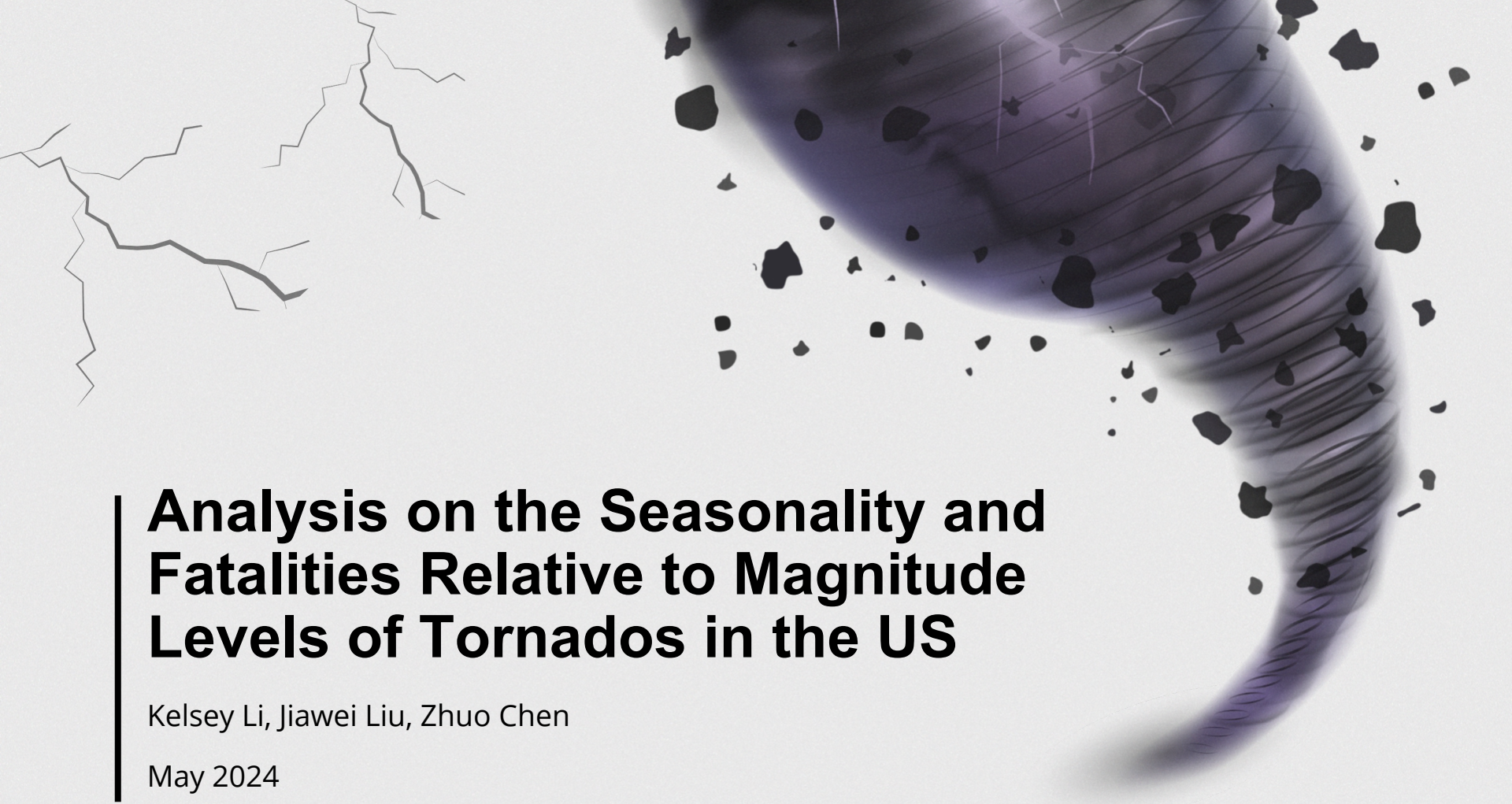


Analysis on the Seasonality and Fatalities Relative to Magnitude Levels of Tornados in the US

Kelsey Li, Jiawei Liu, Zhuo Chen

May 2024

Background

Motivation: Are tornados that scary?

- The data set is about US Tornado 1950-2021, capturing various attributes like date, location, magnitude, injuries and fatalities, and details of the tornado's path.
- The raw data is from **NOAA.** The latest update on Kaggle was in 2023.
- We would like to think the population is all tornadoes that have occurred in the US.

