

Homework 02

Ihsan Kahveci

2020-04-18

Questions

Q1

Table 1: Q1 one-sex closed population

Age	Population (N_x)	Fertility Rate (\tilde{F}_x)	Survival Prob. (s_x)
1	18000	0.0	0.65
2	17000	0.9	0.75
3+	14000	0.2	0.15

Q1.a

The crude birth rate (CBR) is defined as the number of births over the person-years lived in the period $[T_1, T_2]$. Since our period is a single year, we can calculate CBR as:

$$CBR = \sum \frac{N_x \tilde{F}_x}{N_x s_x}$$

where we sum over all age groups. The crude birth rate for this population in the next time period is then **0.682**.

Q1.b

The total fertility rate in the population in the period $[T_1, T_2]$ is defined as the sum of the age-specific fertility rates across all age groups, multiplied by the length of the age interval, n . With $T_2 - T_1 = n = 1$, the total fertility rate represents the single-year cohort total fertility rate:

$$TFR[T_1, T_1 + 1] = \sum {}_1F_x[T_1, T_1 + 1]$$

We can convert between \tilde{F}_x and ${}_1F_x$ using the equation

$$\tilde{F}_x = {}_1F_x \times \frac{1}{1 + SRB} \times \frac{1}{2} \left(1 + s_{x-1} \frac{N_{x-1,t}}{N_{x,t}} \right) \times \left(1 - \frac{q_0}{2} \right)$$

where we assume $SRB = 1.05$ and take $q_0 = 1 - s_0$. After converting to age-specific fertility rates, we calculate a total fertility rate of $0 + 2.649 + 0.52 = \mathbf{3.17}$ for this population.

Q1.c

The Leslie matrix, L , for this population is defined as:

$$L = \begin{bmatrix} \tilde{F}_{A-3} & \tilde{F}_{A-2} & \tilde{F}_{A-1} \\ s_{A-3} & 0 & 0 \\ 0 & s_{A-2} & s_{A-1} \end{bmatrix} = \begin{bmatrix} 0 & 0.9 & 0.2 \\ 0.65 & 0 & 0 \\ 0 & 0.75 & 0.15 \end{bmatrix}$$

Where $(A-1)+$ is the highest age group that can be reached in this population, $3+$, s_x denotes the probability of survival to the next age group for age group x , and \tilde{F}_x is the expected number of female births to a woman age x , who survives to the next time interval.

Q1.d

We can project this population forward using the *cohort-component method of population projection*, which states that the age-specific populations one time period ahead (N_{t+1}) can be calculated from the matrix multiplication of the age-specific population in the current period (N_t) and the Leslie matrix (L) of the population. The population by age one period forward from our given initial population is then:

$$\begin{aligned} N_{t+1} &= LN_t \\ &= \begin{bmatrix} 0 & 0.9 & 0.2 \\ 0.65 & 0 & 0 \\ 0 & 0.75 & 0.15 \end{bmatrix} \begin{bmatrix} 18000 \\ 17000 \\ 14000 \end{bmatrix} \\ &= \begin{bmatrix} 18100 \\ 11700 \\ 14850 \end{bmatrix} \end{aligned}$$

Q1.e

This method can be extended to projecting age-specific population k periods ahead by raising the Leslie matrix to the k^{th} power (L^k). Our given population, projected 2 periods into the future is then:

$$\begin{aligned} N_{t+2} &= L^2 N_t \\ &= \begin{bmatrix} 0 & 0.9 & 0.2 \\ 0.65 & 0 & 0 \\ 0 & 0.75 & 0.15 \end{bmatrix}^2 \begin{bmatrix} 18000 \\ 17000 \\ 14000 \end{bmatrix} \\ &= \begin{bmatrix} 0.585 & 0.15 & 0.03 \\ 0 & 0.585 & 0.13 \\ 0.488 & 0.112 & 0.022 \end{bmatrix} \begin{bmatrix} 18000 \\ 17000 \\ 14000 \end{bmatrix} \\ &= \begin{bmatrix} 13500 \\ 11765 \\ 11002.5 \end{bmatrix} \end{aligned}$$

Q1.f

The crude birth rate for this population from time period 1 to time period 2 is **0.593**.

The total fertility rate between time periods 1&2 is $0 + 2.23 + 0.625 = \mathbf{2.855}$ for this population.

Q1.g

From the theorem that N_t converges to $\lambda^t u$ as t approaches infinity, $\log(\lambda)$ is the *instantaneous rate of increase of the population*. Here, λ is defined as the dominant **right eigenvalue** of the Leslie matrix, or for the equation:

$$Lv = \lambda v$$

it is the eigenvalue λ with the largest magnitude. For our calculated Leslie matrix, the instantaneous rate of increase is **-0.161**.

Q1.h

Again, from the formula $\lambda^t u$, u is the *stable age distribution*, and is defined as the dominant **right eigenvector** of the Leslie matrix, which is the column vector v from the eigendecomposition of L corresponding to the eigenvalue λ with the largest magnitude.

For our calculated Leslie matrix, the stable age distribution is $\begin{bmatrix} 0.666 \\ 0.509 \\ 0.545 \end{bmatrix}$.

Q1.i

The reproductive value vector (v) is a vector of expected the number of future offspring of an individual for each age group. A theorem states that v is the dominant **left eigenvector** of the Leslie matrix for the population. The left dominant eigenvector of a matrix A is equivalent to the right dominant eigenvector of the transpose of matrix, A^\top . So, in the formula:

$$L^\top u = \kappa u$$

the dominant eigenvector u represents the reproductive values. For our Leslie matrix, the reproductive value matrix is then $\begin{bmatrix} -0.598 \\ -0.783 \\ -0.171 \end{bmatrix}$.

Q2

Q2.a

Table 2: Age-specific Mortality Rates of Peru female population, 2015-2020

Age	${}_nq_x$
0	0.012
1	0.001
5	0.000
10	0.000
15	0.000
20	0.001
25	0.001
30	0.001
35	0.002
40	0.002
45	0.003
50	0.004
55	0.005
60	0.008
65	0.013
70	0.022
75	0.036
80	0.062
85	0.108
90	0.172
95	0.251
100	0.367

Q2.b

First, we need to calculate a life table from mortality rates.

