Proposal: The Real Reason For Higher Insurance Rates

DATA 450 Capstone

[Samantha Miller]

February 19, 2025

1 Introduction

Insurance is a price that we all need to pay. Many people have noticed that property insurance prices, both for personal home and commercial businesses have continuously increased in the past years. This has become a topic of conversation for many and what people usually chalk it up to is inflation. However, there are multiple different factors that are likely contributing to the rise in insurance premiums. One significant possibility for the influx in prices is the increase in natural disasters and the damages they are causing. Insurance companies in areas most affected by natural disasters are then forced to raise prices in order to combat for losses and claims that occur in response to the events. So are property insurance prices increasing due to inflation or actually because of the higher risk of natural disasters?

2 Dataset

Inflation Rate Dataset

World Bank, Inflation, consumer prices for the United States [FPCPITOTLZGUSA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FPCPITOTLZGUSA, February 6, 2025.

Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.

- observation_date: yyyy-mm-dd
- FPCPITOTLZGUSA: Inflation, consumer prices (annual %) - United States

Florida Hurricane Dataset

Dataset was obtained from NOAA- National Centers for Environmental Information. Contains 269 Hurricane events reported in Florida between 01/01/2000 and 12/31/2024.

- County/Zone
- St: state
- Date
- Time
- T.Z: timezone
- Type: type of event- Hurricane (typhoon)
- Mag: Magnitude- not applicable to hurricanes
- Dth: Death count
- Inj: Injuries reported
- PrD: Property Damage estimate
- CrD: Crop Damage Estimate

New York Ice Storm Dataset

Dataset was obtained from NOAA- National Centers for Environmentl Information. Contains 270 Ice Storm events reported in New York between 01/01/2000 and 12/31/2024.

- County/Zone
- St: state
- Date
- Time
- T.Z: timezone
- Type: type of event- Ice Storm
- Mag: Magnitude- not applicable to ice storms
- Dth: Death count
- Inj: Injuries reported
- PrD: Property Damage estimate
- CrD: Crop Damage Estimate

Premiums for Homeowner's Insurance Dataset

U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Premiums for Property and Casualty Insurance: Premiums for Homeowner's Insurance [PCU9241269241262], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PCU9241269241262, February 6, 2025.

- observation_date: yyyy-mm-dd
- PCU9241269241262: Producer Price Index (PPI) for Premiums on Property and Casualty Insurance

Premiums for Commercial Insurance Dataset

U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Premiums for Property and Casualty Insurance: Premiums for Commercial Multiple Peril Insurance [PCU9241269241265], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PCU9241269241265, February 6, 2025.

- observation_date: yyyy-mm-dd
- PCU9241269241265: Producer Price Index (PPI) for Premiums on Property and Casualty Insurance

3 Data Acquisition and Processing

- Recode variables in multiple datasets that have the same variable names
- Impute missing values by deleting the row, as there is plenty of data that the value will not likely affect results
- Merge the datasets into a single dataset that combines the insurance rates, inflation rates, and natural disaster data by year
- Overall, the data is pretty tidy but will be thoroughly examined

4 Research Questions and Methodology

Question: Are property insurance rates rising due to inflation or higher risks of natural disasters?

Exploratory data analysis visualizations I plan to answer if insurance rates are rising at the levels of inflation or levels of increased natural disasters.

- Line plot of inflation rates over time
  - X-axis: years
  - Y-axis: inflation rates
  - linear regression model to find slope (rate of change)

Are homeowners insurance and commercial property insurance rates rising at the same rate? - Line plot to compare the rate of increase of homeowners insurance and commercial property insurance over time - X-axis: years - Y-axis: Insurance premiums - 2 different lines (homeowners and commercial insurance premiums) - Fit a linear regression model to both datasets to find the slope for homowners and commercial insurance premiums (rate of change)

- Line plot for the rate of increase of natural disaster occurances

  - two different color lines on graph (1 for Florida hurricanes and 1 for New York ice storms)
  - X axis- years, Y axis- number of
  - Fit a linear regression model to both datasets to find the slope for hurricane and ice storm occurances (rate of change)

- Scatter Plot with Regression Line
  - X axis- inflation rate, Y axis- homeowners insurance premium rates
  - X axis- inflation rate, Y axis- property insurance premium rates
  - Analyze slope of regression lines to check if positive correlation (as inflation rates increase, so do insurance premiums)

Once the data is visualized and trends are able to be depicted in insurance rates, inflation and natural disaster occurrences, I will use the data in a hypothesis test (paired t-test). This will allow me to statistically test whether the rise in insurance rates have a significant correlated with inflation or natural disasters.

