---------------------------------------------------------------

# Load libraries
library(tidyverse, quietly = TRUE)
library(wpp2019)

# Make data
age_range <- c(0, 110)
age_data <- tibble(age = seq(age_range[1], age_range[2], .1))

# Question 1 ---------------------------------------------------------------
pdf = function(x) 0.02*exp(-0.02*x)
cdf = function(x) 1 - exp(-0.02*x)
survf = function(x) exp(-0.02*x)

pdf_plot =
  age_data %>%
  mutate(PDF = pdf(age)) %>%
  ggplot(aes(x=age, y=PDF)) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Probability Density Function",
    x = "Age (years)",
    y = "Probability Density"
    )

median_plot =
  age_data %>%
  mutate(surv = survf(age)) %>%
  ggplot(aes(x=age, y=surv)) +
  geom_line() +
  geom_hline(yintercept = 0.5, linetype="dashed") +
  annotate("text", x=40, y=0.53, label="Median age = 34", color = "blue", size=6) +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Survival Curve",
    x = "Age (years)",
    y = "Survival Probability"
    )
pdf_plot
median_plot
# Question 2 ---------------------------------------------------------------
hazard_fun = function(x) (0.0168*x^2 - 0.668*x + 8)/1000

hazard_plot =
  age_data %>%
  mutate(mortality = hazard_fun(age)) %>%
  ggplot(aes(x=age, y=mortality)) +
```

```
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Mortality over Age",
    x = "Age (years)",
    y = "Hazard rate"
    )

cum_hazard_fun <- function(x) x*(14*x^2 - 835*x + 20000)/25000

chf_plot <-
  ggplot(age_data, aes(x = age, y = cum_hazard_fun(age))) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Cumulative Hazard vs Age",
    x = "Age (years)",
    y = "Cumulative Hazard"
  )

survival_fun <- function(x) exp(-1 * cum_hazard_fun(x))

surv_plot =
  age_data %>%
  mutate(surv = survival_fun(age)) %>%
  ggplot(aes(x=age, y=surv)) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Survival Curve",
    x = "Age (years)",
    y = "Survival Probability"
    )

pdf_fun = function(x) survival_fun(x) * hazard_fun(x)

pdfun_plot <-
  ggplot(age_data, aes(x = age, y = pdf_fun(age))) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Probability density vs Age",
    x = "Age (years)",
    y = "Probability density"
  )


e0 <- integrate(survival_fun, lower = 0, upper = Inf)
e0_val <- round(e0$value, 3)
```

```r
e10 <- integrate(survival_fun, lower = 10, upper = Inf)
e10_val <- round(e10$value / survival_fun(10), digits = 3)

nqx <- function(x, n) (survival_fun(x) - survival_fun(x + n)) / survival_fun(x)
q45_15 <- round(nqx(15, 45), 3)
hazard_plot
chf_plot
surv_plot
pdfun_plot

# Question 3 ---------------------------------------------------------------
## download estimates of males
data("popM")

peru =
  popM %>%
  filter(name == "Peru") %>%
  filter(age == c("70-74", "75-79", "80-84", "85-89", "90-94", "95-99", "100+")) %>%
  select(age, "2015", "2020") %>%
  mutate(ndx = `2015` - lead(`2020`)) %>%
  filter(age != "100+")




peru %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "UN Prospects for Peru")

# Question 4 ---------------------------------------------------------------
peru_LT =
  peru %>%
  mutate(nLx = (5*(`2015`+`2020`)/2)) %>%
  mutate(nMx = ndx/nLx)

peru_LT %>%
 knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
              caption = "UN Life Table for Peru")

## first method
peru_LT %>%
 mutate(nqx = 1-exp(-5*nMx)) %>%
 knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
              caption = "First Method",
             col.names=c("age", "2015", "2020", "$_nd_x$", "$_nL_x$", "$_nM_x$", "$_nq_x$"))


## second method
peru_LT %>%
  mutate(nqx = (5*nMx)/(1+5*nMx/2)) %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "Second Method",
               col.names=c("age", "2015", "2020", "$_nd_x$", "$_nL_x$", "$_nM_x$", "$_nq_x$"))
```

