

Project title

Preregistration of analyses

marvelous echidna

Hypothesis 1

The frequency of landslides has increased over the time period covered in the dataset.

Analysis:

Create a line plot or time series visualization to show the trend in the number of landslide events over time. Fit a simple linear regression model with the date as the independent variable and the count of landslide events as the dependent variable. Evaluate the coefficient and statistical significance of the date variable to determine if there is a significant increase in landslide frequency over time.

Questions for Project Mentor:

Is a simple linear regression model appropriate for analyzing trends in landslide frequency over time, or should I consider alternative time series modeling approaches? Are there any potential confounding factors or seasonal patterns that I should account for in the analysis?

Hypothesis 2

Certain geographic regions or administrative divisions are more prone to experiencing deadlier landslides compared to others.

Analysis:

Group the data by country or administrative division and calculate the average fatality count per landslide event for each location. Identify the top 5-10 locations with the highest average fatality counts. Visualize the geographic distribution of average fatality counts using a choropleth map or other appropriate visualization. Perform a one-way ANOVA or Kruskal-Wallis

test to determine if there are statistically significant differences in average fatality counts between the different locations. If significant differences are found, conduct post-hoc tests to identify which locations are significantly more deadly compared to others.

Questions for Project Mentor:

Should I normalize the fatality counts by population or other factors to account for differences in region size? Are there any specific geographic or environmental factors that I should consider as potential explanations for the differences in landslide deadliness across regions?