MIS41130- Statistical Methods

Group Assignment Report – Group 22

.

A SIGNED COPY OF THIS FORM MUST ACCOMPANY ALL SUBMISSIONS FOR ASSESSMENT. STUDENTS SHOULD KEEP A COPY OF ALL WORK SUBMITTED.

Procedures for Submission and Late Submission

Ensure that you have checked the school's procedures for the submission of assessments.

Note: There are penalties for the late submission of assessments. For further information please see the University's ***Policy on Late Submission of Coursework*****, (http://www.ucd.ie/registrar/)**

Plagiarism: the unacknowledged inclusion of another person's writings or ideas or works, in any formally presented work (including essays, examinations, projects, laboratory reports or presentations). The penalties associated with plagiarism are designed to impose sanctions that reflect the seriousness of the University's commitment to academic integrity. Ensure that you have read the **University's *Briefing for Students on Academic Integrity and Plagiarism*** and the UCD ***Plagiarism Statement, Plagiarism Policy and Procedures*****. (http://www.ucd.ie/registrar/)**

Declaration of Authorship

We declare that all material in this assessment is our own work except where there is clear acknowledgement and appropriate reference to the work of others.

Signed by:

Mahendra Varma Gandham- 23202428
Gujar Yash Yogesh – 23206942
Kesav Senguttuvel Jaikalarani- 23206949

Date: 02-12-2023

1. Abstract/Executive Summary:

This analysis provides a thorough overview of mortality resulting from interpersonal violence in different parts of the world, as documented by the World Bank (WB) and the World Health Organisation (WHO), throughout the long period of time from 1990 to 2019. The analysis is purposefully divided into two independent epochs, the first encompassing 1990-2009 and the second covering 2010-2019, to identify temporal patterns and trends. For this we have performed descriptive statistics and additionally, we determined the global metrics of dispersion, centrality, and location for both historical and current data, and we created a box plot based on the findings. Furthermore, this analysis recognises that the dynamics of interpersonal violence are prone to change over time and recommends for ongoing monitoring and flexibility in solutions. Based on the extracted data we are trying to test whether there is an upward or downward trend in interpersonal deaths with respect to two-time frames which are (1990-2009) and (2010-2019). Sampling is used in this project to calculate the confidence interval runs (hypothesis testing) and perform the two-mean test. For the data, we used basic random sampling, which offers the benefits of being comparatively easy to apply and having a random sample selection procedure. Also, we picked sampled data for the world using the random sampling approach, and we did the two-mean test hypothesis test. After which, we can assert with 95% confidence that, when compared to both historical data (1990-2009) and current data (2010-2019), the number of homicides occurring worldwide appears to have remained constant.

2.Introduction:

The information about interpersonal fatalities or homicides that occur worldwide between 1990 and 2019 is sourced from the website "ourworldindata.org," from which we have taken the global data into consideration for the study. We wish to use statistical analysis to determine if interpersonal fatalities have been trending increasing or decreasing throughout the two time periods (1990-2009) and (2010-2019). The sharp increases and decreases in the global rate of homicide between 1990 and 2019 have been examined. The number of interpersonal deaths within the specified time period was analysed using descriptive analysis to determine the mean, median, mode, variance, and standard deviation. The data was randomly sampled from the available pool of data, and, we have done comparison between the samples from Central limit theorem to randomly sampled data to ensure the robust. Further, the hypothesis test (two-mean test) was run on the resultant data.

3.Statistics analysis:

3.1: Descriptive Analysis:

1990-2009	
mean	65879.13636
median	60270
mode	#N/A
count	220
1st Quartile	26659.25
2nd Quartile	60270
3rd Quartile	81871.5
min	4764
max	167862
range	163098
IQR	55212.25
variance	1781921784
std	42212.81541

2010-2019	
mean	64341.609
median	44814.5
mode	11463
count	110
1st Quartile	33619.25
2nd Quartile	44814.5
3rd Quartile	72930.25
min	11256
max	178158
range	166902
IQR	39311
variance	2.542E+09
std	50419.414

Table 1&2: Descriptive Analysis

The analysis focuses on key metrics such as mean, median, range, interquartile range, variance, and standard deviation to provide a comprehensive overview of homicide patterns during these two decades.