Table 3: Life table for 2015-2020 Peru female population

Age	a_x	${}_5m_x$	${}_1q_x$	${}_1s_x$	${}_5d_x$	l_x	${}_5L_x$	${}_5T_x$	e_x
0	0.1	0.012	0.012	0.989	1150	100000	98948	7907710	79
1-4	1.5	0.001	0.003	0.997	315	98850	394614	7808761	79
5-9	2.5	0.000	0.002	0.998	197	98535	492182	7414148	75
10-14	2.5	0.000	0.002	0.999	148	98338	491320	6921965	70
15-19	2.5	0.000	0.002	0.998	220	98190	490400	6430645	65
20-24	2.5	0.001	0.003	0.997	323	97970	489042	5940245	61
25-29	2.5	0.001	0.004	0.996	443	97647	487128	5451202	56
30-34	2.5	0.001	0.006	0.994	571	97204	484592	4964075	51
35-39	2.5	0.002	0.008	0.992	732	96633	481335	4479482	46
40-44	2.5	0.002	0.010	0.991	912	95901	477225	3998148	42
45-49	2.5	0.003	0.013	0.987	1245	94989	471832	3520922	37
50-54	2.5	0.004	0.018	0.982	1650	93744	464595	3049090	33
55-59	2.5	0.005	0.025	0.975	2350	92094	454595	2584495	28

Age	a_x	${}_5m_x$	${}_1q_x$	${}_1s_x$	${}_5d_x$	l_x	${}_5L_x$	${}_5T_x$	e_x
60-64	2.5	0.008	0.040	0.960	3614	89744	439685	2129900	24
65-69	2.5	0.013	0.061	0.939	5242	86130	417545	1690215	20
70-74	2.5	0.022	0.102	0.897	8291	80888	383712	1272670	16
75-79	2.5	0.036	0.166	0.835	12013	72597	332952	888958	12
80-84	2.5	0.062	0.270	0.730	16366	60584	262005	556005	9
85-89	2.5	0.108	0.426	0.574	18822	44218	174035	294000	7
90-94	2.5	0.172	0.601	0.399	15259	25396	88832	119965	5
95-99	2.5	0.251	0.772	0.228	7821	10137	31132	31132	3
100+	-	0.367	1.000	0.000	2316	2316	-	-	-

Table 4: Peru ${}_ns_x$: 0 and 1-4 combined

Age	${}_nq_x$	${}_ns_x$
0-4	0.01	0.99
5-9	0.00	1.00
10-14	0.00	1.00
15-19	0.00	1.00
20-24	0.00	1.00
25-29	0.00	1.00
30-34	0.01	0.99
35-39	0.01	0.99
40-44	0.01	0.99
45-49	0.01	0.99
50-54	0.02	0.98
55-59	0.03	0.97
60-64	0.04	0.96
65-69	0.06	0.94
70-74	0.10	0.90
75-79	0.17	0.83
80-84	0.27	0.73
85-89	0.43	0.57
90-94	0.60	0.40
95-99	0.77	0.23
100+	1.00	0.00

Then, we will use, we will use ${}_ns_x$ and q_0 values to convert between \tilde{F}_x and ${}_1F_x$ using the equation.

To calculate ${}_5\tilde{F}_x$, we first use the provided proportional age-specific fertility rate and total fertility rate for Peru in 2015 to get age-specific fertility rate with the formula:

$$\frac{TFR[T_1, T_2] \times {}_nPASFR_x}{n} = {}_nF_x[T_1, T_2]$$

where $n = 5$ and $[T_1, T_2] = [2015, 2020]$. Then we use the provided population and previously calculated mortality rates ${}_5s_x$ and ${}_5q_0$ to calculate ${}_5\tilde{F}_x$ using the formula:

$${}_n\tilde{F}_x = {}_nF_x \times \frac{1}{1 + SRB} \times \frac{1}{2} \left(1 + {}_ns_x \frac{{}_5N_{x-1,t}}{{}_5N_{x,t}} \right) \times \left(1 - \frac{{}_nq_0}{2} \right)$$

Note that ${}_5q_0$ was calculated as $1 - {}_5s_0$, which was in turn calculated from ${}_1s_0 \times {}_4s_1$.

With these calculations, our resulting ${}_5\tilde{F}_x$ is:

Table 5: Expected number of live female births per woman per five-year period in Peru, 2015-2020

Age	${}_5\tilde{F}_x$	${}_5F_x$	${}_5N_x$	${}_5s_x$
0-4	0.000	0.000	1370	0.985
5-9	0.000	0.000	1454	0.998
10-14	0.000	0.000	1361	0.999
15-19	0.000	5.688	1338	0.998
20-24	4.876	9.997	1316	0.997
25-29	5.481	10.921	1223	0.996
30-34	4.787	9.568	1141	0.994
35-39	3.115	6.266	1077	0.992
40-44	1.265	2.536	1008	0.991
45-49	0.216	0.424	899	0.987
50-54	0.000	0.000	777	0.982
55-59	0.000	0.000	639	0.975
60-64	0.000	0.000	531	0.960
65-69	0.000	0.000	395	0.939
70-74	0.000	0.000	313	0.897
75-79	0.000	0.000	233	0.835
80-84	0.000	0.000	157	0.730
85-89	0.000	0.000	73	0.574
90-94	0.000	0.000	25	0.399
95-99	0.000	0.000	5	0.228
100+	0.000	0.000	1	0.000

Q2.c

Using our calculated ${}_5\tilde{F}_x$ and ${}_5s_x$, we can build a Leslie matrix for this population:

Table 6: Leslie matrix for Peru females, 2015-2020, (sparse format)

Row	Column	Value
1	5	4.876
1	6	5.481
1	7	4.787
1	8	3.115
1	9	1.265
1	10	0.216
2	1	0.985
3	2	0.998
4	3	0.999
5	4	0.998
6	5	0.997
7	6	0.996
8	7	0.994
9	8	0.992
10	9	0.991
11	10	0.987

