

Statistical Analysis of Alberta Wildfires from (2006 - 2024)

Group4: Johanna Enright - 30303537, Uyoyoumena Orona - 30295200, Agambir Bandesha - 30302510, R

Date: 2025-10-03

For example, here is our data set in-text citation (Forestry & Parks, 2025) and the full citation will appear at the bottom.

Abstract

- maybe we want an abstract (which we will write at the end)?

Introduction

In Canada and across the world, we have seen an increase in forest area burned due to forest fires, with 2023 being an a record breaking year in forest damage for Canada (MacCarthy et al., 2025). Climate change is the main culprit, with changing weather patterns bringing heatwaves and drought which makes forests more easily ignitable than in the past (MacCarthy et al., 2025). Wildfires can be devastating to the forest ecosystem, but they also can have large ramifications on the economy and human health (Sadof & Anderson, 2023). When forests are burned, carbon dioxide is released into the air, further exacerbating our climate crisis and creating a positive feedback loop for future disastrous wildfires (Sadof & Anderson, 2023). Scientists have made significant efforts to understand forest fire dynamics allowing for improved forest fire mitigation and combative tactics *cite?*. However, it is of great importance to continue to research and understand how the forest fire landscape in Canada is changing so we can be better prepared to mitigate the environmental, economic, and social damages they will cause.

Our project looks at forest fire data in Alberta between 2006 and 2024 (Forestry & Parks, 2025). Specifically this data includes wildfires only occurring within the FPA (forest protection area), either provincial or federally owned/managed lands. Information on various fire dynamic factors, including total area burned, and provides time stamps and location coordinates for each wildfire entry (Forestry & Parks, 2025). We are narrowing our focus down to specifically investigate how weather factors, like wind speed, contribute to fire spread rate as well as how the cause of fires (specifically lightning) may be shifting over time due to climate change.

Investigation Goal

Question 1 Our initial area of interest is to investigate if the frequency of lightning-caused forest fires is increasing due to climate change. Research has indicated that the frequency of lightning-caused forest fires is increasing due to climate change (Beverly, 2024). We want to see if the Alberta forest fire data corroborates this claim. We hypothesize that the proportion of lightning-caused forest fires in 2024 will be greater than the proportion of lightning-caused forest fires in 2006.

Question 2: There are a variety of environmental factors which affect how quickly a fire spreads, ultimately contributing to the overall impact and environmental burden that a forest fire has on its surrounding area.

We are investigating how one key weather variable, wind speed, affects the fire spread rate. Our hypothesis is that a wind speed will be directly proportional with the fire spread rate.

Practical Implications

insert your research here

- a little bitty about how climate changing is increasing lightning strikes and making forests drier and more flammable here + practical implications here.

- a bit about wind and wildfire prediction here and models and why it's important here.

#####

##Variables of Interest *Question 1*: The variables we will analyze are fires within the “GENERAL_CAUSE” column of the data set which are “Lightning” (1 Success). All other fires within the “GENERAL_CAUSE” column which are not “Lightning” will be considered a 0 (Failure). We will analyze this variable for fires in the “YEAR” 2024 and compare it to fires from the “YEAR” 2006.

Question 2: For the linear regression, our independent (x) variable will be the numerical variable “WIND_SPEED” (in km/h) and our dependent (y) variable will be the numerical variable “FIRE_SPREAD_RATE” (in m/min).

Data set information

This data comes from the Government of Alberta official website, published by Forestry and Parks. The full data set name is Historical wildfire data: 2006 to 2024 (Forestry & Parks, 2025). It contains information licensed under the Open Government Licence - Alberta. This worldwide, royalty-free, perpetual, non-exclusive license allows us to copy, modify, publish, translate, adapt, distribute or otherwise use the information in any medium, mode or format for any lawful purpose (Government of Alberta). The data collection period was from 2006-01-01 to 2024-12-31. The resource was created on 2020-01-09 and was last modified on 2025-01-31. Updates to this data set are irregular in frequency.