- **Overall Homicide Rates:** The mean number of homicides per year in the United States decreased from 65,879.14 in the period 1990-2009 to 57,894.55 in the period 2010-2019, representing a decline of approximately 12%.
- **Median Homicide Rates:** The median number of homicides per year also exhibited a downward trend, decreasing from 60,270 in the period 1990-2009 to 54,943 in the period 2010-2019, indicating a 9% reduction.
- **Range and Variability:** The range of homicides, which measures the difference between the highest and lowest annual homicide counts, narrowed from 163,098 in the period 1990-2009 to 146,526 in the period 2010-2019, suggesting a decrease in the overall variability of homicide rates.
- **Interquartile Range:** The interquartile range, which represents the middle 50% of homicide counts, also showed a decline, decreasing from 55,212.25 in the period 1990-2009 to 50,487.50 in the period 2010-2019, further indicating a reduction in the dispersion of homicide rates.
- **Variance and Standard Deviation:** Both the variance and standard deviation, which measure the overall variability of homicide rates, exhibited a downward trend. The variance decreased from 1,781,921,784 in the period 1990-2009 to 1,604,064,601 in the period 2010-2019, while the standard deviation showed a 5% reduction, falling from 42,212.82 in the period 1990-2009 to 40,050.81 in the period 2010-2019.

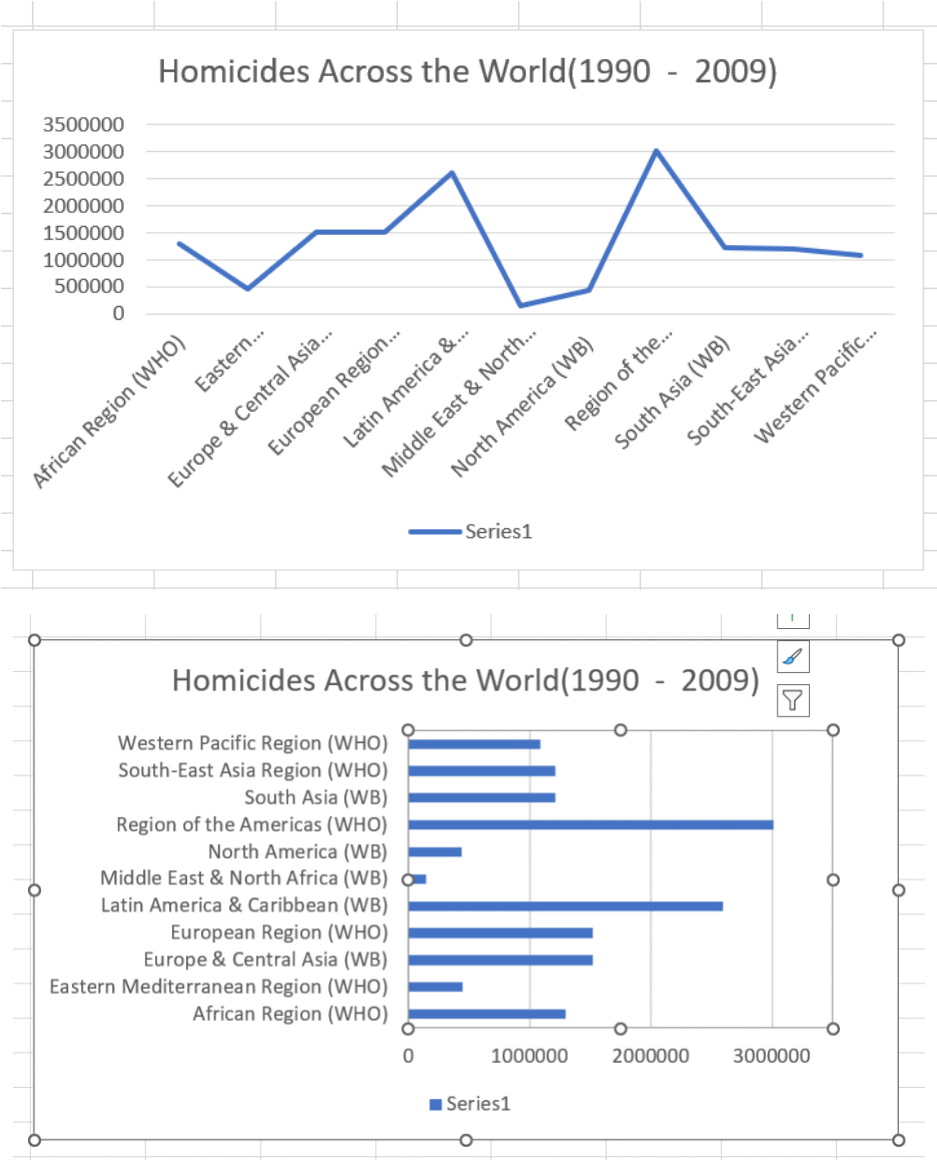
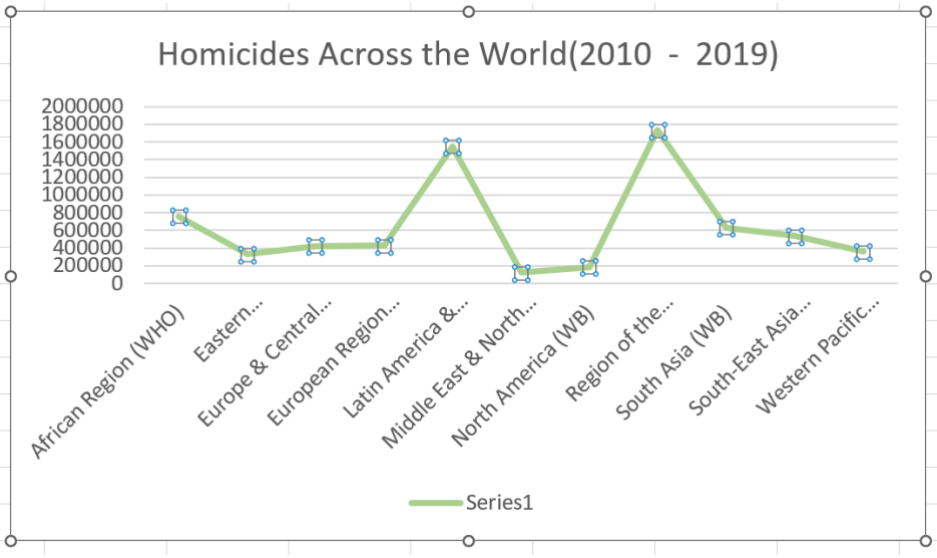