Row	Column	Value
12	11	0.982
13	12	0.975
14	13	0.960
15	14	0.939
16	15	0.897
17	16	0.835
18	17	0.730
19	18	0.574
20	19	0.399
21	20	0.228

Q2.d

Using the given female population in Peru, 2015-2020 (*table 3*), we can calculate the population one 5-year period ahead to 2020-2025, assuming fertility and mortality are constant over time, and that there is no migration.

$$N_{t+1} = LN_t = \begin{bmatrix} 23406 \\ 1350 \\ 1451 \\ 1359 \\ 1335 \\ 1312 \\ 1217 \\ 1135 \\ 1069 \\ 999 \\ 887 \\ 763 \\ 623 \\ 510 \\ 371 \\ 281 \\ 194 \\ 115 \\ 42 \\ 10 \\ 1 \end{bmatrix}$$

Q2.e

Under the same assumptions as the previous question, we can instead project the population 10 years (2 time periods) into the future.

$$N_{t_2} = L^2 N_t = \begin{bmatrix} 24628 \\ 23063 \\ 1347 \\ 1449 \\ 1356 \\ 1331 \\ 1306 \\ 1210 \\ 1126 \\ 1059 \\ 986 \\ 872 \\ 744 \\ 598 \\ 479 \\ 333 \\ 235 \\ 142 \\ 66 \\ 17 \\ 2 \end{bmatrix}$$

Q3

Q3.a

Under the assumption that the age-specific migration rates follow the same distribution as the population distribution, we can calculate age-specific migration rates using the provided net migration rate (after converting the rate from *migrations per 1,000 person-years* to *migrant per person-year*).

Table 7: Age-specific migration rate in Peru, 2015-2020

Age	${}_5G_x$
0-4	4.42e-02
5-9	4.70e-02
10-14	4.40e-02
15-19	4.32e-02
20-24	4.25e-02
25-29	3.95e-02
30-34	3.69e-02
35-39	3.48e-02
40-44	3.26e-02
45-49	2.90e-02
50-54	2.51e-02
55-59	2.07e-02
60-64	1.71e-02
65-69	1.28e-02
70-74	1.01e-02
75-79	7.52e-03
80-84	5.07e-03
85-89	2.37e-03
90-94	7.97e-04

Age	${}_5G_x$
95-99	1.58e-04
100+	1.79e-05

Q3.b

To incorporate migration into the population projection to 2030, I used the assumption that all migration happens half-way through a time interval:

$$\begin{aligned}
N_{t+1} &= LN_t + L^{\frac{1}{2}}G_t \\
&\approx LN_t + \frac{1}{2}(I + L)G^t
\end{aligned}$$

and

$$\begin{aligned}
N_{t+2} &= L^2N_t + L^{\frac{2}{2}}G_t \\
&\approx L^2N_t + LG^t
\end{aligned}$$

where N and G refer to female population and migrants, and I is the identity matrix. Using this equation, we calculate the with-migration female population in 2020 and 2030:

Table 8: Peru female with-migration population projections

Age	2025	2030
0-4	24129	25629
5-9	1651	23953
10-14	1756	1407
15-19	1660	1517
20-24	1635	1416
25-29	1610	1388
30-34	1512	1361
35-39	1427	1258
40-44	1359	1167
45-49	1286	1096
50-54	1171	1018
55-59	1044	897
60-64	901	762
65-69	785	610
70-74	645	487
75-79	553	338
80-84	466	237
85-89	386	143
90-94	313	66
95-99	281	17
100+	272	2

Q3.c

We can compare our population projections for 2020 and 2030 with migration to the same projections without migration:

Table 9: Comparison of Peru female population projections, with and without migration

Age	2025	+Mig. 2025	2030	+Mig. 2030
0-4	23406	24129	24628	25629
5-9	1350	1651	23063	23953
10-14	1451	1756	1347	1407
15-19	1359	1660	1449	1517
20-24	1335	1635	1356	1416
25-29	1312	1610	1331	1388
30-34	1217	1512	1306	1361
35-39	1135	1427	1210	1258
40-44	1069	1359	1126	1167
45-49	999	1286	1059	1096
50-54	887	1171	986	1018
55-59	763	1044	872	897
60-64	623	901	744	762
65-69	510	785	598	610
70-74	371	645	479	487
75-79	281	553	333	338
80-84	194	466	235	237
85-89	115	386	142	143
90-94	42	313	66	66
95-99	10	281	17	17
100+	1	272	2	2

Here we see that the with-migration projections are always larger than the without-migration projections, since Peru has a net positive migration rate in all age groups (*Table 7*).

Q4

Q4.a

I used 5-year age groups based on the Life Table extracted in question 2. A log-linear fit applied to each of the models, defined below, along with the estimated model parameters after fitting our observed data:

Gompertz

$$(\alpha, \beta) = (0.003, 0.097)$$

$$\begin{aligned}\mu(x) &= \alpha e^{\beta x} \\ \log[\mu(x)] &= \log(\alpha) + \beta x\end{aligned}$$