Much of the wild fire information includes the personnel responsible for wildfire assessment and is time stamped (Forestry and Parks, 2025). The “GENERAL_CAUSE” of wildfires is investigated by Alberta wildfire officials. Wildfires under investigation where the cause was not determined were labelled “Undetermined”. “FIRE_SPREAD_RATE” and “WIND_SPEED” were recorded at the time of initial fire assessment. The exact assessors are Alberta Wildfire personnel and are indicated in the “ASSESSMENT_RESOURCE” column of the data set who assessed the fire at the recorded “ASSESSMENT_DATETIME”. Assessors include “air attack officer”, “initial action forces”, “wildfire assessor”, or “other”. More details on these assessment positions can be found in the Historical Wildfire Data Dictionary (Forestry and Parks, 2025).

Data Visualization

- here is where we can put our code to generate plots for distributions and visualizations and such.

Statistical Analyses

Question 1: Is the proportion of lightning-caused wildfires from 2024 significantly greater than those from 2006?

Here we will acquire the proportion data for lightning caused wild fires in 2006 and 2024.

```
##### determining proportion for lightning caused fires in 2006 and 2024 #####
```

```
#2006
```

```
df_2006 <- df %>% filter(YEAR == 2006)
```

```
n_2006 <- nrow(df_2006)
```

```
x_2006 <- df_2006 %>% filter(GENERAL_CAUSE == "Lightning") %>% nrow() #number from 2006 caused by ligh
```

```
p_lightning_2006 <- x_2006/n_2006
```

```
#2024
```

```
df_2024 <- df %>% filter(YEAR == 2024)
```

```
n_2024 <- nrow(df_2024)
```

```
x_2024 <- df_2024 %>% filter(GENERAL_CAUSE == "Lightning") %>% nrow() #x is number of fires caused by l
```

```
p_lightning_2024 <- x_2024/n_2024
```

```
print(paste("The 2006 proportion is", round(p_lightning_2006,4) ," and the 2024 proportion is", round(p
```

```
## [1] "The 2006 proportion is 0.3823 and the 2024 proportion is 0.4528"
```

Next we will confirm that our data follows the assumptions necessary to use a parametric test. (i.e. that n is sufficiently large under the central limit theorem).

```
## confirming it follows assumptions (np > 10) , (nq > 10) # is sufficiently large.
```

```
n_2006*p_lightning_2006 >10 ; n_2006*(1 - p_lightning_2006) > 10 ## greater than 10
```

```
## [1] TRUE
```

```
## [1] TRUE
```

```
n_2024*p_lightning_2024 > 10 ; n_2024*(1 - p_lightning_2024) > 10
```

```
## [1] TRUE
```

```
## [1] TRUE
```

Confirmed here, all values of n are sufficiently large to use the central limit theorem tests.

Testing hypothesis $H_0 : p_{2024} \leq p_{2006}$ and $H_A : p_{2024} > p_{2006}$

```
##### CALCULATING Z-observed #####
```

```
# Zobs = ((phat1 - phat2) - (p1 - p2)) / sqrt(phat*(1-phat)*(1/n1 + 1/n2))
```

```
phat <- (x_2006 + x_2024)/(n_2006 + n_2024)
```

```
p1_min_p2 <- 0 #under the assumption of the null hypothesis
```

```
Zobs <- ((p_lightning_2024 - p_lightning_2006) - p1_min_p2)/sqrt(phat*(1-phat)*(1/n_2024 + 1/n_2006))
```

```
pvalue <- round(1 - pnorm(Zobs),5)
```

```
print(paste("The test statistic is:", round(Zobs,4), "The p-value is:", pvalue))
```

```
## [1] "The test statistic is: 3.9418 The p-value is: 4e-05"
```

Therefore, because our p-value is less than our alpha of 0.05, we can conclude that the proportion of lightning caused wildfires in 2024 is significantly greater than the proportion from 2006.