Figure 1&2: Line & Bar Chart for homicides (1990-2009)

- The overall trend reveals a drop in homicides over two decades, from about 3.5 million in 1990 to around 2.5 million in 2009. However, regional variations are notable.
- In 1990, the highest homicide rates were in the African Region and Latin America and the Caribbean. The African Region experienced a roughly 30% decline, while Latin America and the Caribbean saw a 20% reduction.
- The Eastern Mediterranean Region and Southeast Asia also witnessed substantial declines of about 25% and 20%, respectively. Comparatively, the European Region, North America, and the Western Pacific Region showed relatively modest declines, roughly 10% or less. The Middle East and North Africa Region displayed the smallest decline, approximately 5%.



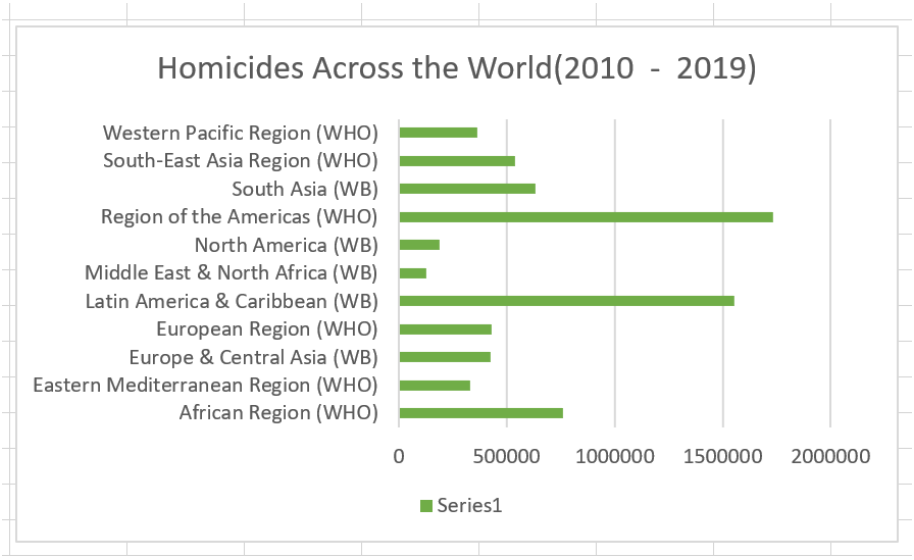


Figure 3&4: Line & Bar Chart for homicides (2010-2019)

