Problem Set 2: Mortality Change in the United States

Due 11:59 PM, Thursday, April 24

In this problem set, you will analyze data from the Vital Statistics of the United States, 1933-2023. The template `Rmd` script loads a dataset that contains counts of deaths and population by age and year. Each row is an age-by-year combination; female, male, and total counts are provided.

Many questions ask you to create tables or graphs. When asked to make quantitative statements about a graph, it is fine to approximate based on looking at the graph.

1. Calculate female, male, and total age-specific mortality rates. Create a table that reports the average age-specific mortality rate by sex. On average, do females or males have higher age-specific mortality rates?

2. Using only 2023 data, create two graphs. In the first, plot the age-specific mortality rate against age separately for females and males. In the second, plot the natural logarithm of the age-specific mortality rate against age separately for females and males. Is the male-female difference in mortality rates between larger among 20-30 year-olds or 90-100 year-olds? Is the male-female ratio of mortality rates larger among 20-30 year-olds or 90-100 year-olds?

3. Calculate the crude mortality rate in each year for the total population. Plot the crude mortality rate over time. Do you think the graph accurately reflects changes in the population's mortality burden? Why or why not?

4. Split off a new data frame that only keeps every tenth age (0, 10, $\cdots$ 90). Use it to draw a graph in which the x-axis is year, the y-axis is the natural logarithm of the age-specific mortality rate, and each age gets its own line plot with a different color.[1] Which ages saw the largest proportional declines in mortality over the period? Which ages show evidence of a mortality increase during the COVID-19 pandemic?

5. Calculate the age-standardized mortality rate using the population age structure in 2023. Plot the age-standardized mortality rate over time. How does the graph for the age-standardized mortality rate compare with the graph for the crude mortality rate in Question 3? Explain.

6. Compute the period life table functions (q, l, d, L, T, e) in each year for the total population. Assume that deaths happen halfway through the year on average, and that anyone who survives to age 110 dies within the next year. Plot life expectancy at birth over time. How much did life expectancy at birth increase between 1933 and 2019 (just before the COVID-19 pandemic)? How much did life expectancy at birth decrease during the pandemic?

[1] Because `age` is a continuous variable, you should specify `color(factor(age))` in the `ggplot` aesthetic, where `factor()` tells R that you want to treat the levels of `age` as discrete.

7. Draw the survivorship curve for the total population in 2023. If a group of people were born and then experienced 2023 age-specific mortality rates over their lives, at roughly what age would half of them have died?

8. Create a table of q_0 in every tenth year (1933, 1943, $\cdots$ 2023). How did the probability of dying in infancy change between 1933 and 2023?

9. Change infant mortality in all years to the 1933 level. Recompute l, d, L, T, and e. Graph the life expectancy at birth over time. How much lower would life expectancy at birth have been in 2023 if infant mortality were at the 1933 level?

10. You calculated *period* life expectancy above. How is it different from *cohort* life expectancy? Calculate cohort life expectancy at age 65 (e_{65}) for as many cohorts as possible. Plot e_{65} against year of birth.