FCDS

Alexandria University

Probability and Statistics II
2024

Life Expectancy Evolution

Charting the Course of Lifespan Growth Over a Century

I. INTRODUCTION

In just over a century, the average lifespan has more than doubled, from a mere 32 years in 1900 to an impressive 71 years by 2021. This remarkable leap in longevity prompts an intriguing exploration into its genesis and implications. While the decline in child mortality has historically bolstered life expectancy, the shift towards longer lifespans is not solely confined to the young. Individuals across all age groups now face reduced mortality risks, contributing to delayed occurrences of death. This striking evolution finds its roots in advancements in healthcare, fortified public health measures, and elevated global living standards. Despite assumptions that the rise in life expectancy was primarily due to reduced child mortality rates, contemporary trends highlight broader declines in mortality across age categories. Here, we want to analyze the complex factors propelling this global shift in life expectancies, uncovering the intricate tapestry that extends human lifespans worldwide.

II. GENERAL RULES

1. ZERO will be assigned to copied reports.
2. You should work alone or with one student as maximum.
3. The report will be evaluated from 10 and bonus marks will be given to outstanding reports.
4. A report must be prepared and include (introduction about your project, your Rcode, results, comments on results, and problems you have faced and how you overcame it
5. A soft copy of the report must be sent to "probabilityii.fcds23@gmail.com" before the final discussions, moreover, a hardcopy from report must be broughtduring the discussion.You must respond to all project questions (parts).
6. Deadline will be at "13/5/2024" with excuses no for lateness

III. QUESTIONS

- Download the necessary data via the URL below

https://drive.google.com/file/d/1W0pJyWi-5dly9dvdQCy4hTQVsRoq7Qtg/view?usp=sharing

- Import the dataset (called " life-expectancy-at-different-ages.csv") into RStudio and store it in a variable called "Life_Expectancy."

1. Generate a precise summary of the data, including details such as the minimum value, maximum value, first quartile (25t percentile), third quartile (75th percentile), median (50th percentile), and mean, using RStudio's built-in summary function for the Life_Expectency data with 9 columns and 20,756 rows.

2. In statistics, there are two main categories of variables: quantitative and qualitative. Can you explain the difference between these two types of variables? Then, categorize the Life_Expectency dataset's variables into these two categories.

3. Draw the histogram for the " Age 0 lifespan " variable.

4. By inspection (from the histogram), which distributions may be employed to provide a good fit for the " Age 0 lifespan " histogram?

5. Assuming that the observations in the dataset represent the population, and that the distribution of ages is normal, determine the true mean and standard deviation for the " Age 0 lifespan " variable.

6. Take a random sample of size 50 from " Age 0 lifespan ". Using this sample, what is your point estimate of the population mean and standard deviation?

7. Since you have access to the entire population, simulate the sampling distribution for the mean " Age 0 lifespan " by taking 30 samples from the population of size 50 and computing 30 sample means. Store these means in a vector called sample means30. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean age of the population

8. Simulate the sampling distribution for the mean " Age 0 lifespan " by taking 75 samples from the population of size 50 and computing 75 sample means. Store these means in a vector called sample_means75. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean age of the population to be?

9. Simulate the sampling distribution for the mean " Age 0 lifespan " by taking 500 samples from the population of size 50 and computing 500 sample means. Store these means in a vector called sample_means500. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean age of the population to be?

10. What happens to the sampling distribution when the number of samples increases?

11. Given your access to the entire population, simulate the sampling distribution for the average " Age 0 lifespan " by drawing 1500 samples from a population of size 20 and calculating 1500 sample means. Save these means in a vector named sample_means_s20. Plot the data and describe the shape of this sampling distribution. Based on this distribution, what would be your estimate for the mean age of the entire population?

12. Simulate the sampling distribution for the average " Age 0 lifespan " by drawing 1500 samples from a population of size 100 and computing 1500 sample means. Save these means in a vector named sample_means_s100. Plot the data and describe the shape of this sampling distribution. Based on this distribution, what would be your estimate for the mean age of the entire population?

13. Simulate the sampling distribution for the average " Age 0 lifespan " by drawing 1500 samples from a population of size 200 and calculating 1500 sample means. Save these means in a vector named sample_means_s200. Plot the data and describe the shape of this sampling distribution. Based on this distribution, what would be your estimate for the mean age of the entire population?

14. Explore the changes in the sampling distribution as the size of each sample increases. Assess the compatibility of the results with the central limit theorem.

15. Since you have access to the entire population, simulate the sampling distribution for the sample variances by drawing 1500 Samples from a population of size 2 and calculating 1500 sample variances. Save these sample variances in a vector named sample_means_s200.plot the data and describes the shape of this sampling distribution. Based on this distribution , what would be your estimate for the mean age of the entire population ?

16. Generate the sampling distribution for the sample variances by drawing 1500 samples from a population of size 50 and computing the sample variance for each. Save these sample variances in a vector named sample_U1500.
    Plot the data and describe the shape of this sampling distribution of variances. Regarding the shape:
    How does the shape of the sampling distribution change?
    How can this sampling distribution be utilized to estimate the population variance?

17. Extract a random sample of size 50 from the " Age 25 lifespan " variable. Utilize this sample to estimate the mean "Age 25 lifespan" of the population using Maximum Likelihood Estimation (MLE) and Method of Moments Estimation (MME). Subsequently, determine the bias of these estimators.

18. Draw a random sample of size 200 from the " Age 25 lifespan " variable. Use this sample to estimate the mean "Age 25 lifespan" of the population with MLE and MME. Analyze whether the bias increases or decreases as the sample size grows. Identify which estimators prove to be the most effective.

19. Extract a random sample of size 10 from the " Age 45 lifespan " variable. Using this sample, calculate a 95% confidence interval for the mean of the ages.

20. Extract a random sample of size 50 from the " Age 45 lifespan " variable. Using this sample, calculate a 95% confidence interval for the mean of the ages.

21. Draw a random sample of size 200 from the " Age 45 lifespan " variable. Using this sample, multiply the ages by 5. Analyze the impact on the expected value (E(age)) and variance (var(age)) after multiplication.

22. Draw a random sample of size 200 from the " Age 45 lifespan " variable. Using this sample, add 5 to the ages. Analyze the impact on the expected value (E(age)) and variance (var(age)) after addition.

23. Additional Task: Research about (p-value) and apply it on two examples