Unit 6: United States Department of Transportation | Comparing the Mean Age of Accident Victims in 2020 and 2021

Student's Name: Avinash Bunga

Course: CIS621 Data Analysis for Business Analytics

Professor: Robert Kao

Date: 09.24.2023

Institution: Park University

Analysis of Road Accident Data

Comparing the Mean Age of Accident Victims in 2020 and 2021 **United States Department of Transportation**

1. Research Question

Did the average age of individuals involved in road accidents significantly differ between 2020 and 2021?

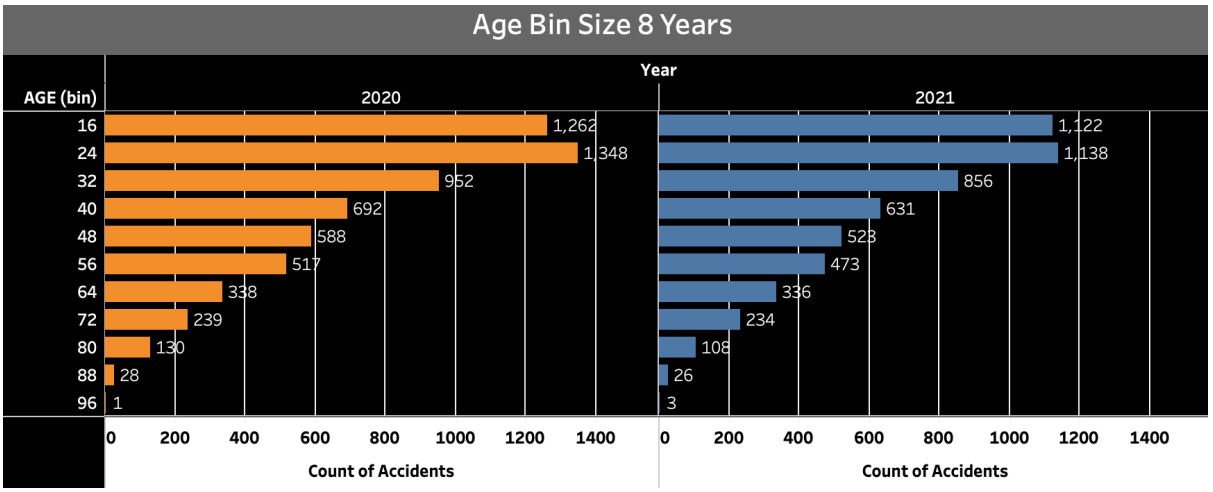
2. Methodology & Tools

T-Test Analysis:

To answer the research question, a two-sample t-test assuming unequal variances was conducted using Excel. This test helps compare the mean values from two independent groups to see if they significantly differ from each other.

Graphical Analysis:

Alongside the t-test, a visual comparison using bar graphs was conducted in Tableau to understand the distribution of ages. This aids in visually capturing any pronounced changes between age groups across the two years.



3. Detailed T-Test Results:

Measure	2020 (Variable 1)	2021 (Variable 2)
Mean	39.57407711	39.98458716
Variance	311.7529216	316.2621959
Observations	6095	5450
Hypothesized Mean Difference	0	
df	11382	
t Stat	-1.242380949	
P(T<=t) one-tail	0.107060819	
t Critical one-tail	1.644987513	
P(T<=t) two-tail	0.214121637	
t Critical two-tail	1.960172429	

4. Findings & Confidence

T-Test Summary: The mean ages between 2020 and 2021 don't have a statistically significant difference. This conclusion is derived from the P-value of 0.214, which exceeds the standard significance level (alpha) of 0.05.

Graphical Insights: Despite similar mean ages, the distribution amongst age groups presents noticeable variations. For instance, while younger age groups (24 to 31) dominated accidents in both years, the age bracket of 40 to 47 saw more accidents in 2021 compared to 2020.

5. Discussion & Potential Bias

To fully appreciate these findings, it's crucial to consider potential biases and factors that might influence the results:

Sample Size Bias: The samples, although large, represent a fraction of the entire population. There could be patterns or anomalies not captured in these samples.

Data Collection Bias: If there are inconsistencies in recording or reporting age data, it could lead to skewed insights.

External Influences: Factors outside the dataset, like changes in traffic regulations, road conditions, or public awareness initiatives, might significantly impact accident rates.

Conclusion

While the analysis indicates minimal changes in the mean age of accident victims between 2020 and 2021, focusing solely on averages could be misleading. The detailed age distributions, as showcased in the bar graphs, hint at shifts in which age groups are more prone to accidents. This disparity between an overarching mean value and individual age group trends underscores the importance of a multifaceted approach to data analysis. Not only should we rely on statistical tests but also explore data visually and be mindful of inherent biases. Only through such a comprehensive examination can we derive nuanced, actionable insights from data.

Reference:

Jain, D. (2023, August 4). *T-Test Formula*. Educba. https://www.educba.com/t-test-formula/

Flerlage, K. (2023, July 26). *Make Your Tableau Bins Dynamic*. Flerlagetwins. https://www.flerlagetwins.com/2023/06/dynamic-bin-size.html

Kim, S. (2022, October 13). *Comparing Datasets Using Statistical Analysis*. Medium. https://medium.com/geekculture/how-to-compare-two-datasets-c0bdc51062c7