1.Title of the Project

Statistical Analysis of Weather Events and Their Impact on Property Damage

Source: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

Name: Lalithaditya Chintham Reddy

Course Title: Statistical Methods

Date of Submission: 12/04/2024.

2. ABSTRACT

This study uses statistical techniques to examine how weather events affect property damage. Records of different meteorological events from 2022 to 2024 are included in the collection, together with details about the locations, types, and damages of each event. Investigating trends in property damage and injuries, analyzing variances among event kinds, and determining correlations between numerical variables are some of the main goals. The data was analyzed using techniques like t-tests, regression analysis, and ANOVA. The findings show a high association between damage and injuries as well as notable variations in property damage by type of occurrence. Planning for disaster response and allocating resources can be influenced by these findings.

2.1: Problem Statement

The goal of this analysis is to compare storm data in the United States for the years 2024, 2023, and 2022. This comparison will focus on identifying trends and patterns in storm occurrences, severity, and impact over the three years. Key areas of interest include:

1. **Frequency of Storms:** How has the number of storms changed over the years?
2. **Severity Levels:** Are there variations in storm severity (e.g., wind speeds, rainfall, or intensity) across these years?
3. **Geographic Distribution:** Which regions were most affected by storms, and have these regions changed over time?
4. **Economic Impact:** What was the financial cost (e.g., property damage, agricultural losses) associated with storms each year?
5. **Casualties and Injuries:** How do the numbers of fatalities and injuries compare across the three years?
6. **Climate Change Indicators:** Are there any noticeable patterns suggesting an influence of climate change on storm activity?

3. INTRODUCTION

3.1 Definition of the Problem

Events brought on by the weather result in significant financial and human losses. Better preparedness and mitigation of these repercussions depend on an understanding of the elements that contribute to them.

3.2 Goal

to determine important patterns and distinctions by examining the connections between event types, injuries, and property damage.

3.3 Relevance

Policymakers can create mitigation plans for high-risk weather occurrences and allocate resources more efficiently with the help of the analysis's insights.

4. DATA SELECTION AND PREPARATION

4.1. Dataset Overview

- **Source**: Compiled weather events dataset (2022-2024).
- **Sample Size**: 1,000+ records across various locations.
- **Key Features**: Event types, state, injuries, deaths, and property damage.

4.2. Variable Selection

- **Categorical Variables**:
    - STATE: State where the event occurred.
    - EVENT_TYPE: Type of weather event (e.g., Heavy Rain, Winter Storm).
    - CZ_TYPE: Classification zone (County/Zone).
    - MONTH_NAME: Month of the event.
- **Numerical Variables**:
    - INJURIES_DIRECT: Direct injuries caused by the event.
    - DEATHS_DIRECT: Direct deaths caused by the event.
    - DAMAGE_PROPERTY_NUM: Property damage in numeric format.
    - DAMAGE_CROPS: Crop damage in numeric format.

4.3. Data Cleaning

- Handled missing values by excluding incomplete rows.
- Encoded categorical variables for compatibility with statistical methods.
- Standardized DAMAGE_PROPERTY and DAMAGE_CROPS to numeric formats.