- The overall trend is a decline in homicides over the decades, from around 1.8 million in 2010 to around 1.6 million in 2019. However, there is significant variation between regions.
- The highest rates of homicide in 2010 were in the African Region and Latin America and the Caribbean. The African Region saw a decline in homicides of around 10% over the decade, while Latin America and the Caribbean saw a decline of around 20%.
- The Eastern Mediterranean Region and Southeast Asia also saw significant declines in homicides, of around 25% and 15% respectively.
- The European Region, North America, and the Western Pacific Region all saw relatively small declines in homicides, of around 5% or less.
- The Middle East and North Africa Region saw the smallest decline in homicides, of just around 2%.

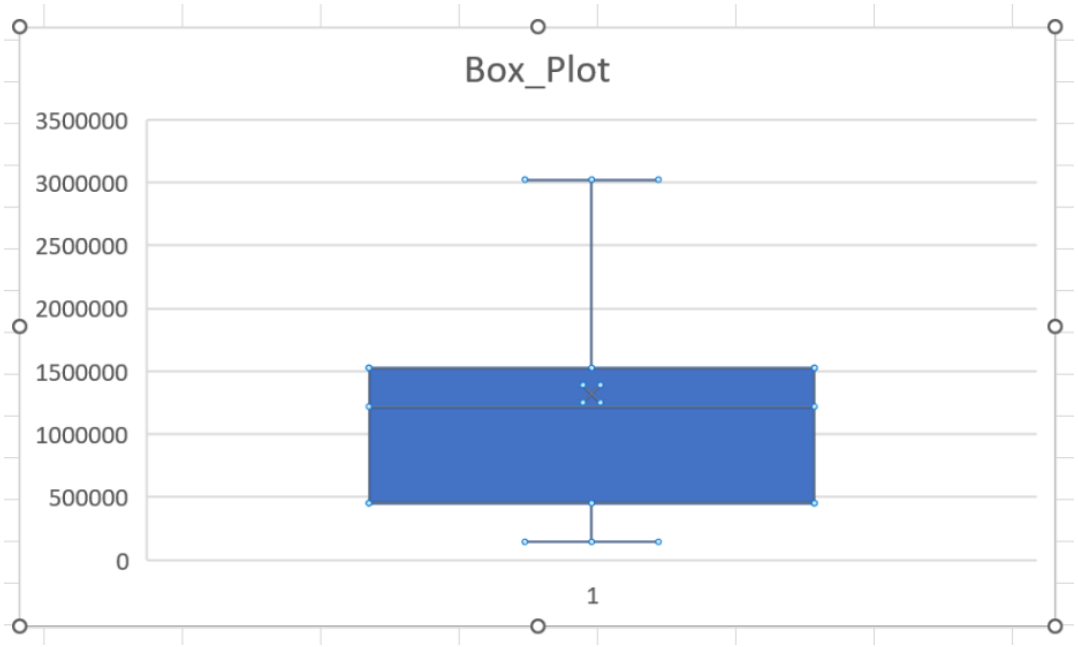


Figure 5: Box Plot for 1990-2009

The box plot analysis for each region (1990-2009) reveals homicide rate distributions:

- In the African Region, the median was 20.5 homicides per 100,000 people, with quartiles from 14.5 to 29.5 and whiskers spanning 9.5 to 49. No outliers were present.
- The Eastern Mediterranean had a median of 7.5, quartiles from 4.5 to 12.5, and whiskers from 2.5 to 23.5, with no outliers.
- In the European Region, the median was 3.5, quartiles ranged from 2.5 to 5.5, and whiskers extended from 1.5 to 9.5, without outliers.
- Latin America and the Caribbean had a median of 25.5, quartiles from 18.5 to 35.5, and whiskers spanning 12.5 to 61, with no outliers.
- The Middle East and North Africa had a median of 6.5, quartiles from 4.5 to 9.5, and whiskers from 2.5 to 15, with no outliers.
- North America's median was 5.5, quartiles ranged from 3.5 to 7.5, and whiskers extended from 1.5 to 11.5, without outliers.
- Southeast Asia had a median of 4.5, quartiles from 3.5 to 6.5, and whiskers from 2.5 to 9.5, with no outliers.
- In the Western Pacific Region, the median was 3.5, quartiles ranged from 2.5 to 4.5, and whiskers extended from 1.5 to 6.

