Project: Outliers
Himalayan Expeditions

Aston Yong, Catherine Ning, and Matthew Hawkins

Introduction and Motivation

- Our topic centers around expeditions that have climbed in the Nepal Himalaya
- Motivation: Discover what are necessary factors potentially leading to a safe expedition to ensure the success of the expedition
- Research question: What expedition factors are most predictive of safe and successful expeditions and what is the probability of success given these predictors?

The Dataset

Our datasets consists of 3 csv files:

- Expeditions - Main dataset
- Members
- Peaks

```
Rows: 8,770
Columns: 11
$ height_metres      <dbl> 6055, 6055, 6055, 6814, 6814, 6814, 6814, 681…
$ member_count       <dbl> 5, 9, 11, 5, 6, 4, 2, 4, 2, 6, 10, 6, 5, 6, 8…
$ hired_staff        <dbl> 0, 1, 2, 0, 1, 2, 2, 0, 2, 0, 3, 0, 2, 2, 0, …
$ oxygen_used        <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FAL…
$ year               <dbl> 2015, 2018, 2015, 2000, 2000, 2000, 2000, 200…
$ season             <chr> "Autumn", "Autumn", "Autumn", "Spring", "Spri…
$ trekking_agency    <fct> Cosmo Treks, Other, Other, Other, Other, Asia…
$ member_deaths      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ hired_staff_deaths <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ mean_age           <dbl> 22.400, 44.333, 49.091, 30.600, 33.167, 40.25…
$ success_expedition <fct> 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, …
```

Data manipulation:

- Join grouped members info and peak height into expeditions
- Create the the variable "success expedition" using termination reason
- Select useful variables and remove null values

Exploratory Data Analysis Highlights

- Through manipulation and classification of data, examined distribution of response variable, the expedition outcome (shown below)
  - 55.1% of expeditions classified as successes
- Analyzed distributions of important predictor variables, such as use of oxygen, number of members in the expedition, and season
- Found important summary statistics to describe the distribution of number of members in expedition
  - Mean: 5.953
  - SD: 5.428
  - IQR: 6
  - Right-skewed distribution
- Distribution of Season:
  - Unsurprisingly, primarily Spring and Summer

Number of Members Histogram

Distribution of Expedition Outcomes
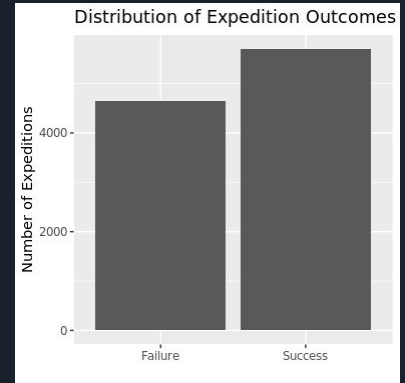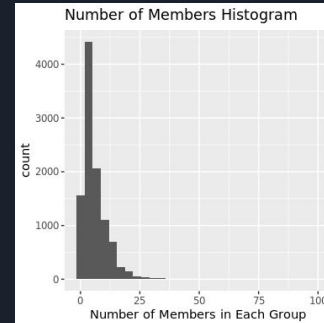
Final Model

- Forward selection BIC with quadratic mean_age_cent

- Mean centered variables

- Conditions checked - drop in deviance, emplogitplots, etc

- Predictive power and fit: AUC = 0.747

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-0.12853	0.11233	-1.14423	0.25253	-0.34879	0.09158
oxygen_usedTRUE	-0.77096	0.25250	-3.05335	0.00226	-1.26556	-0.27532
height_metres_cent	-0.00063	0.00004	-17.59754	0.00000	-0.00071	-0.00056
member_count_cent	0.06468	0.00521	12.41366	0.00000	0.05456	0.07499
mean_age_cent	-0.02315	0.00388	-5.96951	0.00000	-0.03077	-0.01556
mean_age_cent_sq	-0.00219	0.00036	-6.10579	0.00000	-0.00290	-0.00150
hired_staff_deaths	-0.81198	0.16755	-4.84619	0.00000	-1.15379	-0.49903
year	-0.00155	0.00321	-0.48284	0.62921	-0.00785	0.00474
oxygen_usedTRUE:height_metres_cent	0.00098	0.00012	8.18924	0.00000	0.00075	0.00122
oxygen_usedTRUE:year	0.06035	0.00674	8.94971	0.00000	0.04715	0.07359
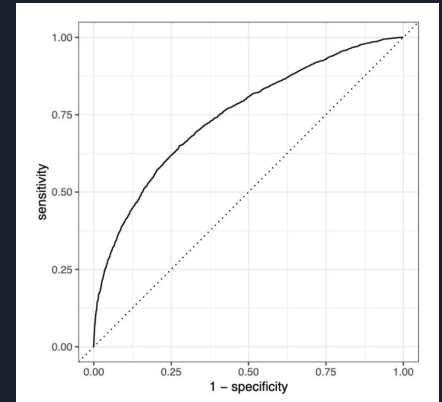
Interesting Findings

- Baseline odds of success are higher than expected

- Log odds decrease with year since 1971

- Oxygen use is a key indicator of expedition failure

- Oxygen use and year interaction term

  - Experience? Technology?

- Impact of mean_age_cent reaches maximum at

-5.29 years (33.2 years)

Conclusions and Future Work

- Significant Predictors:
  - Peak height, number of members, mean member age, staff deaths, year of expedition, and oxygen use
- P-values below 0.05 for all variables but year
  - Drop-in-deviance test to confirm validity of the year variable
- Limitations:
  - Unable to examine all interactions between variables
  - Missing value data points
  - Other potential models with better fit
  - Concerns of independence
    - Expeditions on certain mountains might have higher/lower odds of success
- Future Work:
  - Analyze more of the interactions between different variables
  - Trying different models to explain the variance in the success of an expedition more effectively
  - Examining potential independence problems