Analyzing Factors Influencing Success of Everest Expeditions

Caleb Woessner, Terence Cunningham, Jeffrey Goodstein, and Jane Weissberg

Abstract

This project investigates factors that influence the probability of a successful Mount Everest expedition. Using data from the Himalayan Database and Tidy Tuesdays GitHub page, we compiled a dataset of 1,067 Everest expeditions from 2010 to 2024. After an initial exploratory data analysis and cleaning process, we selected several variables potentially associated with success, including number of camps used past base camp, number of hired personnel, and use of supplemental oxygen. Exploratory analysis revealed that variables like season and face approached did not meaningfully affect success rates and thus were excluded from the model.

We fit a logistic regression model to predict the probability of a successful summit attempt. Our results showed that each additional high camp was associated with approximately a 49% increase in the odds of success, likely due to better altitude acclimatization strategies. The number of personnel hired also had a positive impact, although with diminishing returns as more personnel were added. The most influential factor was the use of supplemental oxygen, which increased the odds of reaching the summit by over 100 times compared to expeditions without oxygen support.

These findings highlight the critical role of logistical support and resource management in achieving successful accents of Everest. Strategic use of camps and hiring additional sherpas significantly enhance success chances, while oxygen use remains an essential tool for overcoming the physiological challenges posed by extreme altitudes. Factors such as total expedition size and the route's face did not statistically impact outcomes, challenging some common assumptions. Our analysis offers practical insights for climbers and expedition planners seeking to improve their chances of success. Future research could further enhance our model by examining additional factors such as climber experience, summit-day crowding, and real-time weather conditions.

Introduction

Scaling Mount Everest, the world's tallest mountain above sea level, is no easy feat. Difficult terrain, quickly changing weather conditions, and lack of atmospheric oxygen at the peak make the climb one of the world's most difficult. Every year, hundreds of dedicated climbers travel to the Himalayas to attempt the climb, and only about 60% of expeditions reach the summit. Climbing Everest is very risky, and very expensive- permits can be upwards of $50,000. It is thus in the climbers' best interest to maximize their probability of successfully reaching the summit.

Several factors may influence the likelihood of a successful climb, including the expedition start date, expedition size, number of hired personnel, number of camps used, use of supplemental oxygen, and the route taken. Our goal was to develop a model that predicts the probability of success based on these variables.

Data and Methods

Describe the data you used in your analysis, including the source, characteristics, and any pre-processing or cleaning steps you performed.

This is a good place to include the results of your exploratory data analysis.

You should also explain the methods you employed. I expect you to define the statistical model you chose for your data in mathematical form, as we did throughout the course.

The data used in our analysis comes from the Himalayan Database, a compilation of records of all expeditions climbed in the Himalayas. The database contains information on over 11,000 expeditions dating back to 1905, but to manage size, the dataset on the Tidy Tuesdays GitHub page only includes 882 expeditions between all mountains from 2020-2024, of which 189 were on Everest.

Due to the Covid-19 pandemic, expeditions on Mount Everest were severely limited, and thus the sample during this period was very small. To rectify this, we downloaded the original data from the Himalayan database website and combined it with the data from the

Tidy Tuesdays page, extending the range to all expeditions on Everest since 2010, of which there were 1,067.

Variables identified as potentially affecting the probability of a successful climb included the season in which the climb began, the face of the mountain approached, the number of camps used beyond base camp, expedition size, the number of hired guides, and whether supplemental oxygen was used.

While season is known to have a significant impact on expedition success, this relationship is already well established. Additionally, because the vast majority of Everest expeditions begin in the spring, including season in the model would offer little explanatory power and was therefore excluded.

