

Analyzing the Distribution of Suicide Rates Across Different Age Groups

By Katelyn Wong

Statement to Accompany the Writing Sample

This short research sample demonstrates my ability to apply statistical methods and data analysis to address real-world problems, skills essential to data science. In this study, I utilized R for data cleaning, visualization, and statistical analysis, leveraging techniques like ANOVA and Tukey's HSD tests to draw meaningful conclusions from a complex dataset about international suicide rates. I also accounted for challenges such as violations of normality assumptions, showcasing my ability to handle data analysis obstacles effectively.

Through this project, I refined my skills in hypothesis testing, data visualization, and presenting actionable insights. These experiences strengthened my commitment to using data science to address societal issues, particularly within social science contexts. This aligns with my goal to pursue graduate studies in data science, where I can build on my expertise and contribute to impactful, interdisciplinary research.

Background

Suicide rates are frequently studied to determine what existing structures can be changed to mitigate these deaths. Since suicide can occur at all stages of life due to various life circumstances and can deeply impact the community around the individual, studying the differences in suicide rates between different life stages can provide insight into suicide trends, causes, and risk factors across age groups. Therefore, this research aims to determine whether there are differences in suicide rates across different age groups. The null hypothesis is that

suicide rates do not differ between different age groups. The alternative hypothesis is that suicide rates do differ between different age groups.

An additional sub-question this research aims to address is whether suicide rates are highest among the 55-74 age group. The sub-alternative hypothesis posits that suicide rates are highest in this age group, based on the rationale that midlife crises often occur around the age of 50, potentially leading to increased suicide rates in subsequent years. The sub-null hypothesis is that suicide rates are not the highest among the 55-74 age group, meaning that there is at least one other age group with a higher suicide rate than the 55-74 age group. These hypotheses can be tested using the post-hoc test comparisons from this analysis.

Method

Participants

This analysis utilizes suicide statistics from the World Health Organization and its Mortality Database online tool. The dataset, which is open access on Kaggle, consists of aggregated sums of suicides from around the world by country, year, age group, and sex from 1979 to 2016. Each entry contains the number of suicides and the total population within a certain country, year, sex group, and age group.

Measures

Because my research explores suicide rates across different age groups, the variables I utilized were age group (categorical independent variable) and the average proportion of suicide count (numerical dependent variable). This dependent variable is a synthetic variable I created using the variables suicide count and population count. I decided to use the proportion numbers because total population count was consistently significantly different across age groups in all

countries, which meant that raw suicide counts could be skewed towards age groups with higher population numbers. For example, the average population count of the 35-54 age group in each country is 2,799,146 people, whereas the average population count of the 75+ age group in each country is 500,690 people. Taking the proportion of suicide numbers per age group (suicide count / total population count) would then normalize the values. Additionally, due to the nature of the dataset, which measured suicide rates per year and country, I decided to average the data entries across years and countries within each age group. This approach allowed me to aggregate the suicide numbers and focus solely on the impact of age group on suicide rates.

To answer the question of whether suicide rates differ across age groups, I utilized a one-way ANOVA test and subsequent post-hoc tests to analyze significant differences between each age group pairing. The dataset failed the initial ANOVA assumptions of normality of residuals (Shapiro-Wilk test results: $W = 0.75, p < .001$) and homogeneity of variances (Levene's Test results: $F(5, 36054) = 1253.3, p < .001$), but because the sample size was quite large ($N = 36054$), the ANOVA test would still most likely be robust to the violations of normality and homogeneity.

Results

A one-way ANOVA was conducted to examine the differences in average suicide rates across six age groups. The results showed that the effect of age on average suicide rates was statistically significant, $F(5, 36054) = 1185, p < .001, \eta^2 = 0.14$. Subsequent post-hoc tests (Tukey's Honestly Significant Difference) indicated that the 5-14 age group had significantly lower suicide rates compared to all other groups. Specifically, the differences were as follows: compared to the 15-24 age group ($M_d = -0.008, 95\% \text{ CI } [-0.009, -0.007], p < .001$), the 25-34 age group ($M_d = -0.012, 95\% \text{ CI } [-0.013, -0.011], p < .001$), the 35-54 age group ($M_d = -0.015, 95\%$