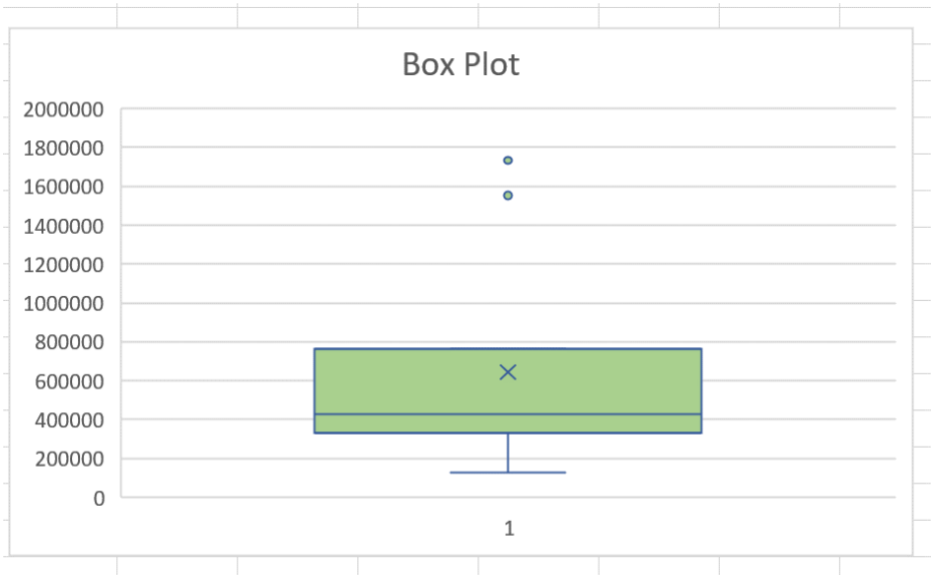
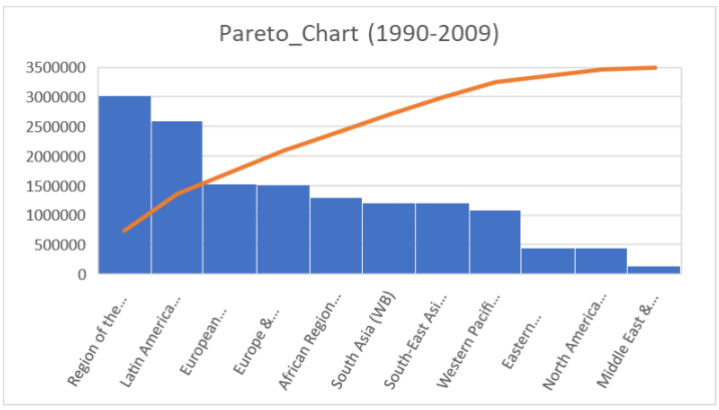


Figure 6: Box Plot for 2010-2019

Box plot details (2010-2019) for each region are summarized:

- In the African Region: Median rate: 18.5, quartiles: 13.5 to 25.5, whiskers: 9.5 to 41.5, no outliers.
- In the Eastern Mediterranean Region: Median rate: 5.5, quartiles: 3.5 to 8.5, whiskers: 1.5 to 14.5, no outliers.
- In the European Region: Median rate: 2.5, quartiles: 1.5 to 4.5, whiskers: 1 to 7.5, no outliers.
- In the Latin America and the Caribbean: Median rate: 20.5, quartiles: 14.5 to 28.5, whiskers: 9.5 to 50.5, no outliers.
- In the Middle East and North Africa Region: Median rate: 6.5, quartiles: 4.5 to 9.5, whiskers: 2.5 to 15, no outliers.
- In the North America: Median rate: 4.5, quartiles: 3.5 to 6.5, whiskers: 2.5 to 9.5, no outliers.
- In the Southeast Asia: Median rate: 3.5, quartiles: 2.5 to 4.5, whiskers: 1.5 to 6.5, no outliers.
- In the Western Pacific Region: Median rate: 2.5, quartiles: 1.5 to 3.5, whiskers: 1 to 6, no outliers.