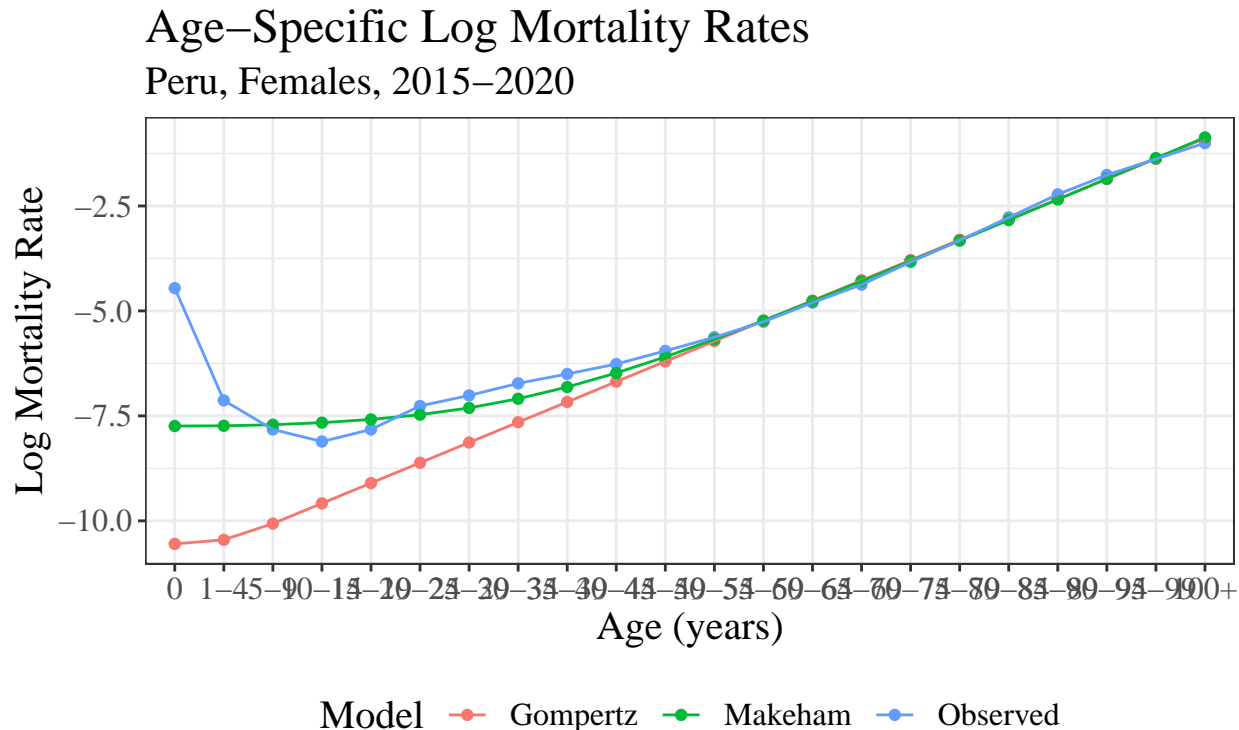
Gompertz-Makeham

$$(\alpha, \beta, \gamma) = (0.0027, 0.0987, 4 \times 10^{-4})$$

$$\begin{aligned}\mu(x) &= \gamma + \alpha e^{\beta x} \\ \log[\mu(x) - \gamma] &= \log(\alpha) + \beta x\end{aligned}$$

Note: both of these models were fit using the MortalityLaws package, optimizing the function $\log^2(\frac{est.}{obs.})$.

Q4.b



Plotting both fitted rates against the observed rates, we see that the additional constant in the *Gompertz-Makeham* enables the model to follow the observed trends in child and young adult mortality much closer. This presents a good case for using *Gompertz-Makeham* over just *Gompertz*.

Q4.c

The *Heligman-Pollard* model uses three terms to capture child mortality, the adult accident bump, and old-age mortality to predict odds of death:

$$\frac{q_x}{1 - q_x} = A^{(x+B)^2} + De^{-E(\log(x) - \log(F))^2} + GH^x$$

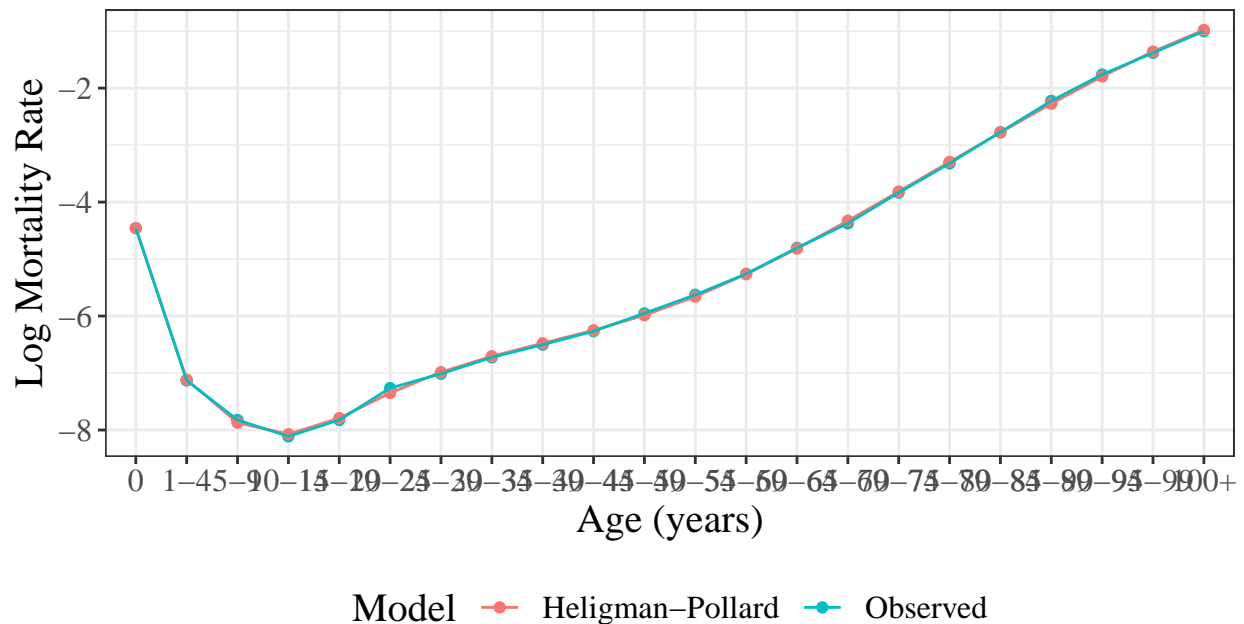
Fitting this model with our observed mortality rates, we get the model parameters:

(A, B, C, D, E, F, G, H) = 0.000802, 0.00071, 0.0651, 0.00105, 1.61, 44, 9.35e-06, 1.12

Note: this model was fit using the MortalityLaws package, optimizing the function $\log^2(\frac{est.}{obs.})$, and using M_x as an input (internal M_x to q_x conversion).