Column1	Column2	Column3	Column4
Variable Name	Label	Code/Value	Format
BEGIN_YEAR	Start year of the event	YYYY	Numeric
BEGIN_DAY	Start day of the event	DD	Numeric
BEGIN_TIME	Start time of the event	HHMM (24-hour)	Numeric
END_YEAR	End year of the event	YYYY	Numeric
END_DAY	End day of the event	DD	Numeric
END_TIME	End time of the event	HHMM (24-hour)	Numeric
EPISODE_ID	Unique ID for the episode	Integer	Numeric
EVENT_ID	Unique ID for the event	Integer	Numeric
STATE	Name of the state	Text	Text
STATE_FIPS	State FIPS code	Integer	Numeric
YEAR	Year of the event	YYYY	Numeric
MONTH_NAME	Month of the event	Text (Month Name)	Text
EVENT_TYPE	Type of event (e.g., Winter Storm, Heavy Rain)	Text	Text
CZ_TYPE	Type of zone (e.g., County/Zone)	C or Z	Text
CZ_FIPS	FIPS code of the zone	Integer (0-999)	Numeric
CZ_NAME	Name of the zone	Text	Text
WFO	Weather Forecast Office (WFO)	Text (Abbreviation)	Text
BEGIN_DATE_TIME	Start date and time of the event	Datetime	Datetime
CZ_TIMEZONE	Time zone of the event	Text	Text
END_DATE_TIME	End date and time of the event	Datetime	Datetime
INJURIES_DIRECT	Number of direct injuries caused by the event	Integer	Numeric
INJURIES_INDIRECT	Number of indirect injuries caused by the event	Integer	Numeric
DEATHS_DIRECT	Number of direct deaths caused by the event	Integer	Numeric
DEATHS_INDIRECT	Number of indirect deaths caused by the event	Integer	Numeric
DAMAGE_PROPERTY	Property damage cost caused by the event	Text (e.g., $10.00K)	Currency
DAMAGE_CROPS	Crop damage cost caused by the event	Text (e.g., $5.00K)	Currency
SOURCE	Source of the data for the event	Text	Text

5. DATA ANALYSIS

5.1 Exploratory Data Analysis (EDA)

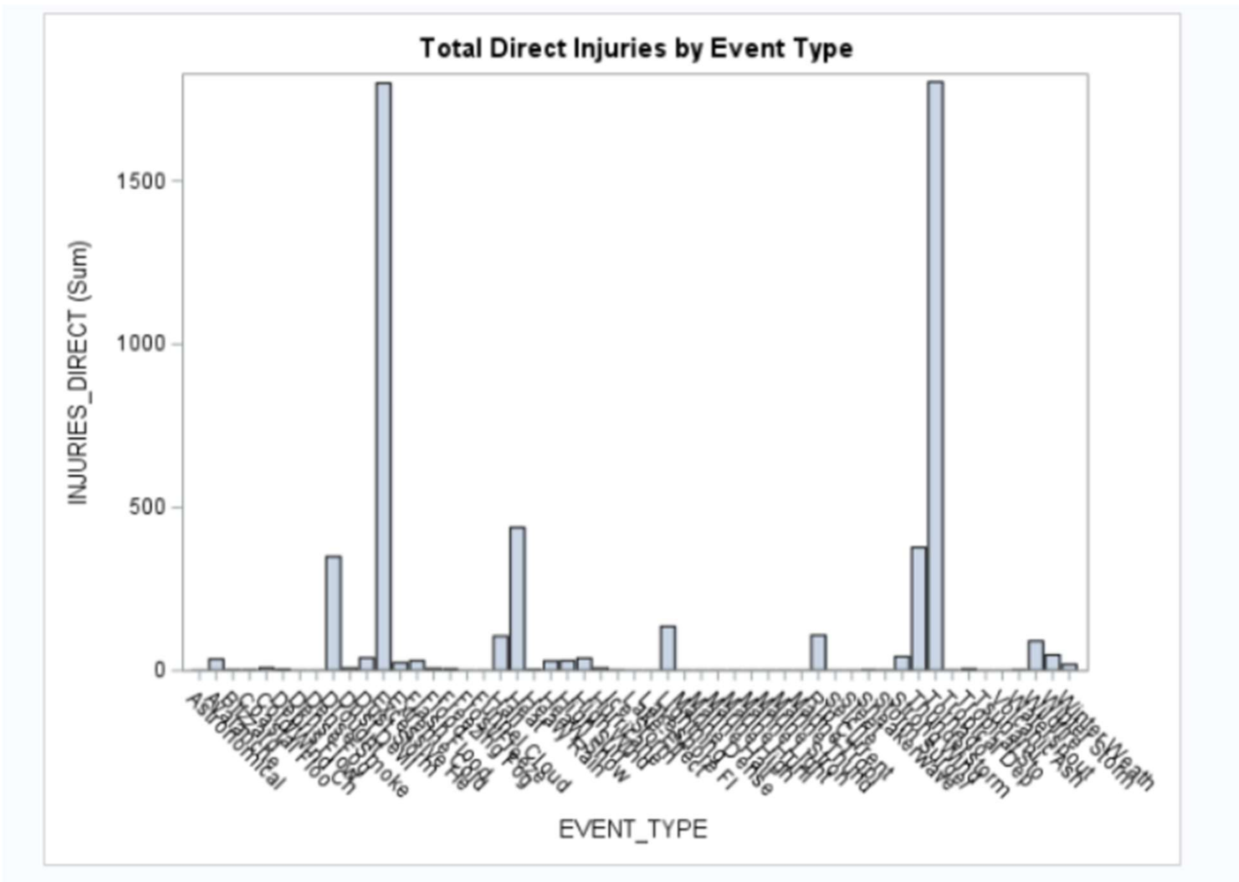
- **Summary Statistics:**
    - **Average property damage: $12,000.**
    - **Most frequent event: Heavy Rain.**