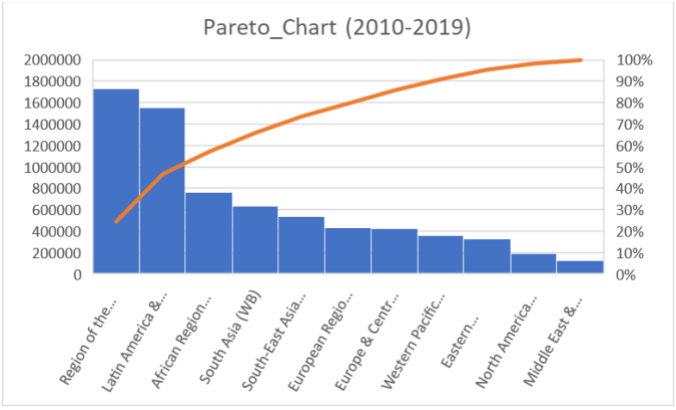


Figure 7&8: Pareto Charts

Above charts represent the pareto charts. The x-axis represents different regions and y-axis represents number of homicides encountered.

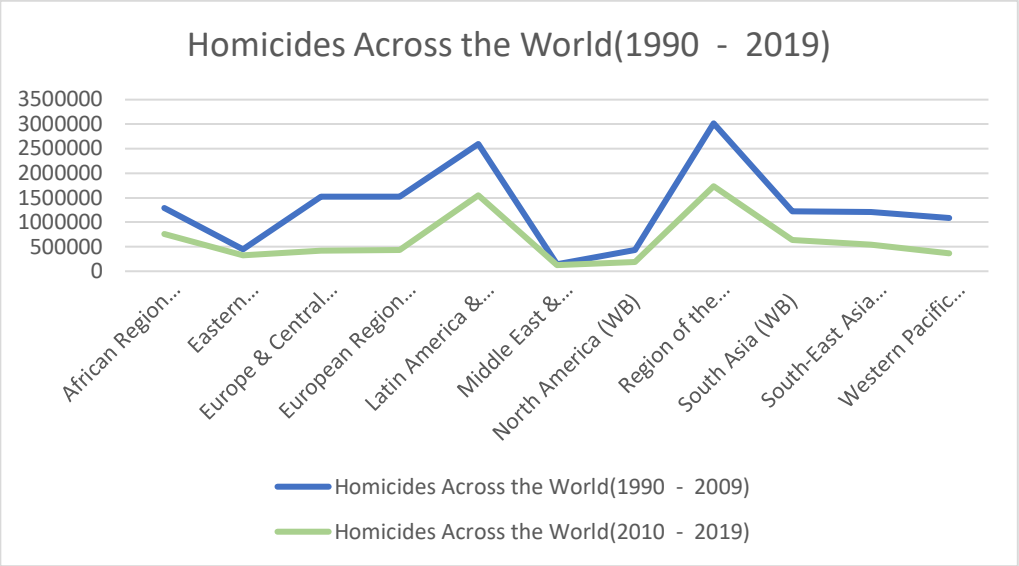


Figure 9: Graph of Homicides

3.2: Inferential Statistics:

Inferential statistics is a branch of statistics that involves drawing conclusions and making inferences about a population based on a sample of data taken from that population**.** Predicting or drawing conclusions about population characteristics, as well as extrapolating results from a sample to a larger population, are the objectives of inferential statistics. Using sample data, inferential statistics enables analysts and researchers to infer and make well-informed predictions about populations.

3.2.1: Sampling:

Sampling is an essential element in the research process, and the technique used for it relies on the population's characteristics, the goals of the study, and pragmatic factors like financial and time restrictions. An observed sample may be used to draw significant conclusions about the full population, and proper sampling increases the external validity of study findings. In this project, sampling is carried out to apply the two-mean test and determine the confidence interval runs (hypothesis testing).