Age-Specific Log Mortality Rates

Peru, Females, 2015–2020



This model does a great job of capturing the high child mortality and gets closer to capturing the adult accident hump, following the observed data closely.

Q4.d

The *Brass relational model* estimates mortality given a “standard” reference mortality ($q^*(x)$) as an input. This standard mortality comes from the a *Coale-Demeny West* model life table, where the life table that was selected was the one with the lowest RMSE compared to our observed mortality (life table index #25). The model and estimated parameters are shown below:

$$(\alpha, \beta) = -0.702, 0.677$$

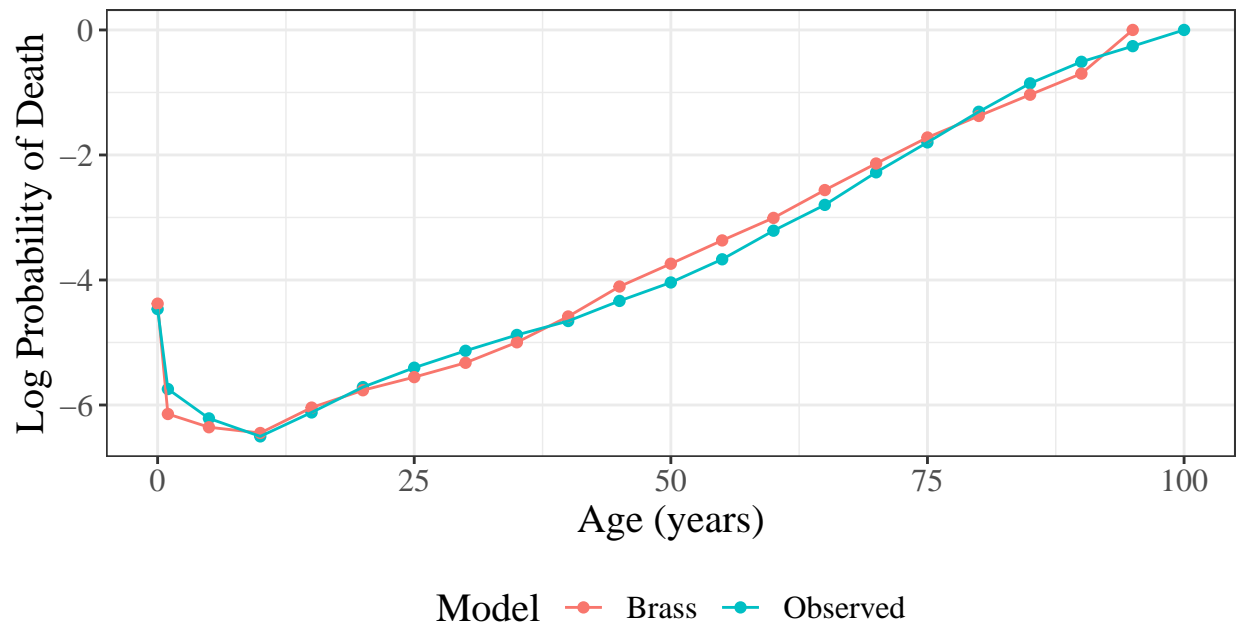
$$\text{logit}(q_{\alpha, \beta}(x)) = \alpha + \beta \times \text{logit}(q^*(x))$$

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

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

Age-Specific Log Probability of Death

Peru, Females, 2015–2020



Plotting the observed and estimated q_x values, we see that the *Brass* model also does a good job of capturing high child mortality, but underestimates the adult accident hump. In the old ages this model loses validity, since it was fix with a lower terminal age than is present in the observed data.

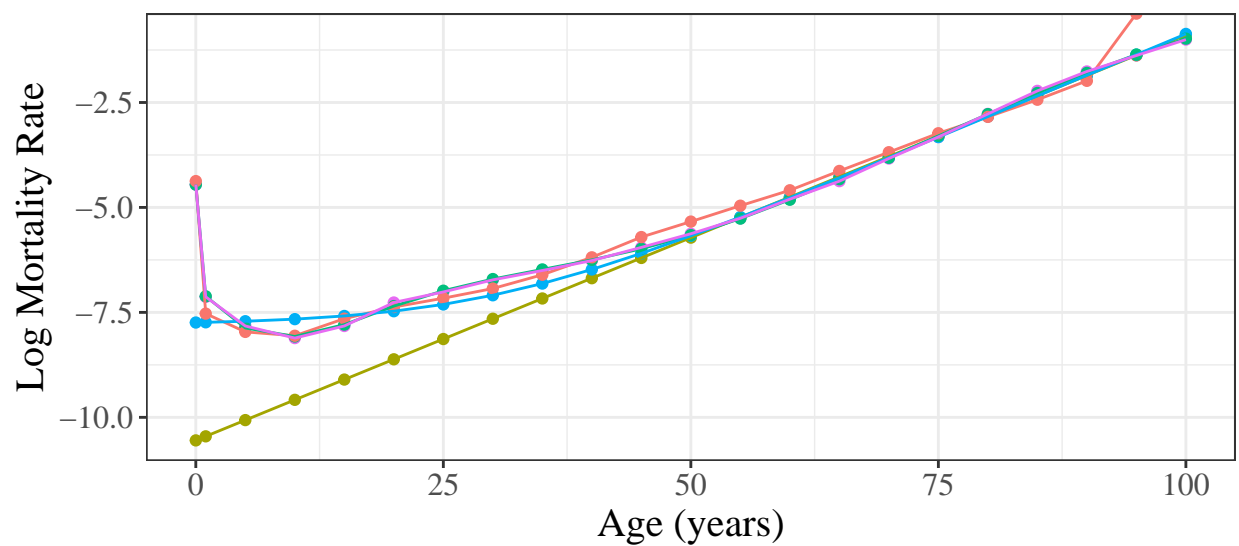
Q4.e

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

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

Age-Specific Log Mortality Rates

Peru, Females, 2015–2020



Model ● Brass ● Gompertz ● Heligman–Pollard ● Makeham ● Obser

Table 10: $RMSE$ of estimated $\log(m_x)$ against observed $\log(m_x)$

Model	RMSE
Brass	5.762
Heligman-Pollard	5.771
Makeham	6.016
Gompertz	6.976
n	7.239

Comparing all four of these models to the observed data, both visually and via the root mean-squared-error of log-transformed m_x , shows that the *Brass* model most closely follows the observed data.