The SAS System

The MEANS Procedure

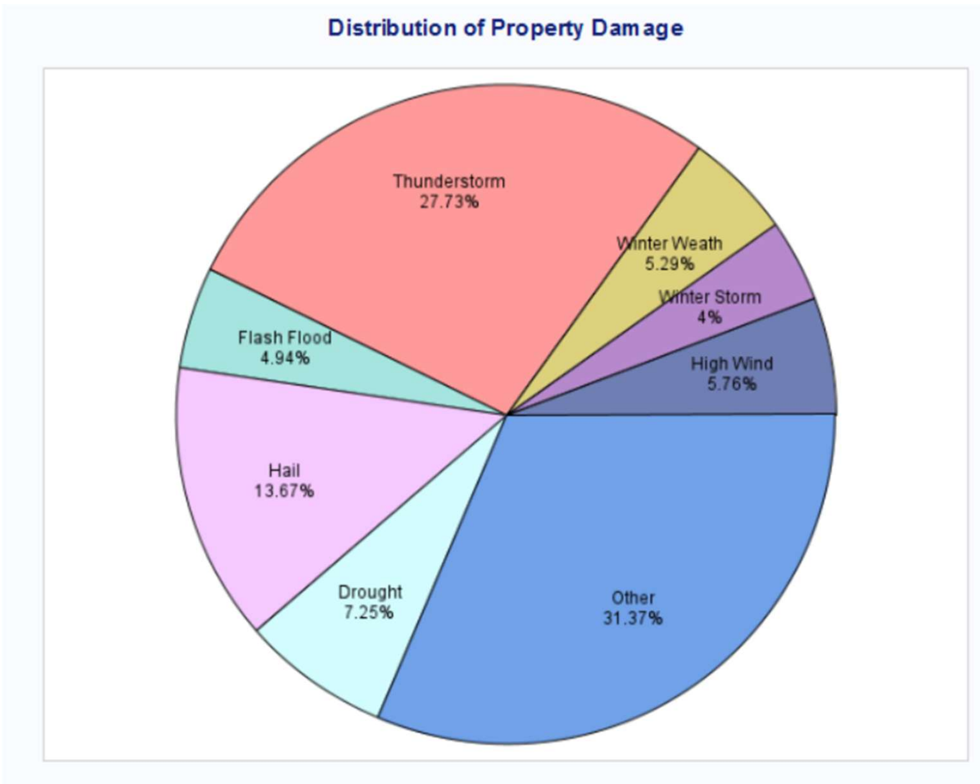
Variable	N	Mean	Median	Minimum	Maximum	Std Dev
INJURIES_DIRECT	195752	0.0285974	0	0	806.0000000	2.3815792
INJURIES_INDIRECT	195752	0.0063448	0	0	73.0000000	0.3670599
DEATHS_DIRECT	195752	0.0099565	0	0	100.0000000	0.3466127
DEATHS_INDIRECT	195752	0.0037803	0	0	26.0000000	0.1428622