We chose this data for 4 major reasons:

1. Large enough for us to perform any analysis.
2. Result of meteorological data collection—no ethical issues.
3. A good combination of quantitative variables and categorical (or potential categorical) variables.
4. Real data from official government organizations—relatively credible and clean.

About Dataset - Overview

- Original dataset: 14 variables, 67558 observations
- Used the Fujita Scale before February, 2007 and then switched to the Enhance Fujita Scale afterwards—excluded any observations before 2008
- converted **mo** and **mag** from *int* to *fac*

Enhanced Fujita Scale

EFU	Unknown	No surveyable damage
EF0	65–85 mph	Light damage
EF1	86–110 mph	Moderate damage
EF2	111–135 mph	Considerable damage
EF3	136–165 mph	Severe damage
EF4	166–200 mph	Devastating damage
EF5	>200 mph	Incredible damage

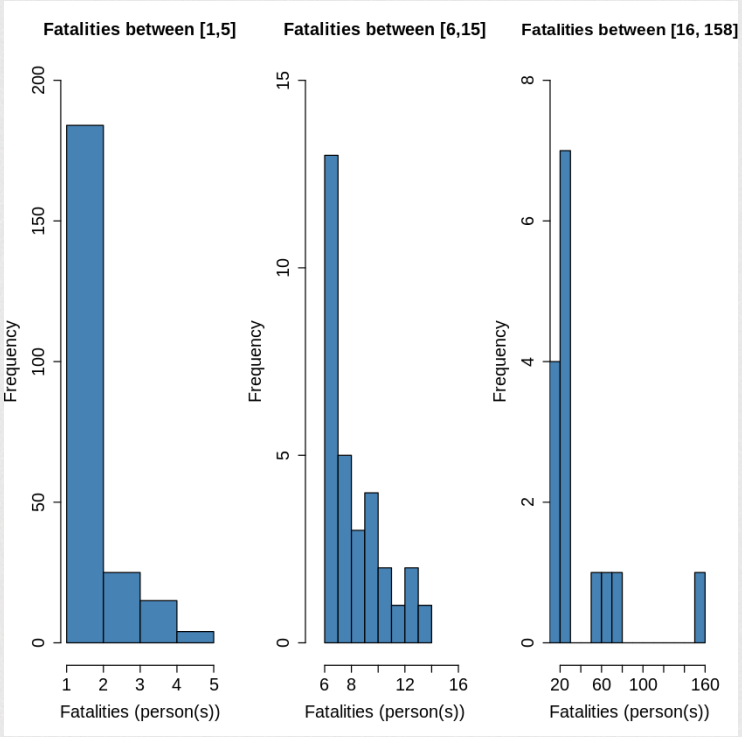
About Dataset - Overview

```
'data.frame':   17167 obs. of  4 variables:
 $ mo : Factor w/ 12 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ mag: Factor w/ 7 levels "0","1","2","3",..: 1 1 2 1 2 4 1 1 1 1 ...
 $ fat: int  0 0 0 0 0 0 0 0 0 0 ...
 $ len: num  0.1 0.1 2.6 2.35 3 13.2 0.07 0.17 0.3 0.25 ...
```

```
      mo        mag             fat                  len          
 5      :3765   0 :8941   Min.   :  0.00000   Min.   :  0.010  
 4      :3000   1 :5713   1st Qu.:  0.00000   1st Qu.:  0.390  
 6      :2507   2 :1455   Median :  0.00000   Median :  1.470  
 7      :1377   3 : 366   Mean   :  0.07212   Mean   :  3.559  
 3      :1217   4 :  79   3rd Qu.:  0.00000   3rd Qu.:  4.140  
 8      :1136   5 :   8   Max.   :158.00000   Max.   :168.530  
 (Other):4165   -9: 605                                        
```