Appendix

```
# Prep work -----

# Load libraries
library(tidyverse, quietly = TRUE)
library(wpp2019)
library(LifeTables)
library(demogR)
library(MortalityLaws)
library(Metrics)

# helper functions
```

```

## matrix printing for latex
write_matex <- function(x, digits = 3) {
  # From: https://stackoverflow.com/a/54088015/8866058
  x <- round(x, digits = digits)
  mat_string <- apply(x, 1, function(y) paste(y, collapse = "&"))
  paste("\\begin{bmatrix}", paste0(mat_string, collapse = "\\&"), "\\end{bmatrix}")
}

## shortcut for matrix multiplication and exponents
"%^%" <- function(A, n) {
  if (n == 1) {
    A
  } else {
    A %*% (A %^(n - 1))
  }
}

## creating a Leslie matrix from fertility and survival rates
make_leslie_matrix <- function(f, s) {
  if (length(f) != length(s)) {
    stop("f and s must be the same length")
  }

  n_size <- length(f)
  l_mat <- matrix(0, nrow = n_size, ncol = n_size)

  l_mat[1, ] <- f
  diag(l_mat[-1, ]) <- s[1:(n_size - 1)]
  l_mat[n_size, n_size] <- s[n_size]

  l_mat
}

## plotting log mortality models
plot_log_mort <- function(data, ...) {
  ggplot(data, aes(...)) +
    geom_point() +
    geom_line() +
    theme_bw() +
    theme(
      text = element_text(family = "serif", size = 15),
      legend.position = "bottom"
    ) +
    labs(
      title = "Age-Specific Log Mortality Rates",
      subtitle = "Peru, Females, 2015-2020",
      x = "Age (years)",
      y = "Log Mortality Rate",
      color = "Model"
    )
}

# Question 1 -----

```

```

pop_table <- tibble(
  age = c("1", "2", "3+"),
  pop = c(18, 17, 14) * 1000,
  fr = c(0, .9, .2),
  surv = c(.65, .75, .15)
)

knitr::kable(
  pop_table,
  booktabs = TRUE,
  caption = "Q1 one-sex closed population",
  col.names = c(
    "Age",
    "Population ($N_x$)",
    "Fertility Rate ($\\tilde{F}_x$)",
    "Survival Prob. ($s_x$)"
  ),
  eval = FALSE
)

# Question 1a -----

CBR <- pop_table %>%
  mutate(births = pop * fr,
         person_years = pop * surv) %>%
  summarise(cbr = sum(births) / sum(person_years)) %>%
  pull(cbr)

# Question 1b -----
## function for conversion fertility rate to expected number of female births
f_tilde_2_asfr <- function(F_tilde, srb, Sxm1, Nx1, Nx, q0) {
  F_tilde * (1 + srb) * 2 / (1 + Sxm1 * (Nx1/Nx)) / (1 - q0/2)}

pop_asfr_1 <- 0
pop_asfr_2 <- f_tilde_2_asfr(0.9, 1.05, .65, 18000, 17000, 1 - 0.65)
pop_asfr_3 <- f_tilde_2_asfr(0.2, 1.05, .75, 17000, 14000, 1 - 0.65)

pop_asfr <- c(pop_asfr_1, pop_asfr_2, pop_asfr_3)

TFR <- sum(pop_asfr)
tfr_eqn <- paste0(round(pop_asfr, 3), collapse = " + ")

# Question 1c -----

pop_leslie <- make_leslie_matrix(pop_table$fr, pop_table$surv)

# Question 1d -----

pop_t0 <- matrix(pop_table$pop)
pop_t1 <- pop_leslie %*% pop_t0

```



```

# Question 1e -----
pop_t2 <- (pop_leslie %~% 2) %*% pop_t0

# Question 1f -----

CBR_t2 <- pop_table %>%
  mutate(
    pop = as.vector(pop_t1),
    births = pop * fr,
    person_years = pop * surv
  ) %>%
  summarise(cbr = sum(births) / sum(person_years)) %>%
  pull(cbr)

pop_asfr_t2_1 <- 0
pop_asfr_t2_2 <- f_tilde_2_asfr(0.9, 1.05, .65, pop_t1[1], pop_t1[2], 1 - 0.65)
pop_asfr_t2_3 <- f_tilde_2_asfr(0.2, 1.05, .75, pop_t1[2], pop_t1[3], 1 - 0.65)

pop_asfr_t2 <- c(pop_asfr_t2_1, pop_asfr_t2_2, pop_asfr_t2_3)

TFR_t2 <- sum(pop_asfr_t2)
tfr_eqn_t2 <- paste0(round(pop_asfr_t2, 3), collapse = " + ")

# Question 1g -----

pop_right_eigen <- eigen(pop_leslie)
dominant_right_index <- which.max(abs(pop_right_eigen$values))

pop_iroi <- log(pop_right_eigen$values[dominant_right_index])

# Question 1h -----

pop_sad <- matrix(pop_right_eigen$vectors[, dominant_right_index])

# Question 1i -----

pop_left_eigen <- eigen(t(pop_leslie))
dominant_left_index <- which.max(abs(pop_left_eigen$values))

pop_repv <- matrix(pop_left_eigen$vectors[, dominant_left_index])

# Question 2 -----
## Question 2a -----
data(mxF)
peru_mx <- mxF %>%
  filter(name == "Peru") %>%
  select(age, mx = `2015-2020`)

