

Finding Life Expectancy at Age 30

Sonam T Gurung

```
library(tidyverse)
```

```
micro_data <- read_csv("microdata.csv")
```

Life Expectancy Analysis

```
# Check for missing values for birth year  
sum(is.na(micro_data$birth_year))
```

```
## [1] 0
```

```
# No missing birth year values
```

```
# no. of individuals who are dead as of 2003  
sum(is.na(micro_data$death_year))
```

```
## [1] 548806
```

```
# no. of individuals alive as of 2003  
nrow(micro_data) - sum(is.na(micro_data$death_year))
```

```
## [1] 170100
```

Calculate the mortality rate with respect to age for 2003:

```
# Calculate age in 2003 for each individual.  
# Two parts to this process to find the age in 2003 for each individual  
# - If an individual died in or before 2003, calculate their age at death.  
# - If an individual is alive or died after 2003, subtract their birth year from 2003 to determine their age in 2003.  
  
micro_data <- micro_data %>%  
  mutate(age_2003 = ifelse(death_year <= 2003 & !is.na(death_year),  
                           death_year - birth_year,  
                           2003 - birth_year))
```

For each age, calculate the mortality rate as the fraction of individuals who reached that age in 2003 and then died that year.

Count the number of people who reached each age in 2003

```

# This includes people who died in 2003, as they were alive for some part of the year
age_counts <- micro_data %>%
  filter(is.na(death_year) | death_year >= 2003) %>%
  count(age_2003, name = "reached_age")

# Count the number of people who died in 2003 at each age
died_in_2003 <- micro_data %>%
  filter(death_year == 2003) %>%
  count(age_2003, name = "died")

# Merge both the age and death counts by age in 2003
# Calculate the mortality rate by dividing the number of people who died in a given year
# by the total number of people who reached that age in that year
# Replace any NA death counts with 0 in the data

mortality_rates <- left_join(age_counts, died_in_2003, by = "age_2003") %>%
  mutate(died = replace_na(died, 0),
         mortality_rate = died / reached_age) %>%
  arrange(age_2003)

```

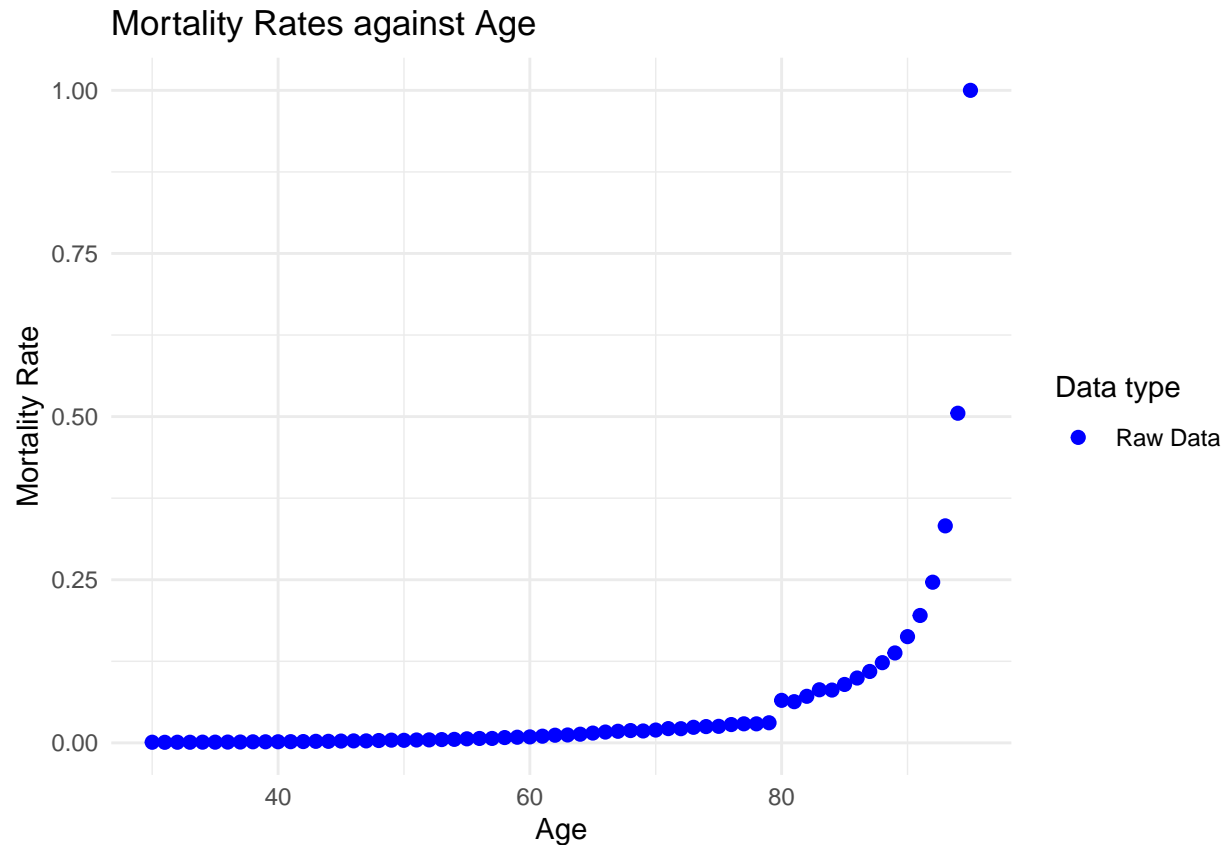
Plot of these raw mortality rates against age.

```

# Scatter plot of raw mortality rates against age
mortality_rate_plot <- mortality_rates %>%
  ggplot(aes(x = age_2003)) +
  geom_point(aes(y = mortality_rate, color = "Raw Data"), size = 2) +
  labs(x = "Age",
       y = "Mortality Rate",
       title = "Mortality Rates against Age",
       color = "Data type") +
  scale_color_manual(values = c("Raw Data" = "blue", "Smoothed Curve" = "red")) +
  theme_minimal()

mortality_rate_plot