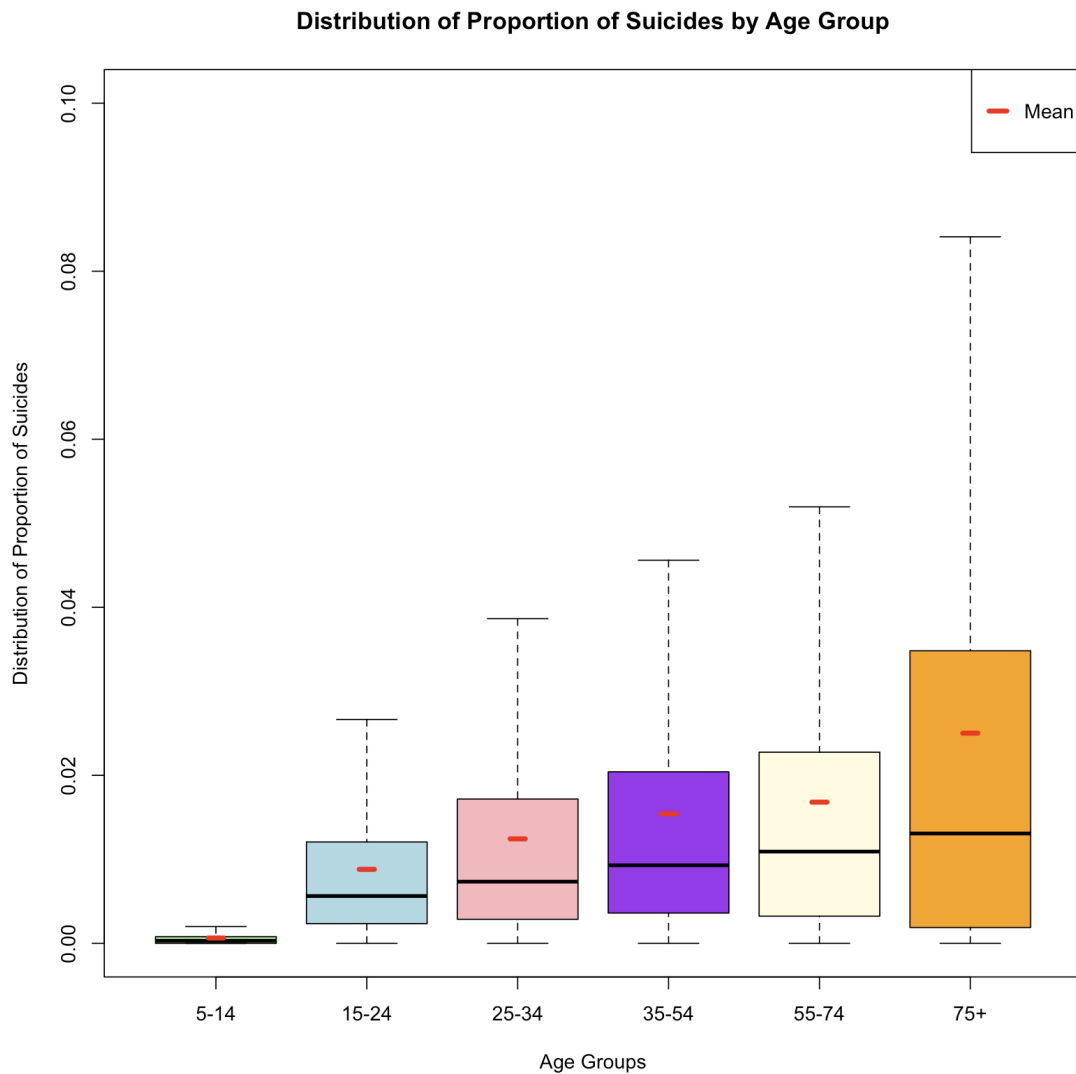
CI [-0.016, -0.014], $p < .001$), the 55-74 age group ($M_d = -0.016$, 95% CI [-0.017, -0.015], $p < .001$), and the 75+ age group ($M_d = -0.024$, 95% CI [-0.025, -0.023], $p < .001$). The post-hoc tests (Tukey's Honestly Significant Difference) also indicated that the 75+ age group had significantly higher suicide rates compared to all other groups. Specifically, the differences were as follows: compared to the 5-14 age group ($M_d = 0.024$, 95% CI [0.023, 0.025], $p < .001$), the 15-24 age group ($M_d = 0.016$, 95% CI [0.015, 0.017], $p < .001$), the 25-34 age group ($M_d = 0.013$, 95% CI [0.012, 0.014], $p < .001$), the 35-54 age group ($M_d = 0.010$, 95% CI [0.009, 0.011], $p < .001$), and the 55-74 age group ($M_d = 0.008$, 95% CI [0.007, 0.009], $p < .001$ (Table 1).

Table 1

Means and Standard Deviations for Suicide Proportion % and Population Count by Age Group

Age Group	<i>N</i>	Suicide Proportion (%)		Population Count	
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
5-14	6010	.001	.001	1738938	3384599
15-24	6010	.009	.010	1761706	3396100
25-34	6010	.012	.015	1693976	3276127
35-54	6010	.015	.019	2799146	5601422
55-74	6010	.017	.019	1705519	3582171
75+	6010	.025	.032	500960	1213275

Note. N = the number of data entries per country and year for the specified age group.



Conclusion

To answer the question of whether suicide rates differ across age groups, my results reject the null hypothesis that suicide rates do not differ between different age groups and conclude that there are statistically significant differences in the average suicide rates between age groups at an alpha level of $\alpha = 0.01$. This is likely due to the common life circumstances and situations that each life stage goes through, which differ from other life stages. This finding highlights the

importance of quickly identifying life stages with higher average suicide rates and implementing measures for individuals in those life stages who are at higher risk of committing suicide.

For my sub-question of whether suicide rates are the highest among the 55-74 age group, my findings failed to reject the null hypothesis that average suicide rates are not the highest among the 55-74 age group. In fact, the results show statistical evidence that the average suicide rate of the 75+ age group is significantly higher than the average suicide rate of the 55-74 age group at an alpha level of $\alpha = 0.01$. Some reasons for this surprising finding could be chronic illnesses, loneliness from a lack of social connections, and depression from grief, which seem to occur at higher levels in older adults. In addition, studies show that older adults tend to plan suicide more carefully and are therefore more successful in carrying it out (National Council on Aging). These findings reveal how crucial it is to implement suicide intervention and preventative measures for older adults and the importance of quickly implementing policies and practices that provide the support that this age group needs.