```

```

# Builds a life table by using the mortality rate
peru_LT =
  LifeTables::lt.mx(peru_mx$mx, age = peru_mx$age)$lt %>%
  as_tibble() %>%
  mutate(age =
    paste0(Age, "-", Age + 4) %>%
    magrittr::inset(c(1, 2, length(Age)), c("0", "1-4", "100+")), .before=1) %>%
  mutate(age = as_factor(age))

peru_mx %>%
  knitr::kable(
    booktabs = TRUE,
    escape = TRUE,
    digits = 3,
    col.names = c("Age", "$_nq_x$"),
    eval = FALSE,
    caption = "Age-specific Mortality Rates of Peru female population, 2015-2020")

# Question 2b -----
LT_colnames <- c(
  "Age",
  "$a_x$",
  "$_{5}m_x$",
  "$_{1}q_x$",
  "$_{1}s_x$",
  "$_{5}d_x$",
  "$l_x$",
  "$_{5}L_x$",
  "$_{5}T_x$",
  "$e_x$"
)

peru_LT %>%
  select(-Age) %>%
  knitr::kable(
    booktabs = TRUE,
    col.names = LT_colnames,
    eval = FALSE,
    digits = c(0, 1, 3, 3, 3, 0, 0, 2, 0, 0, 2),
    caption = "Life table for 2015-2020 Peru female population")

# Create standard 5-year age group npx and nqx mortality
peru_0_to_5 = peru_LT %>%
  filter(Age < 5) %>%
  summarise(age = "0-4", nqx = 1 - prod(npx), npx = prod(npx))

peru_q0 <- peru_0_to_5$nqx[1]

peru_nqx <- peru_LT %>%
  filter(Age >= 5) %>%
  select(age, nqx, npx) %>%
  bind_rows(peru_0_to_5, .)

```

```

peru_nqx%>%
  knitr::kable(
    booktabs = TRUE,
    col.names = c("Age", "$_nq_x$", "$_ns_x$"),
    eval = FALSE,
    digits = c(0, 2, 2),
    caption = "Peru $_ns_x$: 0 and 1-4 combined")
data(popF)
data(sexRatio)
data(tfr)
data(percentASFR)

peru_pop <- popF %>%
  filter(name == "Peru") %>%
  select(age, pop = `2015`) %>%
  mutate(row_id = row_number(), .before=1)

peru_srb <- sexRatio %>% filter(name == "Peru") %>% pull(`2015-2020`)
peru_tfr <- tfr %>% filter(name == "Peru") %>% pull(`2015-2020`)

asfr_2_f_tilde <- function(asfr, srb, Sxm1, Nxm1, Nx, q0) {
  asfr * (1 / (1 + srb)) * .5 * (1 + Sxm1 * (Nxm1 / Nx)) * (1 - q0 / 2)}

peru_Fx <- percentASFR %>%
  filter(name == "Peru") %>%
  select(age=age, pasfr = `2015-2020`) %>%
  mutate(asfr = pasfr * peru_tfr / 5) %>%
  right_join(peru_pop, by = "age", ) %>%
  replace_na(list(pasfr = 0, asfr = 0)) %>%
  left_join(peru_nqx, by = "age") %>%
  mutate(
    f_tilde =
      asfr_2_f_tilde(asfr, peru_srb, lag(npx), lag(pop), pop, peru_q0)
  ) %>%
  replace_na(list(f_tilde = 0)) %>%
  arrange(row_id) %>%
  select(-row_id)

knitr::kable(
  select(peru_Fx, age, f_tilde, asfr, pop, npx),
  booktabs = TRUE,
  col.names =
    c("Age", "${}_5\\tilde{F}_x$", "${}_5F_x$", "${}_5N_x$", "${}_5s_x$"),
  eval = FALSE,
  digits = c(0, 3, 3, 0, 3),
  caption = paste(
    "Expected number of live female births per woman per five-year period",
    "in Peru, 2015-2020"
  )
)
peru_leslie <- with(peru_Fx, make_leslie_matrix(f_tilde, npx))

```

```

peru_leslie_idx <- which(peru_leslie != 0, arr.ind = TRUE)

peru_leslie_tbl <-
  peru_leslie_idx %>%
  as_tibble() %>%
  mutate(value = peru_leslie[peru_leslie_idx]) %>%
  arrange(row, col)
knitr::kable(
  peru_leslie_tbl,
  booktabs = TRUE,
  col.names = c("Row", "Column", "Value"),
  digits = c(0, 0, 3),
  caption = "Leslie matrix for Peru females, 2015-2020, (sparse format)"
)
# Question 2d -----
peru_pop_t1 <- peru_leslie %*% matrix(peru_Fx$pop)

# Question 2.e -----
peru_pop_t2 <- (peru_leslie %^% 2) %*% matrix(peru_Fx$pop)

# Question 3 -----
## Net migration is per 1000 person-years
data(migration)

peru_net_mig <- migration %>%
  filter(name == "Peru") %>%
  pull(`2015-2020`)

peru_mig <- peru_pop %>%
  mutate(
    mig_rate = (peru_net_mig / 1000) * (pop / sum(pop)),
    mig_rate_fmt = formatC(mig_rate, digits = 2, format = "e"),
    mig_num = pop * mig_rate)

peru_mig %>%
  select(age, mig_rate_fmt) %>%
  knitr::kable(
    booktabs = TRUE,
    col.names = c("Age", "${}_{{5}}G_x$"),
    eval = FALSE,
    caption = "Age-specific migration rate in Peru, 2015-2020")
# Question 3b -----
peru_pop_mig_t1 <- `+`(
  (peru_leslie %*% matrix(peru_mig$pop)),
  (.5 * (1 + peru_leslie)) %*% matrix(peru_mig$mig_num)
)

peru_pop_mig_t2 <- (peru_leslie %^% 2) %*% matrix(peru_mig$pop) + (peru_leslie %^% 2) %*% matrix(peru_mig$mig_num)

peru_pop_mig_tbl <- tibble(
  age = peru_pop$age,
  mig_pop_2025 = peru_pop_mig_t1,
  mig_pop_2030 = peru_pop_mig_t2
)