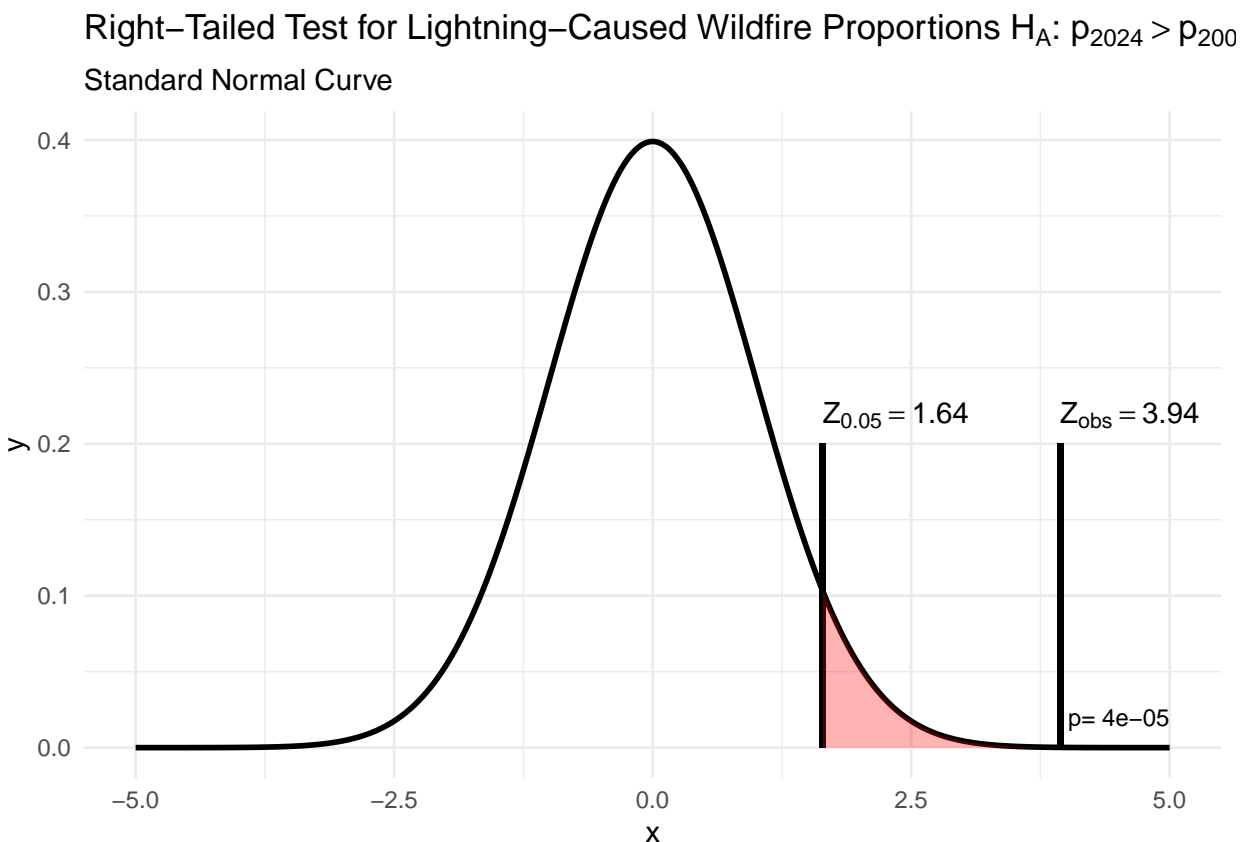
```

#set up dataframe
x <- seq(-5, 5, 0.001)
df_norm <- tibble(x = x, y = dnorm(x, mean = 0, sd = 1))

#labels for lines
labels_df <- tibble(x = c(qnorm(0.95), Zobs), y = 0.21,
                    label = c(bquote(Z[0.05] == .(round(qnorm(0.95), 2))), bquote(Z["obs"] == .(round(Zobs, 2)))))

p <- df_norm %>%
  ggplot(aes(x = x, y = y)) +
    geom_line(linewidth = 1) +
    geom_segment(aes(x = qnorm(0.95), xend = qnorm(0.95), y = 0, yend = 0.2),
                linewidth = 1, color = "black") +
    geom_segment(aes(x = Zobs, xend = Zobs, y = 0, yend = 0.2),
                linewidth = 1, color = "black") +
    geom_text(data = labels_df, aes(x = x, y = y, label = label), vjust = 0, hjust = 0, color = "black")
#shade in areas
geom_area(data = subset(df_norm, x > qnorm(0.95)), aes(x = x, y = y), fill = "red", alpha = 0.3) +
geom_area(data = subset(df_norm, x > Zobs), aes(x = x, y = y), fill = "blue", alpha = 1) + #shade
annotate(geom = "text", x = 5, y = 0.02, label = paste("p=", pvalue), size = 3, hjust = 1) +
ggtitle(bquote("Right-Tailed Test for Lightning-Caused Wildfire Proportions H[A] * ": p"[2024] > "p"
            subtitle = "Standard Normal Curve") +
theme_minimal()
print(p)