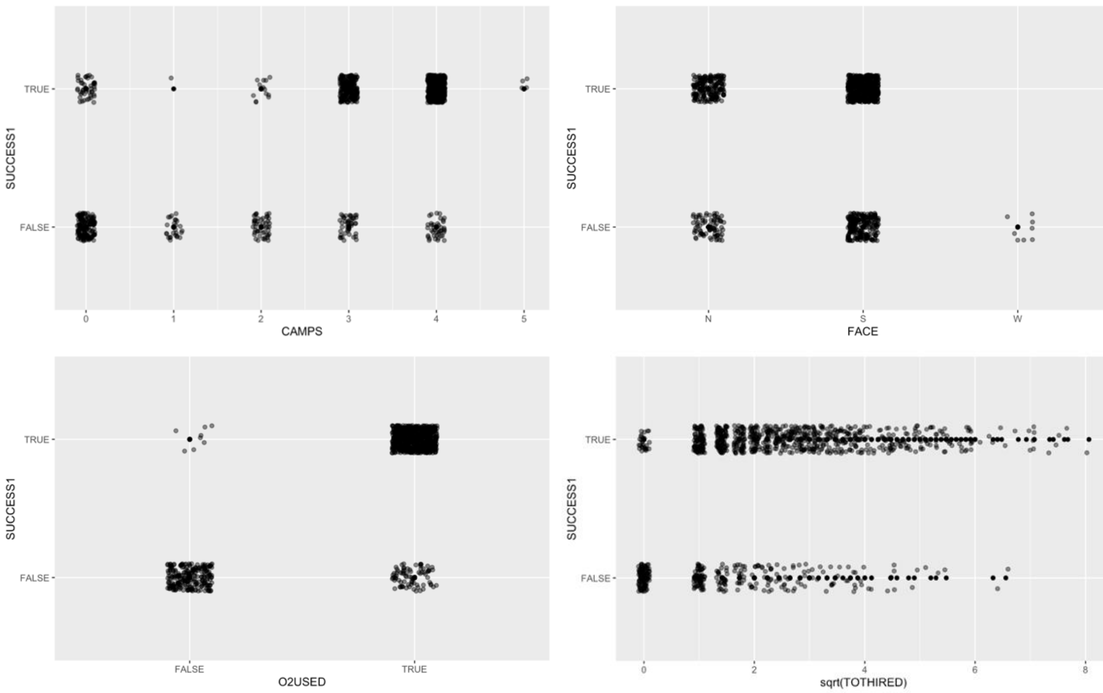
To simplify the many distinct routes used, we created a new variable, **face**, which categorizes each route by the face of the mountain it approaches (north, east, or west). However, exploratory data analysis indicated that face does not have a strong effect on the probability of success.

In contrast, the number of camps used beyond base camp showed a potential relationship with success, based on a scatter plot analysis. As a result, this variable was included in the model.

The distribution of the number of hired guides is right-skewed, with many natural zero values, making a log transformation inappropriate. To address the skew while preserving the zeroes, we applied a square root transformation to this variable in the model.

Exploratory analysis suggested that the total expedition size does not have a meaningful impact on the probability of success, so it was excluded from the model.

In contrast, the use of supplemental oxygen showed a clear association with increased success rates and was therefore included as a predictor.



Based on the results of our exploratory analysis and variable selection process, the following logistic regression model was constructed to estimate the probability of a successful climb:

$$Y_i \sim \text{Bernoulli}(p_i), \quad \text{for } i = 1, \ldots, 1067$$

$$\text{logit}(p_i) = \beta_0 + \beta_1 \cdot \text{CAMPS}_i + \beta_2 \cdot \sqrt{\text{TOTHIRED}_i} + \beta_3 \cdot I(\text{O2USED}_i = \text{TRUE})$$

Results

Present your results in this section, including figures, tables, and any other relevant outputs from your analysis.

In this section, discuss and interpret your results. Explain the implications and significance of your findings and relate them to your research question and existing literature or theories.

Table 1: Logistic Regression Summary

	Est.	SE.	Z val.	p
Intercept	-4.81499	0.46254	-10.410	<2e-16***
CAMPS	0.39650	0.08452	4.691	2.72e-06***
Sqrt(TOTHIRED)	0.51120	0.09597	5.327	1.00e-07***
O2USEDTRUE	4.72027	0.41963	11.249	<2e-16***

Null deviance: 1291.3 on 1066 degrees of freedom

Residual deviance: 536.17 on 1061 degrees of freedom

These results suggest that the number of camps, number of hired guides and whether supplemental oxygen was used all have significant effects on the probability of an expedition successfully reaching the summit. Factors like size of expedition, face approached from, and member deaths did not have a significant effect on a successful expedition.

Model Interpretations:

Our logistic regression model identifies several key factors that significantly impact the likelihood of a successful Everest summit. Specifically, each **additional camp** used increases the odds of success by approximately **49%**.