EDA– Quantitative

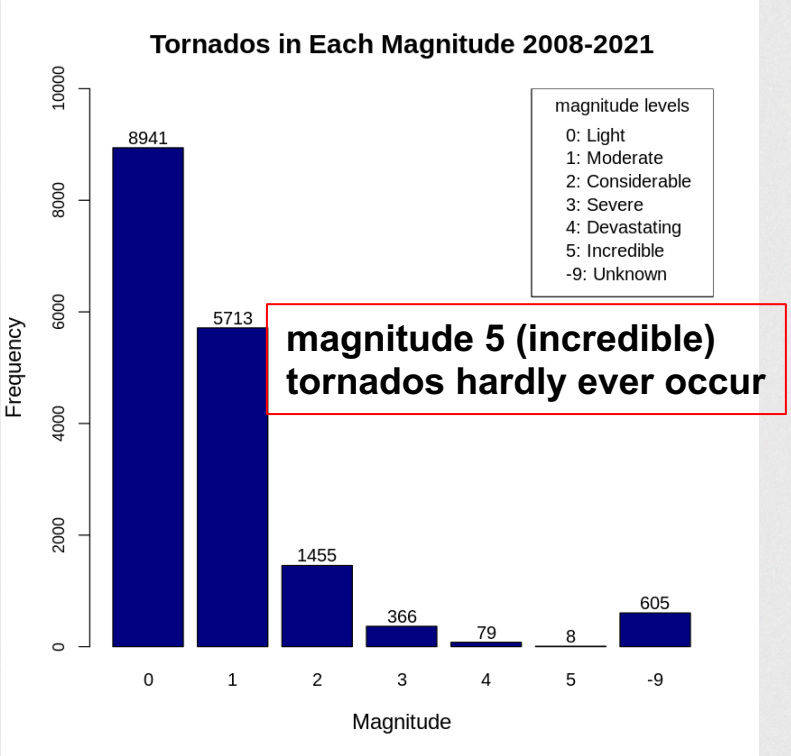
Number of tornados that did not produce any fatalities : `16893`

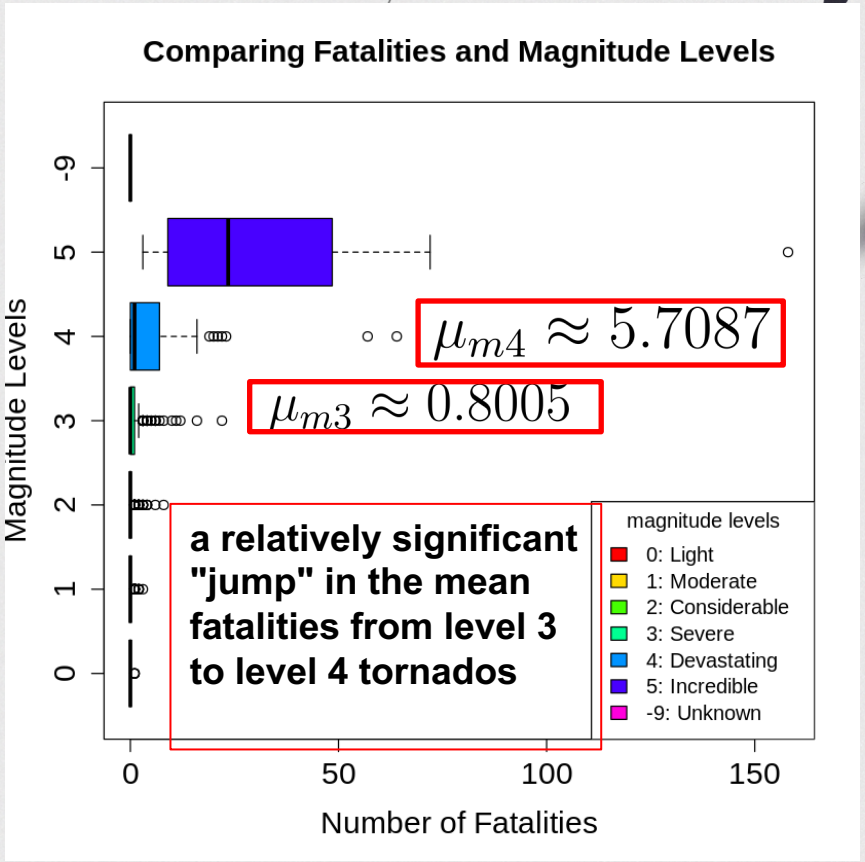


About fatalities:

