Avalanches

Mitchell Pudil

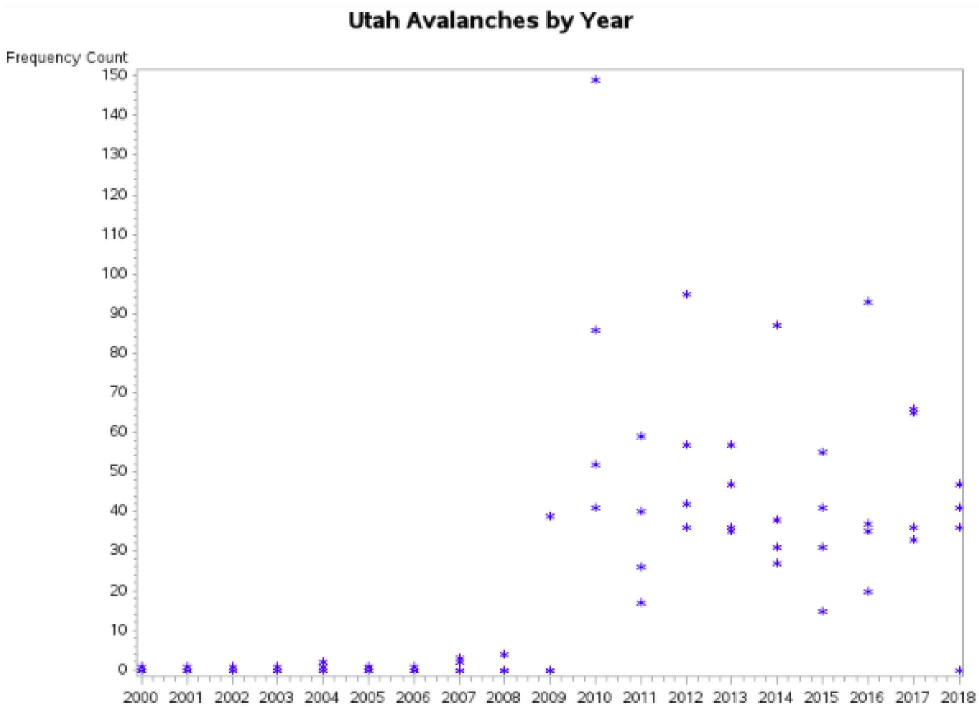
March 2019

1 Introduction

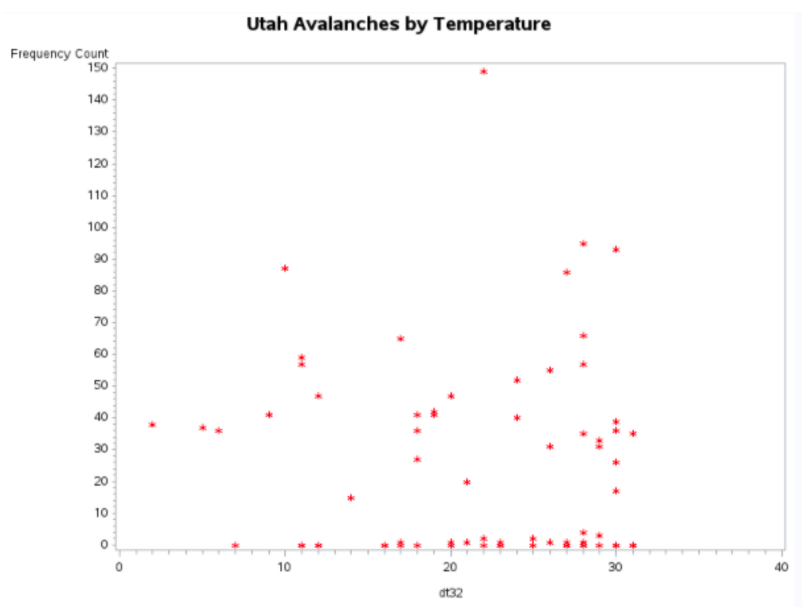
In this data analysis, we use merged data from Utah avalanche and weaher data in an attempt to predict the number of Utah avalanches. The data for each reported avalanche in Utah is at the Utah Avalanche Center webpage: https://utahavalanchecenter.org/observations. The data appears on over 200 webpages for each reported avalanche since 1965.