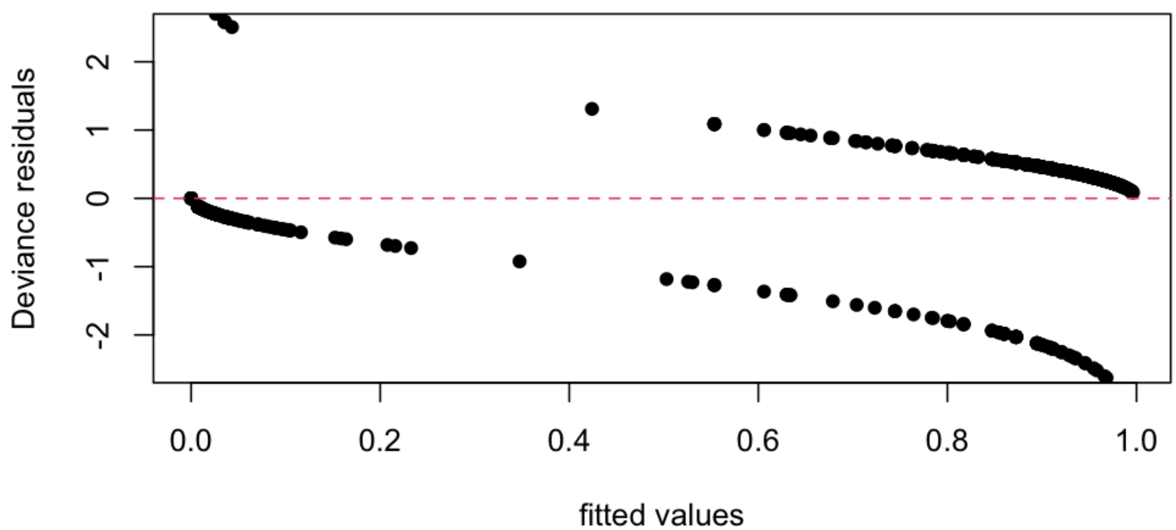
A one-unit increase in the **square root of total hired personnel** increases the odds by about **67%**, holding all other variables constant. However, we see **diminishing returns** as the number of personnel hired increases. An increase from 4 to 5 sherpas increases the odds by about **13%** whereas an increase from 9 to 10 sherpas only increases the odds of success by **9%**.

The use of **supplemental oxygen** has the most substantial effect, increasing the odds of success by over **100** times compared to expeditions that do not use it. These findings highlight the importance of strategic planning and resource use in improving expedition outcomes.

Confidence Intervals:

A 95% confidence interval for the number of **camps** used reveals that the true value of beta 1 is (0.228, 0.56). Exponentiating this result, the effect of a one-unit increase in the number of camps results in a **(1.256,1.75)** multiplier on the odds of success.

A 95% confidence interval for the square root of the number of hired guides reveals that the true value of beta 2 is (0.32881, 0.70586). Exponentiating this result, the effect of a one-unit increase in the square root of hired guides results in a **(1.389, 2.025)** multiplier on the odds of success.



The graph of deviance residuals vs fitted values shows expected behavior for binomial models.

Conclusion

Summarize your main findings and conclusions in this section. Highlight the key takeaways and contributions of your analysis and provide recommendations or suggestions for future research or practical applications based on your results.

Our analysis provides clear evidence that several logistical decisions significantly impact the success of Mount Everest expeditions. Using a logistic regression model, we found that three factors: number of high camps used, number of hired personnel (sherpas), and use of supplemental oxygen are all strong predictors of an expedition's probability of reaching the summit.

Each additional camp beyond base camp was associated with a higher probability of success, likely reflecting more strategic altitude acclimatization. Similarly, expeditions that hired more personnel had greater success, suggesting the value of additional experience and support in both carrying supplies and ensuring safety. Most importantly, the use of supplemental oxygen had a dramatic effect, increasing the odds of summiting by more than 100 times. This aligns with the known physiological challenges climbers face in the low-oxygen environment above 8,000 meters.

By contrast, factors such as expedition size, route (face approached), and whether a member died during the expedition did not significantly influence success in our model, highlighting that not all intuitively important factors bear out statistically.

These findings can inform both individual climbers and expedition organizers by emphasizing the importance of support infrastructure (camps and sherpas) and oxygen use. Future research could incorporate additional variables such as weather conditions, climber experience, and crowding on summit day to refine predictive accuracy. Moreover, similar analyses could be extended to other Himalayan peaks or stratified by expedition nationality or leadership to uncover more nuanced patterns.

Overall, our model offers valuable insights into what contributes most to a successful Everest expedition and can help climbers make more informed, safer decisions on the mountain.

References

Appendix (Optional)

If you want to include your R code as an appendix, you can create a new code chunk and set #| echo: true to show the code and #| eval: false to avoid that to be run. You can also provide scripts separately as supplementary material.