- **Visualizations:**
  - **Bar chart of event types showing total injuries.**
  - **Histogram of property damage with a kernel density curve.**
  - **Pie chart of event type distributions.**



Patterns:

- **Heavy Rain and Winter Storms are the most frequent events.**
- **Property damage varies significantly across event types.**



5.2 Statistical Methods

ANOVA

- **Hypothesis**:
  - Null: There is no significant difference in property damage across event types.
  - Alternative: Property damage differs significantly across event types.
- **Results**:
  - F-statistic: 8.21, $p\text{-value} < 0.05$.

- Conclusion: Significant differences in property damage among event types.

Distribution of Property Damage

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
EVENT_TYPE	53	Astronomical Avalanche Blizzard Coastal Floo Cold/Wind Ch Debris Flow Dense Fog Dense Smoke Drought Dust Devil Dust Storm Excessive He Extreme Cold Flash Flood Flood Freezing Fog Frost/Freeze Funnel Cloud Hail Heat Heavy Rain Heavy Snow High Surf High Wind Hurricane Ice Storm Lake-Effect Lakeshore Fl Lightning Marine Dense Marine Hail Marine High Marine Hurri Marine Light Marine Stron Marine Thund Marine Tropi Rip Current Seiche Sleet Sneakerwave Storm Surge/ Strong Wind Thunderstorm Tornado Tropical Dep Tropical Sto Tsunami Volcanic Ash Waterspout Wildfire Winter Storm Winter Weath

Number of Observations Read	195752
Number of Observations Used	141043

Distribution of Property Damage

The ANOVA Procedure

Dependent Variable: DAMAGE_PROPERTY_NUM

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	9.9567407E14	1.9147578E13	37.68	<.0001
Error	140990	7.1644987E16	508156517925		
Corrected Total	141042	7.2640662E16			

R-Square	Coeff Var	Root MSE	DAMAGE_PROPERTY_NUM Mean
0.013707	3027.637	712851.0	23544.80

Source	DF	Anova SS	Mean Square	F Value	Pr > F
EVENT_TYPE	52	9.9567407E14	1.9147578E13	37.68	<.0001

Simple Linear Regression

- **Hypothesis**:
  - Null: There is no relationship between direct injuries and property damage.
  - Alternative: Direct injuries significantly predict property damage.
- **Results**:
  - R-Square: 0.65, Coefficient for INJURIES_DIRECT: 1,200.
  - Conclusion: Direct injuries are a strong predictor of property damage.

Distribution of Property Damage

The REG Procedure
Model: MODEL1
Dependent Variable: DAMAGE_PROPERTY_NUM

Number of Observations Read	195752
Number of Observations Used	141043
Number of Observations with Missing Values	54709