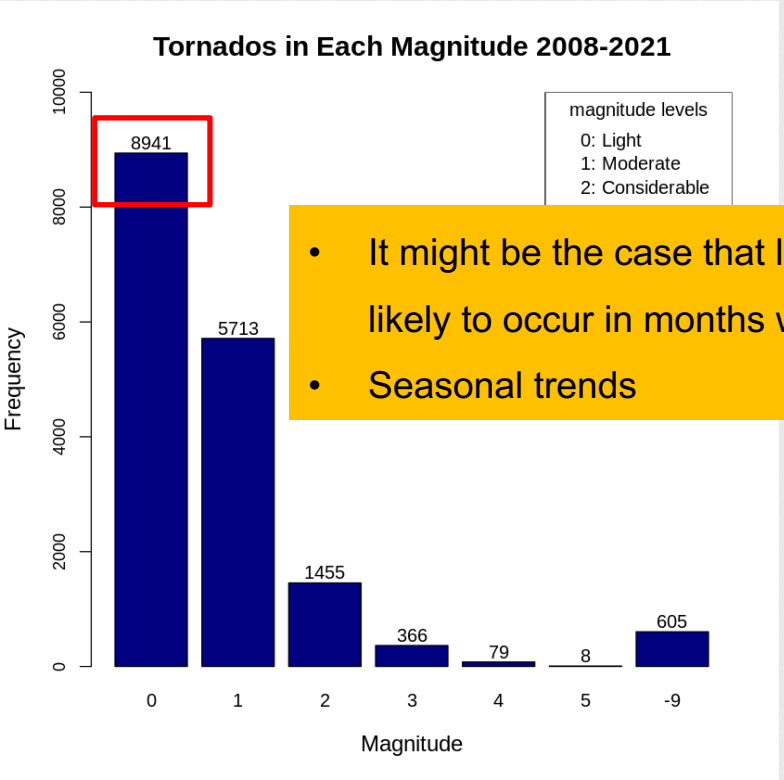
- Most tornadoes do not result in major fatalities
- However, the **maximum value of 158** indicates that tornadoes in extreme cases can still cause serious injuries and deaths

