Otzi's Adventures

Hailey Clark, Michael Girard, Brian Johnson

Datasets

We utilize two datasets in our project: the Himalayan Database and weather data from Mt. Everest.

The Himalyan Database is a digitized repository of data concerning all climbing expeditions in the Nepalese Himalayas from 1905 until spring of 2021. It contains detailed information about each expedition and the climbers. We pull a subset from the database; specifically, we focus on records from 2009-2019 on the eight 8,000m or greater peaks. We restrict our time frame to exclude any COVID-induced changes in climbing, and the "8000-er" peaks carry prestige in the mountaineering community. The dataset we extract contains 18,662 climber-expedition-level observations. Each observation contains variables to link a climber to a specific expedition, with demographic information, and with information pertaining to the climb itself, such as use of hired help, route taken, summit times, and supplemental oxygen utilization. The majority of these variables are boolean, while some are freetext or factors. Very few of the variables take on continuous numeric values.

The weather data from Mt. Everest is an initiative of National Geographic. National Geographic and Rolex sponsored an expedition in 2019 to install weather stations at 5 different locations on Mt. Everest. The data was intended to be used for research purposes and for climbers to access real-time weather conditions while on their expeditions. We utilize the downloadable dataset from Base Camp at 5315m. This dataset contains 14591 day-hour observations from November 2019 until Jun 2021. Each observation records the average temperature, humidity, and barometric pressure and total precipitation over the course of the hour. All of these variables take on continuous numeric values.

Overview of Analyses

We conduct a series of descriptive analyses to understand the climber dataset and to inform our later analyses. We tabulate climber demographic variables, expedition-level variables, and variables involving climbing complications. We also construct a number of survival plots to ascertain how climbing success might differ by demographic or time of expedition. We then attempt to fit geometric, gamma, and beta distributions to variables that already existed in the climber dataset or were derived from the climber dataset.

We turn our focus to summiting these peaks. We run a permutation test to see if the time that a climber attempts to summits is related to the probability of a climber dying. We also compare the success rates of summiting by occupation with a permutation test and a chi-squared test on a contingency table. We employ a regression-based approach to understand the impact that supplemental oxygen may have on summit success for climbers with multiple attempts. We adapt this model into a logistic model to estimate the impact of supplemental oxygen on the odds of summiting on a single attempt.

We build and compare linear models for predicting barometric pressure with temperature as our main predictor. We construct 95% confidence intervals for the average temperature at Everest basecamp during the spring hiking season. Then, we conduct Fourier analysis to capture the seasonal variation in temperature.

We run power calculations to determine the number of randomized expeditions we would need to send up with either mandatory oxygen usage or optional oxygen usage in order to detect the treatment effect of oxygen on climbing success. This analysis is run specifically to compare the results from classical methods and from a simulation-based approach.