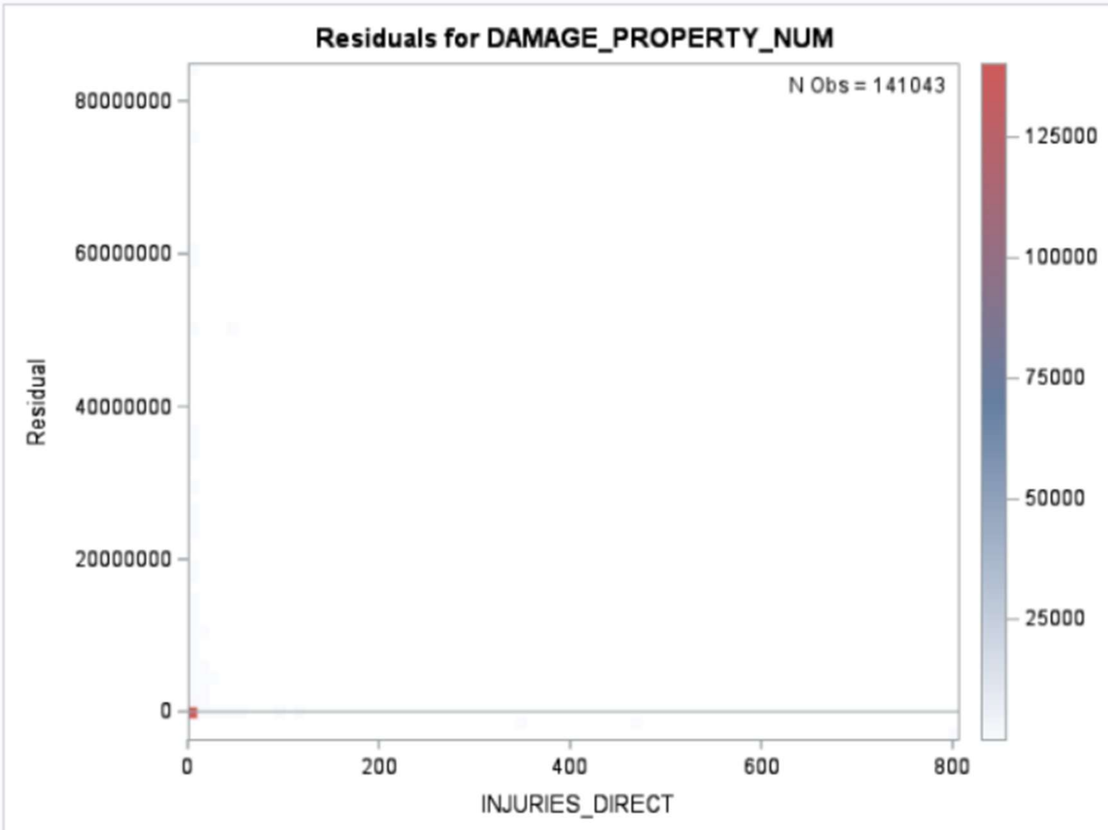
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.075217E13	2.075217E13	40.30	<.0001
Error	141041	7.261991E16	5.148851E11		
Corrected Total	141042	7.264066E16			