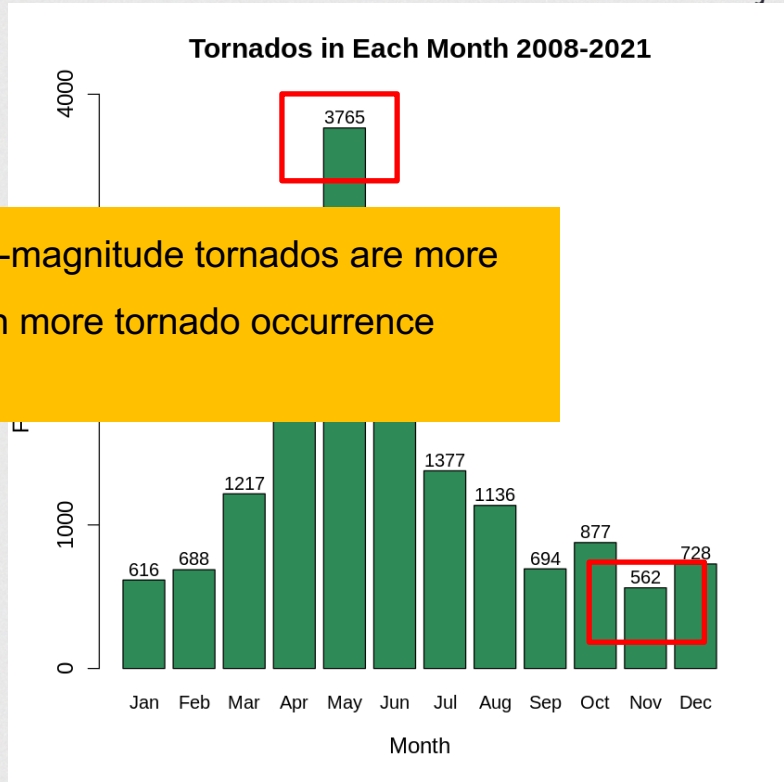
EDA– Quantitative





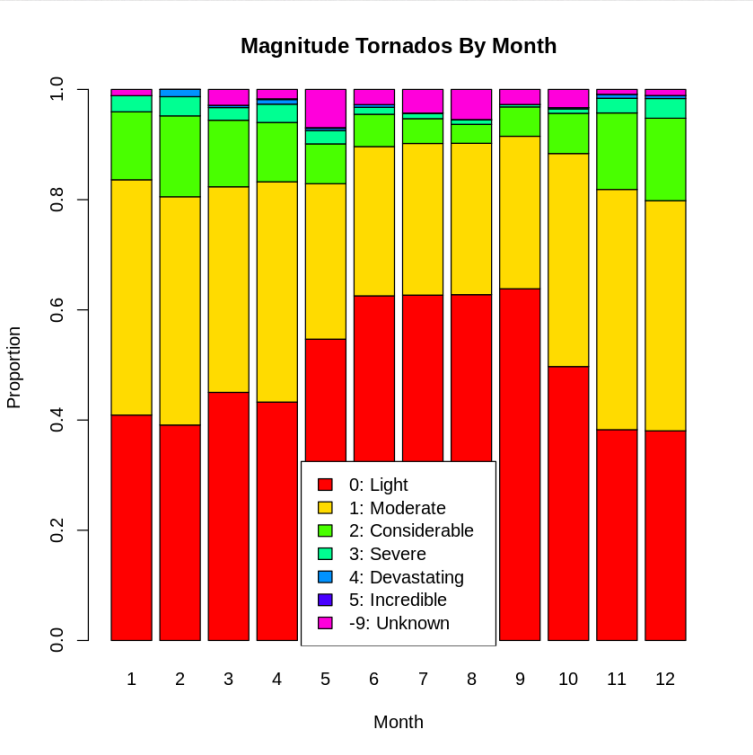
EDA – Qualitative





- It might be the case that low-magnitude tornados are more likely to occur in months with more tornado occurrence
- Seasonal trends

EDA – Qualitative



September has the "highest" red bar (highest proportion) which is a proportion of just over 60%.

May has one of the relatively higher red bars and November has what might be one of the shortest red bars.

2 specific proportions we are interested in are: p_{May}=0.5469, p_{Nov}=0.3826

Test for a Difference in Two Means

Does magnitude level 4 (Devastating) tornados necessarily cause more fatalities than magnitude level 3 (Severe) tornados on average in the US?

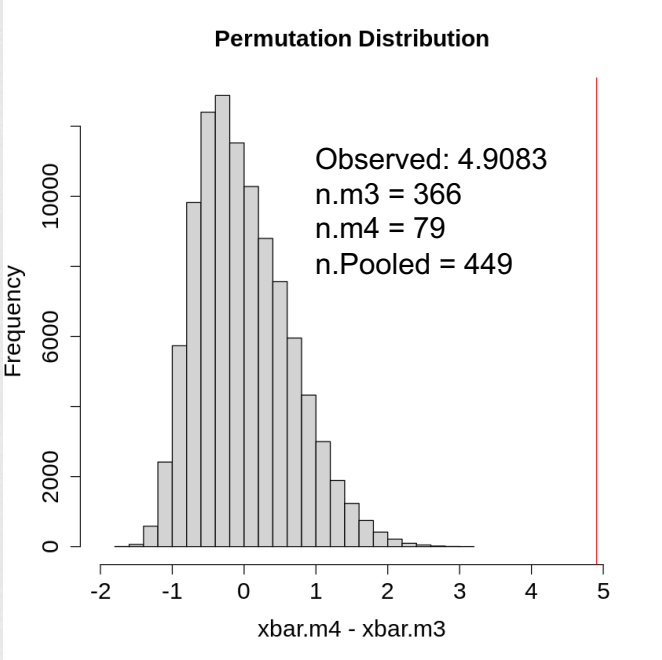
$$H_0 : \mu_{m4} - \mu_{m3} = 0$$

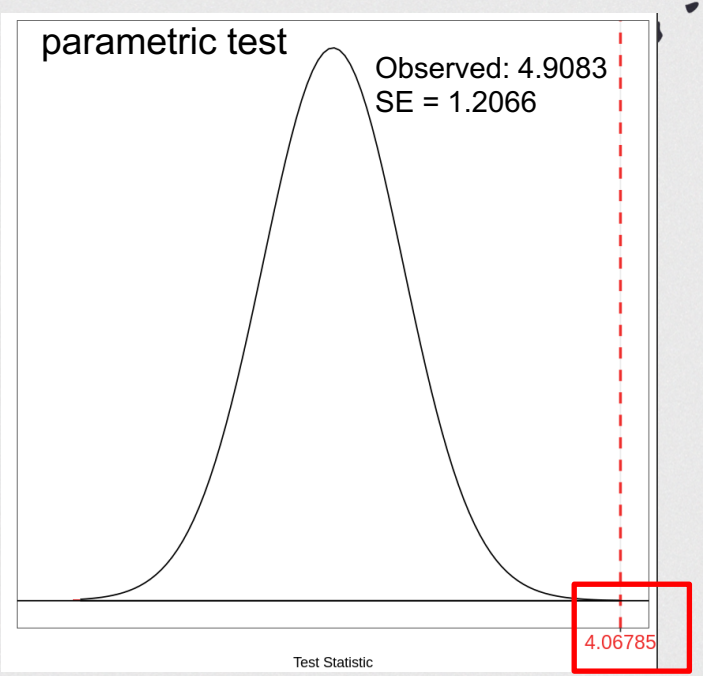
Magnitude level 4 (Devastating) tornados cause the same amount of fatalities as magnitude level 3 (Severe) tornados