- Null hypothesis: insurance rates are increasing at the same rate as Inflation
- Alternate hypothesis: insurance rates are increasing at a significantly higher rate, suggesting other contributing factors such as natural disasters.
- calculate the yearly percentage changes in both homeowners insurance premiums and inflation rates
- use t- test to get p- value, to determine whethere or not to reject null hypothesis
- Line Plot for Yearly Percentage Changes
  - X-axis: year, Y-axis: percentage change
  - 2 lines, 1 for % change insurance premiums, 1 % change inflation

I will use multiple linear regression in to quantify the relationships. It will determine how much the insurance rates are influenced by inflation and natural disasters to find which or if both factors have a strong correlation with the increase in insurance premiums.

Further evaluate Natural Disaster (Hurricane and Ice Storm data):

- Stacked Bar Plot (total property damages by year for both hurricanes and ice storms)
  - X-axis: Year
  - Y-axis: Total Property Damage
  - Stacked segments: Hurricanes and Ice Storms

- Make a simulation/ annimation of ice storm occurances in NY and hurricane occurances in FL on a map of each state that shows where within the state (by county) storms occured. The animation will go throughout time (2000-2024) and show markers of where/when storms occured. This will show whether or not storms were more common in some areas and if they got more frequent on the animation.

- Calendar heatmap plot
  - 2 rows (one for FL hurricanes and one for NY ice storms)
  - each row containing 12 squares, one for each calendar month
  - color represents the total property damage for all storms in that month
- Comparison of deaths and injuries for FL vs. NY with side by side line charts
  - X-axis: Year
  - Y-axis: Number of Deaths or Injuries
  - Two different color lines lines (FL and NY) for each plot
  - two plots: one for deaths, one for injuries

5 Work plan

Week 4 (2/10 - 2/16):

- Past project research (0.5 hours)
- Revise proposal (5 hours)
- Data tidying and recoding (1.5 hours)

Week 5 (2/17 - 2/23):

- Data tidying, recoding and merging dataset (4 hours)
- inflation line plot and regression model (2 hours)
- Start Homeowners insurance line plot and regression model (1 hour)

Week 6 (2/24 - 3/2):

- Finish Homeowners insurance line plot and regression model (1 hour)
- Commercial insurance line plot and regression model (2 hours)
- Florida Hurricanes line plot and regression model (2 hours)
- New York ice storm line plot and regression model (2 hours)

Week 7 (3/3 - 3/9):

- Start hypothesis test (paired t-test) (3 hours)
- Presentation prep and practice (4 hours)

Week 8 (3/10 - 3/16): *Presentations given on Wed-Thu 3/12-3/13.*

- Finish working on hypothesis test and visualization (1.5 hours)
- Poster prep (4 hours)
- Presentation peer review (1.5 hours)

Week 9 (3/24 - 3/30): *Poster Draft 1 due Monday morning 3/24 at 9am. Poster Draft 2 due Sunday night 3/30.*

- Multiple Linear Regression (3.5 hours)
- Peer feedback (2 hours)
- Poster revisions (1.5 hours)

Week 10 (3/31 - 4/6): *Final Poster due Sunday 4/6.*

- Natural disaster stacked bar plot (1.5 hours)
- Start working on animation (2 hours)
- Peer feedback (1.5 hours)
- Final Poster revisions (2 hours)

Week 11 (4/7 - 4/13):

- Work on animation (7 hours)

Week 12 (4/14 - 4/20):

- Calander Heat Map (3 hours)
- Prepare for DMC fair- practice presenting (4 hours)

Week 13 (4/21 - 4/27): *Blog post draft 1 due Sunday night 4/28.* [All project work should be done by the end of this week. The remaining time will be used for writing up and presenting your results.]

- Side by side line charts (3 hours)
- Draft blog post (4 hours).

Week 14 (4/28 - 5/4):

- Peer feedback (3 hours)
- Blog post revisions (4 hours)

Week 15 (5/5 - 5/8): *Final blog post due Tues 5/7. Blog post read-throughs during final exam slot, Thursday May 8th, 8:00-11:20am.*

- Blog post revisionsion (2 hours)
- Peer feedback (2 hours)

6 References

Here's an example of citing a source (see Phillips 1999, 33–35). Be sure the source information is entered in "BibTeX" form in the `references.bib` file.

[The bibliography will automatically get generated. Any sources you cite in the document will be included. Other entries in the `.bib` file will not be included.]

Phillips, T. P. 1999. "Possible Influence of the Magnetosphere on American History." *J. Oddball Res.* 98: 1000–1003.