```

```

)

knitr::kable(
  mutate_if(peru_pop_mig_tbl, is.numeric, round),
  booktabs = TRUE,
  col.names = c("Age", "2025", "2030"),
  caption = "Peru female with-migration population projections"
)
peru_pop_all_tbl <- peru_pop_mig_tbl %>%
  mutate(
    pop_2025 = peru_pop_t1,
    pop_2030 = peru_pop_t2) %>%
  select(age, pop_2025, mig_pop_2025, pop_2030, mig_pop_2030)

knitr::kable(
  mutate_if(peru_pop_all_tbl, is.numeric, round),
  booktabs = TRUE,
  col.names = c("Age", "2025", "+Mig. 2025", "2030", "+Mig. 2030"),
  caption = paste(
    "Comparison of Peru female population projections,",
    "with and without migration"
  )
)
)
# Question 4a -----
model_gompertz2 <- lm(log(nmx) ~ age, data = filter(peru_LT, Age >= 50))

model_gompertz <- with(
  filter(peru_LT, Age >= 50),
  MortalityLaws::MortalityLaw(x = Age, mx = nmx, law = "gompertz")
)

model_makeham <- with(
  filter(peru_LT, Age >= 50),
  MortalityLaw(x = Age, mx = nmx, law = "makeham")
)

coef_gompertz <- coef(model_gompertz)
coef_makeham <- coef(model_makeham)

# Question 4b -----

peru_LT_model <- peru_LT %>%
  rename(obs_qx = nqx, obs_mx = nmx) %>%
  mutate(
    Gompertz = predict(model_gompertz, x = Age),
    Makeham = predict(model_makeham, x = Age)
  ) %>%
  select(age, everything())

peru_LT_model %>%
  select(-obs_qx, age, Observed = obs_mx, everything()) %>%

```

```

mutate(age = as_factor(age)) %>%
pivot_longer(-age, names_to = "model", values_to = "mx") %>%
filter(model %in% c("Observed", "Gompertz", "Makeham")) %>%
plot_log_mort(x = age, y = log(mx), color = model, group = model)

# Question 4c -----
n =
model_HP_mx <- with(peru_LT, MortalityLaw(x = Age, mx = nm, law = "HP"))
model_HP_qx <- with(peru_LT, MortalityLaw(x = Age, qx = nqx, law = "HP"))

peru_LT_model <- peru_LT_model %>%
  mutate(
    n = lead(Age) - Age,
    `Heligman-Pollard` = predict(model_HP_mx, x = Age),
    `Heligman-Pollard2` = -1 * log(1 - predict(model_HP_qx, x = Age)) / n)

coef_HP <- coef(model_HP_mx)

peru_LT_model %>%
  select(-obs_qx, age, Observed = obs_mx, everything()) %>%
  pivot_longer(-age, names_to = "model", values_to = "mx") %>%
  filter(model %in% c("Observed", "Heligman-Pollard")) %>%
  plot_log_mort(x = age, y = log(mx), color = model, group=model)

# Question 4d -----
coaleDemenyLTW <- demogR::cdmltw(sex = "F")
peru_mx_0to95 <- peru_LT %>% filter(Age < 100) %>% pull(nm)

best_match_lt <-
  coaleDemenyLTW[["nm"]] %>%
  apply(1, function(x) rmse(x, log(peru_mx_0to95))) %>%
  which.min()

peru_LT_model <- peru_LT_model %>%
  mutate(
    obs_qx_95 = obs_qx,
    standard_qx = c(coaleDemenyLTW[["nqx"]][best_match_lt, ], NA)
  )

peru_LT_model[["obs_qx_95"]][21:22] <- c(1, NA)

model_brass <- lm(
  qlogis(obs_qx_95) ~ qlogis(standard_qx),
  data = filter(peru_LT_model, Age < 95)
)

peru_LT_model <- peru_LT_model %>%
  mutate(
    brass_pred_fit_qx = c(plogis(model_brass$fitted.values), 1, NA),
    brass_pred_obs_qx = plogis(predict(model_brass, data.frame(standard_qx = obs_qx))),
    brass_pred_fit_mx = -1 * log(1 - brass_pred_fit_qx) / n
  )

```

```

coef_brass <- coef(model_brass)

peru_LT_model %>%
  select(
    Age, Observed = obs_qx,
    Brass = brass_pred_fit_qx, Brass2 = brass_pred_obs_qx
  ) %>%
  pivot_longer(-Age, names_to = "model", values_to = "qx") %>%
  filter(model %in% c("Observed", "Brass")) %>%
  plot_log_mort(x = Age, y = log(qx), color = model, group=model) +
  labs(
    title = "Age-Specific Log Probability of Death",
    y = "Log Probability of Death"
  )

# Question 4.e -----
peru_LT_compare <- peru_LT_model %>%
  mutate(n = lead(Age) - Age) %>%
  select(
    Age, n,
    Observed = obs_mx,
    Gompertz,
    Makeham,
    `Heligman-Pollard`,
    Brass = brass_pred_fit_mx
  )

# knitr::kable(
#   peru_LT_compare,
#   booktabs = TRUE,
#   digits = 3,
#   caption = "Comparison of models against observed Mx in peru females, 2015-2020"
# )

peru_LT_compare %>%
  pivot_longer(!c(Age, n), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Gompertz", "Makeham", "Heligman-Pollard", "Brass")) %>%
  plot_log_mort(x = Age, y = log(Mx), color = model, group=model)

peru_LT_compare_fit <- peru_LT_compare %>%
  filter(Age < 95) %>%
  select(-Age) %>%
  mutate_all(log) %>%
  summarise_all(~rmse(., Observed)) %>%
  select(-Observed) %>%
  pivot_longer(everything(), names_to = "Model", values_to = "RMSE") %>%
  arrange(RMSE)

knitr::kable(
  peru_LT_compare_fit,
  booktabs = TRUE,
  digits = 3,
  caption = "$RMSE$ of estimated $log(m_x)$ against observed $log(m_x)$"
)

```

)