```
peru_LT %>%
  mutate(nLx = 5*`2015`-(1/2)*5*ndx) %>%
  mutate(nmx = ndx/nLx) %>%
  select(-nMx) %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "Life Table $_nm_x$",
               col.names = c("age", "2015", "2020", "$_nd_x$", "$_nL_x'$", "$_nm_x$"))

## comparison
peru_LT %>%
  mutate(nqx_1 = 1-exp(-5*nMx)) %>%
  mutate(nqx_2 = (5*nMx)/(1+5*nMx/2)) %>%
  select(age, nqx_1, nqx_2) %>%
  knitr::kable(digits = 3, booktabs = TRUE, escape = FALSE,
               caption = "Comparison", col.names = c("age", "$_nq_x^1$", "$_nq_x^2$"))
peru_LT %>%
  mutate(nLx2 = 5*`2015`-(1/2)*5*ndx) %>%
  mutate(nmx = ndx/nLx2) %>%
  select(age, `2015`, ndx, nLx, nLx2, nMx, nmx) %>%
  knitr::kable(digits = 3, booktabs = TRUE, escape = FALSE,
               caption = "Comparison: $_nM_x$ and $_nm_x$",
               col.names = c("age", "2015", "$_nd_x$", "$_nL_x$", "$_nL_x'$" , "$_nM_x$", "$_nm_x$"))

# Question 5 ---------------------------------------------------------------
## Life Table using first method
LT_1 = peru_LT %>%
  mutate(nqx = 1-exp(-5*nMx)) %>%
  select(age, nqx) %>%
  mutate(n = 5, npx = 1-nqx) %>%
  bind_rows(a=list(age="100+", n=5, nqx=1, npx=0))

npx <- LT_1$npx

age <- seq(70,100,5)
lx <- 10000

for (a in age[-length(age)]) {
  l <- lx[which(age == a)] * npx[which(age == a)]
  lx <- c(lx, l)
}

LT_1 <- LT_1 %>%
  mutate(
    x = age,
    lx = lx,
    ndx = lx*nqx,
    Lx = n*lx - n*ndx/2,
    Tx = rev(Lx) %>% coalesce(0) %>% cumsum() %>% rev(),
    ex = Tx / lx
)

## Life Table using second method
```

```r
LT_2 = peru_LT %>%
  mutate(nqx = (5*nMx)/(1+5*nMx/2)) %>%
  select(age, nqx) %>%
  mutate(n = 5, npx = 1-nqx) %>%
  bind_rows(a=list(age="100+", n=5, nqx=1, npx=0))

npx <- LT_2$npx

age <- seq(70,100,5)
lx <- 10000

for (a in age[-length(age)]) {
  l <- lx[which(age == a)] * npx[which(age == a)]
  lx <- c(lx, l)
}

LT_2 <- LT_2 %>%
  mutate(
    x = age,
    lx = lx,
    ndx = lx*nqx,
    Lx = n*lx - n*ndx/2,
    Tx = rev(Lx) %>% coalesce(0) %>% cumsum() %>% rev(),
    ex = Tx / lx
)
LT_1 %>%
  select(-n, -x) %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "Life Table - First Method",
               col.names=c("age", "$_nq_x$", "$_np_x$", "$_nd_x$", "$l_x$", "$L_x$", "$T_x$", "$e_x$"))


LT_2 %>%
  select(-n, -x) %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "Life Table - Second Method",
               col.names=c("age", "$_nq_x$", "$_np_x$", "$_nd_x$", "$l_x$", "$L_x$", "$T_x$", "$e_x$"))

LT_1 %>%
  select(age, ex_1 = ex) %>%
  cbind(ex_2 = LT_2$ex) %>%
  knitr::kable(digits = 2, booktabs = TRUE, escape = FALSE,
               caption = "Life Expectancy Comparison", col.names = c("age", "$e_x^1$", "$e_x^2$" ))
```