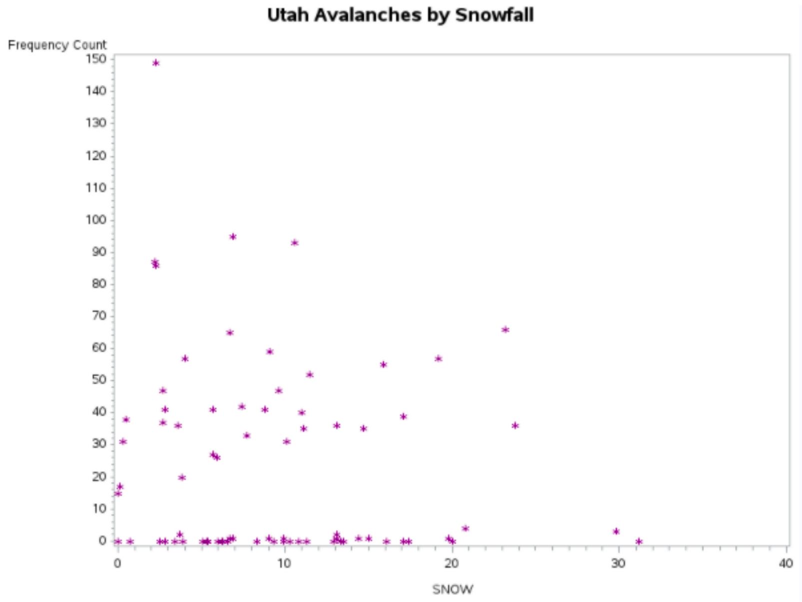
2 EDA

We begin the analysis by considering the relationship between the response variable, the mean number of avalanches, and all other variables. First, however, we look at the response variable alone, the number of avalanches, by year. A time series plot is shown below.



It appears that after 2009, the number of avalanches in Utah increased pretty dramatically. To further explain this phenomenon, additional plots which show the relationship between avalanches and the expanatory variables, snowfall and days below 32°. These plots are shown below.





Noticing the discrete values of avalanches, we appear to have a contradiction to our normality assumption. This is because normality suggests that any number, whether a discrete number or not, has a probability of being chosen if within the specified range. In this case, normality would require there to be a positive probability of having 2.5 avalanches, for example. However, this is of course not the case since the number of avalanches in a given year must be discrete. Hence, we have a violation of the normality assumption.

Looking at the above plot of DT32 against avalanches, we notice that there is slightly increasing variance and curvature. Generally speaking, in these situations, it is common to take the log of the dependent variable, which would transform the multiplicative model into an additive model. However, since there are many instances in which we observe 0 avalanches, if we were to take the log of the number of avalanches, we would not be able to use that data. For these reasons, we use a poisson regression, which allows discrete values of our dependent variable.

3 Analysis

The response variable in this case is the mean number of avalanches, and the explanatory variables are the snowfall and number of days below 32°. The poisson regression model we use is mathematically represented by:

$$\text{Avalanches} \mid \text{Snow, TMIN, D32} \sim \text{Poisson}(\exp(\hat{\beta}_0 + \hat{\beta}_1\text{DT32} + \hat{\beta}_2\text{TMIN}))$$

(1)

Running this regression results in a $\beta 1$ estimate of -0.0279 and a standard error of 0.0046. This suggests that for every extra inch of snow, we decrease the risk of avalanches by -.0279 or -2.79%, holding all else constant. A 95% confidence interval for how much the average number of avalanches will decrease by with an additional inch of snow is (-0.037, -0.019).

The Poisson regression also results in a β_{DT32} estimate of -0.0017 and a standard error of 0.0077. This means that for every additional day in which we observe a temperature below 32°, we decrease the risk of avalanches by -0.0017, or -0.17%. The 95% confidence interval here is (-0.0167, 0.0133), However, this value is not significant (p-value 0.822), so there doesn't appear to be a significant relationship between the number of days below 32°and the number of avalanches observed.

We also find that a 1°increase in temperature for the lowest temperature of the month decreases the number of avalanches by about -0.0279, or -2.79%, holding all else constant. Again, this value is not significant with a 95% confidence interval of (-0.0369, -0.0188).

The primary null hypothesis of interest is that snowfall does not have an effect on the number of avalanches. The χ^2 test statistic for that test is 36.40, and the p-value is less than 0.0001. Thus we conclude that snowfall does have a significant effect on the number of avalanches. We also find, as mentioned, that temperature does not have a statistically significant effect on the number of avalanches. The null hypothesis here is that there is no effect of temperature (either DT32 or TMIN) on the number of avalanches. Performing this ANOVA test results in a p-value <0.0001, suggesting that it is likely that DT32 and TMIN jointly have no effect.

We end the analysis be determining the expected number of avalanches for the months December through March. For December, the median is 4, for January it is 1, for February it is 2, and for March it is 0.5.

4 Conclusion

Overall, we find that temperature is not a good predictor of how many avalanches there will be. However, snowfall is a decent measure. A similar project to this would be determining how temperature affects the number of mold spores present in a house.