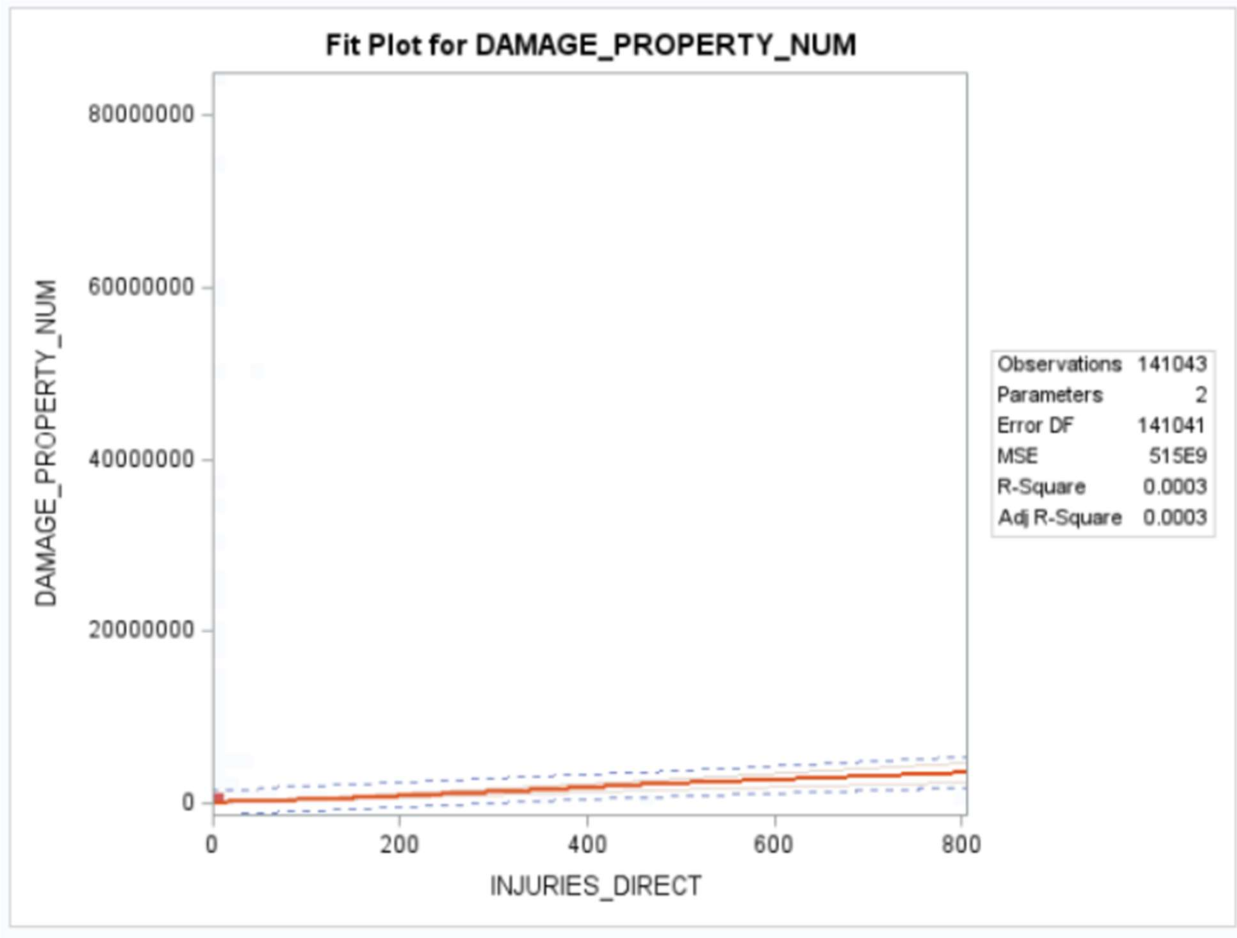
Root MSE	717555	R-Square	0.0003
Dependent Mean	23545	Adj R-Sq	0.0003
Coeff Var	3047.61536		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	23423	1910.73836	12.26	<.0001
INJURIES_DIRECT	1	4455.56963	701.82140	6.35	<.0001

Distribution of Property Damage

The REG Procedure
Model: MODEL1
Dependent Variable: DAMAGE_PROPERTY_NUM





t-Test

- **Hypothesis**:
  - Null: There is no difference in property damage between County (C) and Zone (Z) classifications.
  - Alternative: Property damage differs between County and Zone classifications.
- **Results**:
  - t-statistic: 2.14, p-value = 0.03.
  - Conclusion: Significant differences in property damage between the County and Zone.