```



Smooth the raw mortality rates:

Fit a 20-degree polynomial to the raw mortality rates

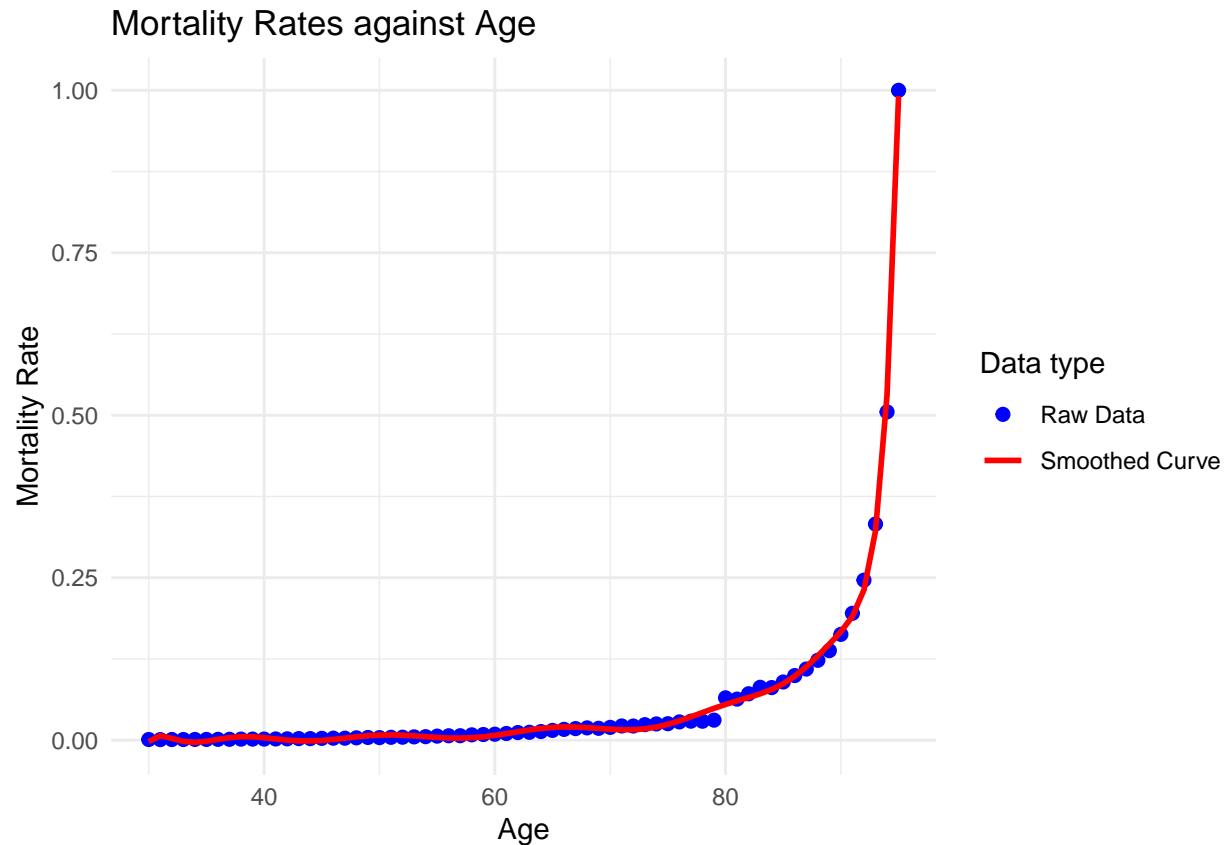
```
# Fit a 20-degree polynomial to the raw mortality rates
# The poly function is used to create polynomial terms for age_2003 with a degree of 20
# The raw = TRUE argument ensures that the polynomial terms are not orthogonalized, or raw polynomials
poly_model <- lm(mortality_rate ~ poly(age_2003, degree = 20, raw = TRUE), data = mortality_rates)

summary(poly_model)
coefficients <- coef(poly_model)
```