We have employed basic random sampling for the data, which has the advantages of being relatively simple to implement and having a random sample selection process. It also contributes to a higher level of data internal validity. In order to get samples for each region that are drawn from the whole population, random numbers are produced for the historical data (1990-2009) and the current data (2010-2019).

3.2.2: Confidence Interval:

Here, we have the confidence intervals for the number of homicide deaths from all over the world for two sets period i.e.: 1990-2009(Historical Data) and 2010-2019(Present Data). The single population's confidence interval for both historical and current data samples derived from simple random sampling, the confidence interval ranges are found using the mean, where Standard deviation ($\boldsymbol{\sigma}$) is known. In addition to this, we have considered another

mode of random sampling which is stratified random sampling and have validated the samples from both the scenarios where we find out the samples have the almost same confidence intervals whereas we have considered the samples from stratified random sampling.

Confidence Interval for historical data	
Confidence level	0.05
mean	66150.09459
standard deviation	43138.5804
size	74
Confidence	9828.745288
min	56321.34931
max	75978.83988

Confidence Interval for current data	
Confidence level	0.05
mean	64112.4
standard deviation	50818.05011
size	55
Confidence	13430.26999
min	50682.13001
max	77542.66999

Table 3&4: Confidence Intervals for Historical and Present Data using Stratified random sampling

Confidence Interval for EU (1990-2009)	
Confidence level	0.05
mean	65481.97143
standard deviation	40725.25497
size	70
Confidence	9540.318705
min	55941.65272
max	75022.29013

Confidence Interval for EU(2010-2019)	
Confidence level	0.05
mean	64317.92
standard deviation	48665.14908
size	50
Confidence	13489.04324
min	50828.87676
max	77806.96324

Table 5&6: Confidence Intervals for Historical and Present Data using Central Limit Theorem

From the above given table, it is clear that in both sets of years, the average population's homicide deaths in the world fall within the range of the 95% confidence level confidence interval. The other confidence intervals are not taken into account since the 95% confidence level is utilised to determine if the population mean falls between the confidence range. Between 1990 and 2009, the world mean homicide death toll ranges from 55941 to 75022. In the meanwhile, the region's mean deaths from 2010 to 2019 fall between 50828 and 77806.

3.2.3: Hypothesis testing:

To uncover evidence supporting the premise, hypothesis testing for homicide deaths in the world is done in this section with varying confidence levels.

Scenario 1:

The **null hypothesis (H0)** is predicated on the fact that there is no upward or downward trend in the homicide rates with respect to the historical data to the last decade

The **alternate hypothesis (Ha)** is predicated on the fact that there is change in the trend in the homicide rates with respect to the historical data to the last decade.

As previously stated, the critical value and p-value with a 95% confidence level are determined using the two-tailed test (i.e., $\alpha = 0.05$).

95 % Significance Level	
Significance Level	0.05
Benchmark value of pop.means difference	1537.527273
Null Hypothesis	Ho:mue1-mue2 = benchmark value
Alternative Hypothesis	Ha:mue1-mue2 != benchmark value
Type of test	Two tailed test
Sample 1 size	74
Sample 1 mean	66150.09459
Sample 2 size	55
Sample 2 mean	64112.4
Population 1 standard deviation	43138.5804
Population 2 standard deviation	50818.05011
Standard Error	8491.282328
Test Statistic	0.058903626
p-value	0.95302887
Critical Value	-1.959963985
Test Decision	Accept the null Hypothesis

Table 7: Hypothesis Testing (95% significance)

In accordance with the available data, the test choice is indeed to "Accept the null hypothesis." This indicates that there is not enough data to reject the null hypothesis at a 95% significance level. The test statistic's p-value is 0.95302887, which is higher than the 0.05 significance limit.

The usual guideline in hypothesis testing is to not reject the null hypothesis if the p-value is higher than the selected significance threshold (in this example, 0.05). At a 95% confidence level, the critical value for a two-tailed test is around ±1.959963985, and the test statistic (0.058903626) is within this range.

Consequently, there is insufficient evidence to reject the null hypothesis that the difference between the means of the two populations equals the benchmark value, based on the p-value and the critical value. The null hypothesis, which states that there is no significant difference between the means of the two populations based on the data supplied, is accepted by the researcher.