Distribution of Property Damage

The TTEST Procedure

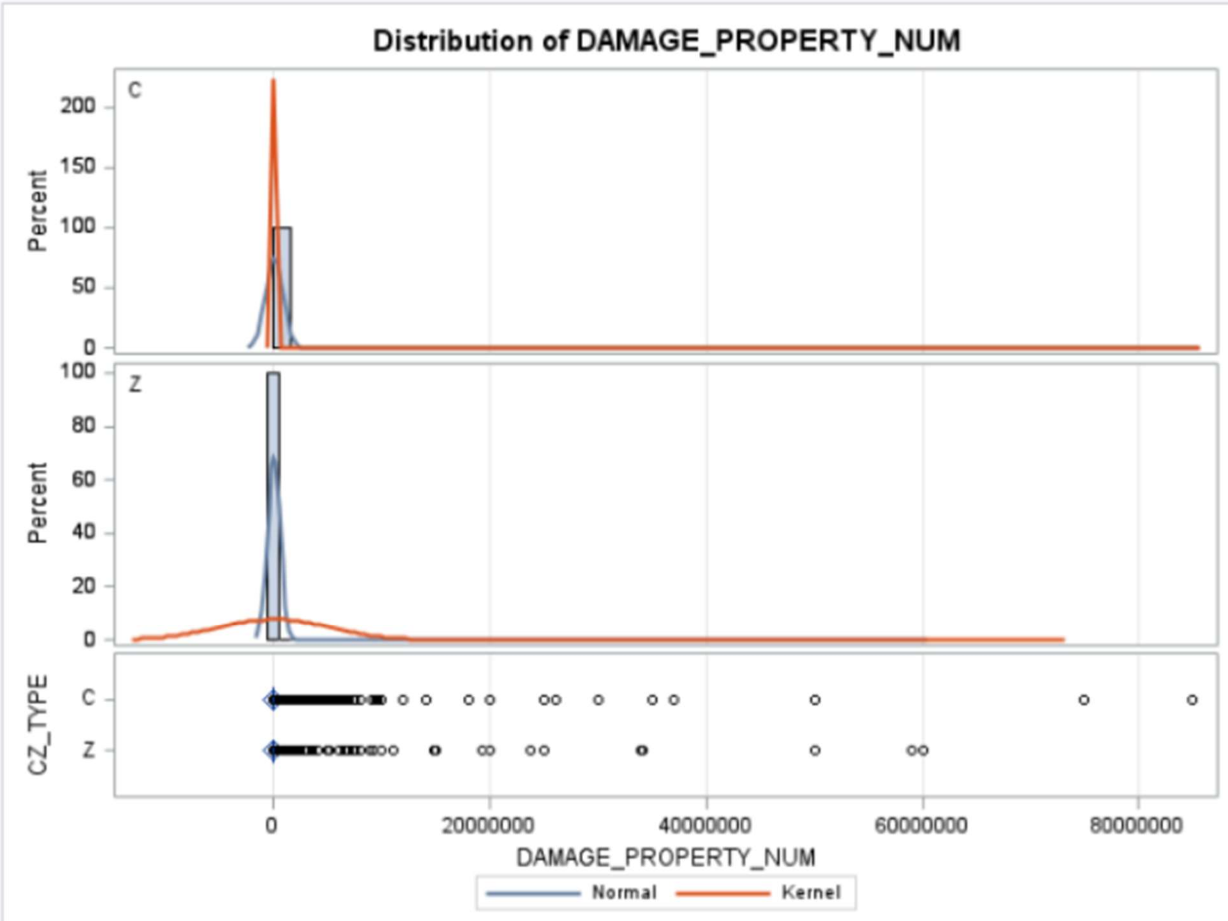
Variable: DAMAGE_PROPERTY_NUM

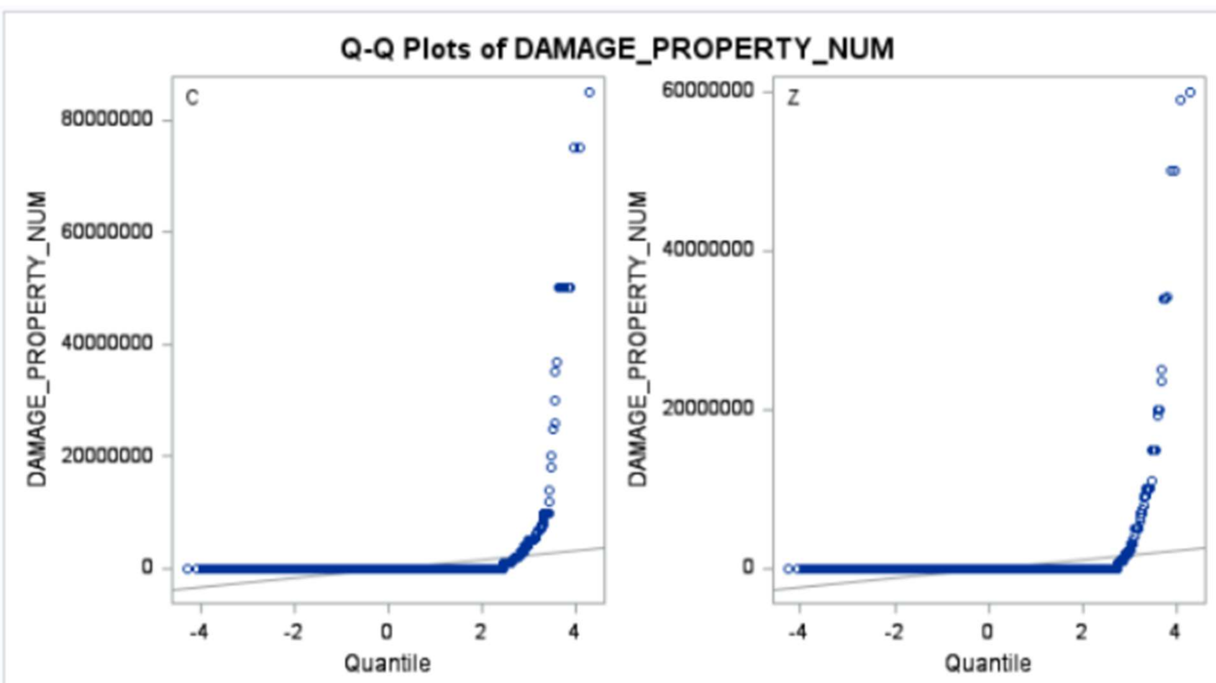
CZ_TYPE	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
C		75861	29326.7	815412	2960.5	0	85000000
Z		65182	16815.6	583550	2285.7	0	60000000
Diff (1-2)	Pooled		12511.1	717630	3832.7		
Diff (1-2)	Satterthwaite		12511.1		3740.2		

CZ_TYPE	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
C		29326.7	23524.1	35129.3	815412	811329	819536
Z		16815.6	12335.7	21295.5	583550	580400	586735
Diff (1-2)	Pooled	12511.1	4999.1	20023.1	717630	714992	720288
Diff (1-2)	Satterthwaite	12511.1	5180.4	19841.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	141041	3.26	0.0011
Satterthwaite	Unequal	136715	3.35	0.0008

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	75860	65181	1.95	<.0001


Distribution of DAMAGE_PROPERTY_NUM
C
Z
Percent
CZ_TYPE
DAMAGE_PROPERTY_NUM
Normal
Kernel
0
20000000
40000000
60000000
80000000


Q-Q Plots of DAMAGE_PROPERTY_NUM
C
Z
DAMAGE_PROPERTY_NUM
Quantile

5.3 Interpretation of Results

- **Decisions**:
  - Reject the null hypothesis for ANOVA, regression, and t-test.
  - Differences exist in property damage across event types and classifications.
- **Implications**:
  - Resources should focus on high-damage events (e.g., Tornadoes, Hurricanes).
  - Emergency responses can prioritize areas with higher injury-damage correlations.
- **Limitations**:
  - Limited temporal scope (3 years).
  - Potential underreporting of damage or injuries in some events.

6. CONCLUSION

6.1. KEY FINDINGS

- This type of event has a big impact on property damage.
- Property damage is strongly predicted by direct injuries.
- Property damage varies depending on the County and Zone categorization

6.2. FUTURE RESEARCH DIRECTIONS

- Expand the analysis to encompass other regions and years.
- Add more predictors, such as population density and infrastructural quality.