$$H_a : \mu_{m4} - \mu_{m3} > 0$$

Magnitude level 4 (Devastating) tornados cause the same amount of fatalities as magnitude level 3 (Severe) tornados

Does magnitude level 4 (Devastating) tornados necessarily cause more fatalities than magnitude level 3 (Severe) tornados on average in the US?





$$\text{p-value}_{permutation} = 0.00001$$

$$\text{p-value}_{parametric} \approx 0.00005631403$$

Does magnitude level 4 (Devastating) tornados necessarily cause more fatalities than magnitude level 3 (Severe) tornados on average in the US?

$$\text{p-value}_{permutation} = 0.00001$$
$$\text{p-value}_{parametric} \approx 0.00005631403$$

- On a 5% significance level, both tests were statically significant and thus provide evidence for H_A that magnitude level 4 tornadoes result in more fatalities than magnitude level 3 tornadoes on average in the U.S.
- The distribution of the variable fat (fatalities) is highly skewed right. And while our permutation distribution captures that, our t-distribution doesn't since it is always symmetric—the permutation test might be more accurate.

Test for a Difference in Two Props

In the US, are magnitude level 0 (Light) tornados more likely to occur in May or November?

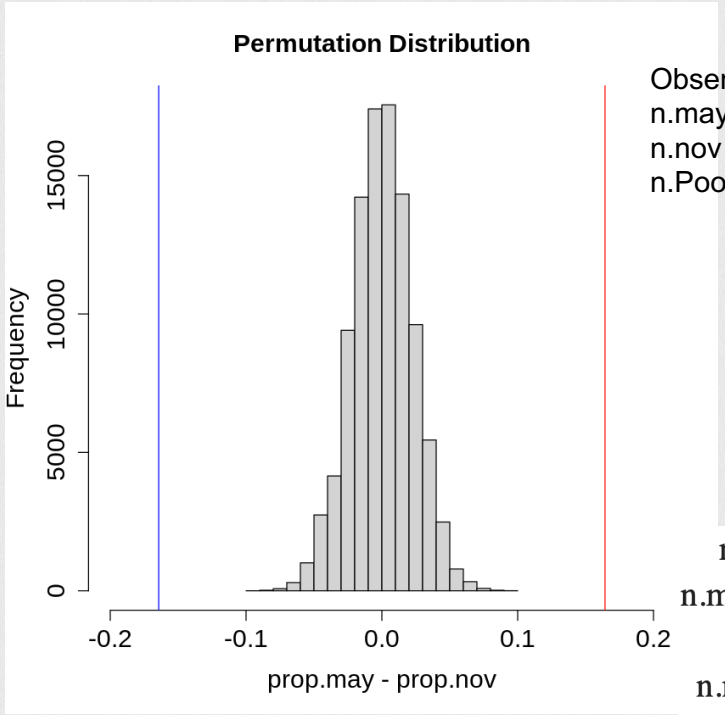
$$H_0 : p_{May} - p_{Nov} = 0$$

Magnitude level 0 (Light) tornados are **equally likely** to occur in May and November

$$H_a : p_{May} - p_{Nov} \neq 0$$

The likelihood of magnitude level 0 (Light) tornados occurring in May and November are **not the same**

In the US, are magnitude level 0 (Light) tornados more likely to occur in May or November?