```



Next, we will confirm our findings with a 95% Agresti-Coull Confidence interval for the difference in propor-

tions, $p_{2024} - p_{2006}$.

```
##### USING THE AGRESTI-COULL (+1/+2 METHOD) 95% CI #####
# (ptilde1 - ptilde2) +/- (z (1 - alpha/2)) * sqrt((ptilde1*ptilde1/(n1 + 2)) + (ptilde2*ptilde2)/(n2 + 2))
#p_tilde = (x + 1)/ (n + 2)

p_tilde_2024 <- (x_2024 + 1)/(n_2024 + 2)
p_tilde_2006 <- (x_2006 + 1)/(n_2006 + 2)

ptilde_diff <- p_tilde_2024 - p_tilde_2006
z <- qnorm(1 - 0.05/2)
SE <- sqrt(((p_tilde_2024)*(1-p_tilde_2024)/(n_2024 + 2)) +
            ((p_tilde_2006)*(1-p_tilde_2006)/(n_2006 + 2))
            )

lb <- ptilde_diff - z * SE
ub <- ptilde_diff + z * SE
paste(round(lb, 4), "-", round(ub, 4))

## [1] "0.0353 - 0.1057"
```

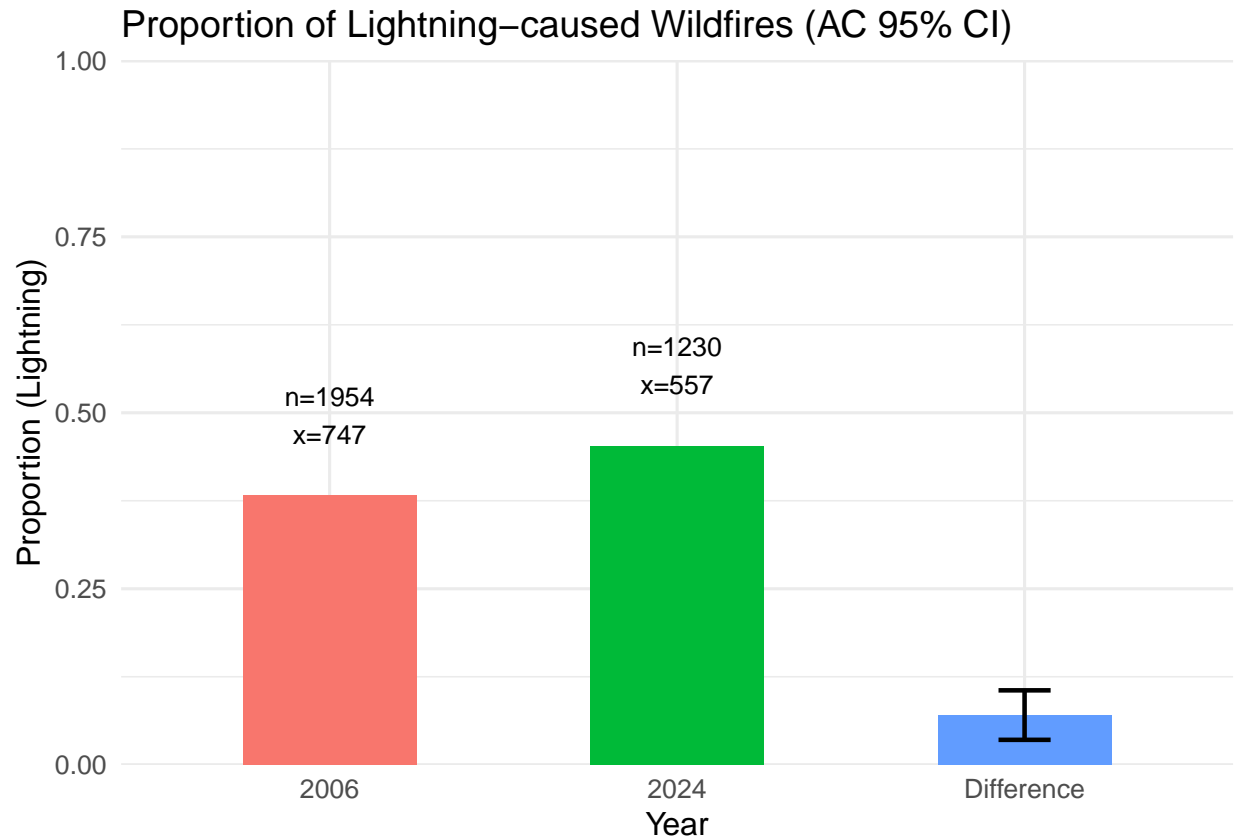
CONCLUSION : 2024 HAS A SIGNIFICANTLY HIGHER PROPORTION OF LIGHTNING-CAUSED WILDFIRES THAN 2006