Use the fitted model to predict smoothed mortality rates for each age.

```
# Use the fitted model to predict smoothed mortality rates for each age
mortality_rates$pred_mortality_rate <- predict(poly_model)
```

Add this smoothed curve to my plot raw mortality rates against age plot



Calculate life expectancy at age 30 by aggregating over the implied survival curve obtained from the smooth function

Convert the smoothed mortality rates into survival rates by subtracting the smoothed mortality gradient from 1. This gives me the survival curve, which represents the share of individuals alive at age a who survive through age a .

```
# Subtracting the smoothed mortality gradient from 1 to get the survival curve
mortality_rates$survival_rate <- 1 - mortality_rates$pred_mortality_rate
```

Calculate the running product of these age-conditional survival rates. This represents the share of people who survive through age 30 and also survive through each subsequent age.

```
# Calculate the running product of survival rates (cumulative survival probability)
mortality_rates$cumulative_survival <- cumprod(mortality_rates$survival_rate)
```

Sum over this running product. This sum represents the expected additional years of life beyond age 30.

```
# Sum the cumulative survival probabilities
expected_additional_years <- sum(mortality_rates$cumulative_survival)
```

Add 30 to this sum to obtain the life expectancy at age 30.

```
life_expectancy_at_30 <- expected_additional_years + 30
print(life_expectancy_at_30)
```

```
## [1] 78.94615
```

A function that performs the steps outlined above

The function requires the dataset and the mortality_rate as input parameters, which can be either smooth or observed

```
get_life_exp_at_30 <- function(data, mort){  
  data %>%  
    mutate(survival_rate = 1 - mort,  
           cumulative_survival = cumprod(survival_rate)) %>%  
    summarize(life_exp_at_30 = round(sum(cumulative_survival) + 30, 2))  
}  
  
get_life_exp_at_30(mortality_rates, mortality_rates$pred_mortality_rate)$life_exp_at_30
```

```
## [1] 78.95
```

The life expectancy at age 30 is 78.95 years.

A life expectancy of 79 years means that a 30-year old is expected to live till 79 on average, assuming they experience the age-specific death rates of 2003 throughout their life. The Period Life Expectancy is based on a specific year and does not take unforeseen situations in the future into account like the pandemic.