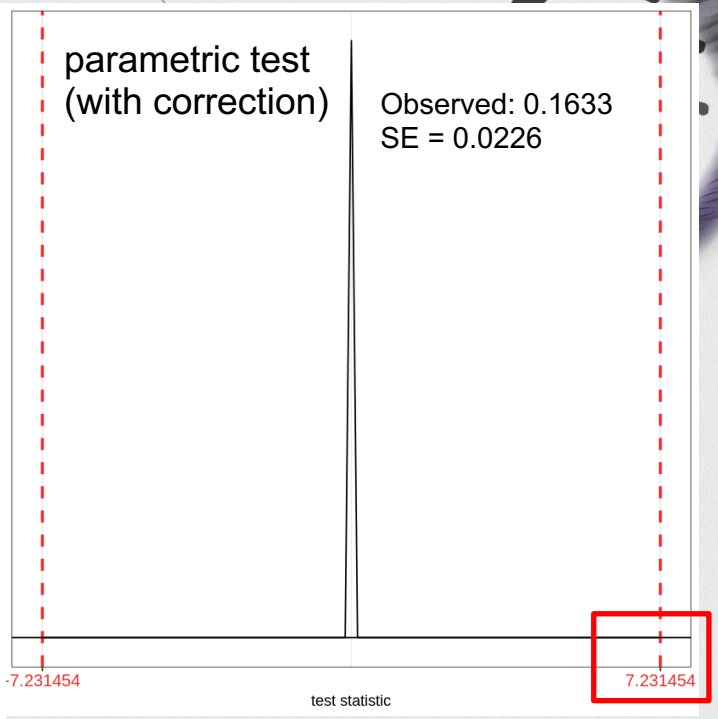
Observed: 0.1643
n.may = 3765
n.nov = 562
n.Pooled = 4327

$\text{n.may}\hat{p}_{may} = 2059 > 10$

$\text{n.may}(1 - \hat{p}_{may}) = 1706 > 10$

$\text{n.nov}\hat{p}_{nov} = 215 > 10$

$\text{n.nov}(1 - \hat{p}_{nov}) = 347 > 10$



$$\text{p-value}_{permutation} = 0.00002$$

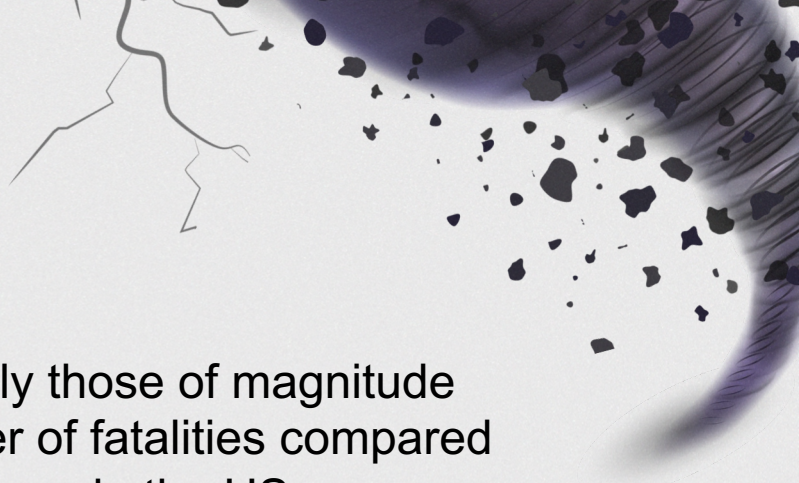
$$(\text{with correction})\text{p-value}_{parametric} = 4.778482403021 63 \times 10^{-13}$$

In the US, are magnitude level 0 (Light) tornados more likely to occur in May or November?

$$\text{p-value}_{permutation} = 0.00002$$
$$(\text{with correction})\text{p-value}_{parametric} = 4.77848240302163 \times 10^{-13}$$

- On a 5% significance level, both tests were statically significant and thus provide evidence for H_A that magnitude level 0 (Light) tornados are indeed more likely to occur in May than November on average in the U.S.
- In our permutation test, we did the resample only $N=10^5$ times, which means that for a two-tail test, the $\text{p-value}_{permutation}$ can only go as small as 2 x $\frac{1}{10^5}$ which is 0.00002 .

Conclusion



- Higher magnitude tornadoes, specifically those of magnitude level 4, indeed result in a higher number of fatalities compared to magnitude level 3 tornadoes on average in the US.
- Light tornadoes (magnitude level 0) are more likely to occur in May than in November on average in the US.
- These results have potential implications for climate policy and emergency management strategies.

Possible Future Work

- Can we generalize the trends in May and November into seasonal trends with more climatology research and analysis?
- Longitudinal trends in tornado magnitude and frequency - Investigating whether climate change is influencing the intensity and frequency of tornadoes.
- Impact of geographical factors on tornado outcomes - Include geographical features to model the path, magnitude, and impact of tornadoes.
- Are fatalities linked to any other geographical characteristics of the tornado, such as length of track and width?