Our 95% confidence interval for $p_{2024} - p_{2006}$ lies between 0.0353 - 0.1057. Both lower and upper bounds of our confidence interval are greater than 0, so this supports the above conclusion that the proportion of lightning caused wildfires from 2024 is significantly greater than that of 2006.

Here is a visualization for our 95% confidence interval.

```
# Visualization
plot_df <- tibble(
  YEAR = factor(c(2006, 2024, "Difference")),
  prop = c(p_lightning_2006, p_lightning_2024, p_lightning_2024-p_lightning_2006),
  lower = c(NA, NA, lb),
  upper = c(NA, NA, ub),
  n = c(n_2006, n_2024, NA),
  x = c(x_2006, x_2024, NA)
)

ggplot(plot_df, aes(x = YEAR, y = prop, fill = YEAR)) +
  geom_col(width = 0.5, show.legend = FALSE) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = 0.15, linewidth = 0.8) +
  geom_text(aes(label = ifelse(YEAR %in% c(2006, 2024), paste0("n=", n, "\nx=", x), NA)),
            vjust = -0.9, size = 3.5) +
  scale_y_continuous(limits = c(0, 1), expand = c(0, 0)) +
  labs(
    title = "Proportion of Lightning-caused Wildfires (AC 95% CI)",
    x = "Year",
    y = "Proportion (Lightning)"
  ) +
  theme_minimal(base_size = 12)
```



Question 2

- Agam and Mena, you guys can do your analyses here.

References

- Beverly, D., J. and Schroeder. (2024). Alberta's 2023 wildfires: Context, factors, and futures. *Canadian Journal of Forest Research*. <https://doi.org/10.1139/cjfr-2024-0099>
- Forestry and Parks. (2025). *Historical wildfire data dictionary: 2006 to 2024 [data dictionary]*. Open Government Alberta; Government of Alberta. <https://open.alberta.ca/opendata/wildfire-data#summary>
- Forestry, & Parks. (2025). *Historical wildfire data: 2006 to 2024 [data set]*. Open Government Alberta; Government of Alberta. <https://open.alberta.ca/opendata/wildfire-data#summary>
- MacCarthy, J., Richter, J., Tyukavina, A., & Harris, N. (2025). *The latest data confirms: Forest fires are getting worse*. World Resources Institute. <https://www.wri.org/insights/global-trends-forest-fires>
- Sadof, A., & Anderson, J. (2023). *5 ways wildfires affect people near and far*. World Resources Institute. <https://www.wri.org/insights/effects-wildfires-cities>