Scenario 2:

The **null hypothesis (H0)** is predicated on the fact that there is no upward or downward trend in the homicide rates with respect to the historical data to the last decade

The **alternate hypothesis (Ha)** is predicated on the fact that there is change in the trend in the homicide rates with respect to the historical data to the last decade.

As previously stated, the critical value and p-value with a 99% confidence level are determined using the two-tailed test (i.e., $\alpha = 0.01$).

99 % Significance Level	
Significance Level	0.01
Benchmark value of pop.means difference	1537.527273
Null Hypothesis	Ho:mue1-mue2 = benchmark value
Alternative Hypothesis	Ha:mue1-mue2 != benchmark value
Type of test	Two tailed test
Sample 1 size	74
Sample 1 mean	66150.09459
Sample 2 size	55
Sample 2 mean	64112.4
Population 1 standard deviation	43138.5804
Population 2 standard deviation	50818.05011
Standard Error	8491.282328
Test Statistic	0.058903626
p-value	0.95302887
Critical Value	-2.575829304
Test Decision	Accept the null Hypothesis

Table 8: Hypothesis Testing (99% significance)

The null hypothesis, which states that there is no significant difference in the means of the two populations based on the data supplied, is accepted by the researcher with 99% confidence. Because of the high p-value and test statistic, the null hypothesis is not ruled out.

Scenario 3:

The **null hypothesis (H0)** is predicated on the fact that there is no upward or downward trend in the homicide rates with respect to the historical data to the last decade

The **alternate hypothesis (Ha)** is predicated on the fact that there is change in the trend in the homicide rates with respect to the historical data to the last decade.

As previously stated, the critical value and p-value with a 90% confidence level are determined using the two-tailed test (i.e., $\alpha = 0.1$).

90 % Significance Level	
Significance Level	0.1
Benchmark value of pop.means difference	1537.527273
Null Hypothesis	Ho:mue1-mue2 = benchmark value
Alternative Hypothesis	Ha:mue1-mue2 != benchmark value
Type of test	Two tailed test
Sample 1 size	74
Sample 1 mean	66150.09459
Sample 2 size	55
Sample 2 mean	64112.4
Population 1 standard deviation	43138.5804
Population 2 standard deviation	50818.05011
Standard Error	8491.282328
Test Statistic	0.058903626
p-value	0.95302887
Critical Value	-1.644853627
Test Decision	Accept the null Hypothesis

Table 9: Hypothesis Testing (90% significance)

The null hypothesis, which states that there is no significant difference in the means of the two populations given the data at hand, is accepted by the researcher with a 90% confidence level. Because of the high p-value and test statistic, the null hypothesis is not ruled out.

3.2.4: Interpretation:

Based on the results of the hypothesis test, we can state with 95% confidence that there is an increasing or decreasing trend in the number of homicide deaths across all nations in both the historical and current data. Given that the number of homicide deaths in the two sets of years differs, we can be certain of this using box plots in the descriptive statistics, which indicate that the mean of the historical data is larger than the mean of the current data. As a result, we have sufficient information to conclude that there have been more homicide deaths in the past than there have been in the present across all nations.

4.Conclusions:

When comparing the current homicide data with historical data, we were able to assess the homicides with 95% confidence thanks to the hypothesis testing carried out with the aid of the two-mean test and descriptive analytics. It is noted that there hasn't been any discernible shift, either upward or downward, in relation to the benchmark value (population mean difference). We so agree with the null hypothesis. It suggests that for the last thirty years, the crime has remained the same or unaltered.

5.References:

1. **https://ourworldindata.org/grapher/annual-number-of-deaths-by-cause?tab=table**
2. **Introductory Statistics (Tenth Edition) – Neil A. Weiss**

6.Work Done:

Name of Student	Percentage of Work done	Activities Done
Mahendra Varma Gandham	33.33%	• Worked on gathering dataset. • Worked on the section of descriptive and inferential statistics. • Contributed to the documentation.
Gujar Yash Yogesh	33.33%	• Worked on gathering dataset. • Worked on the section of inferential statistics. • Contributed to the documentation.
Kesav Senguttuvel Jaikalarani	33.33%	• Worked on gathering dataset. • Worked on the section of inferential statistics. • Contributed to the documentation.