Risk and Success Factors in Himalayan Mountaineering*

A Comprehensive Analysis of Historical Expedition Data (1905-2019)

Jiaqi Lai

December 13, 2024

This study examines the key factors influencing safety and success in Himalayan mountaineering expeditions. Using comprehensive data from the Himalayan Database covering expeditions from 1905 through Spring 2019, we analyze various factors affecting expedition outcomes through Bayesian modeling approaches. The findings provide insights into risk management and success factors in high-altitude mountaineering.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/lizl66/himalayan_expedition/tree/main

1 Introduction

Himalayan mountaineering represents an important area of human exploration and self-challenge. This sport not only tests climbers' physical fitness and skills, but also requires careful strategic planning and risk management. In this context, a key question emerges: What factors affect the success rate and safety of mountaineering activities? This study is dedicated to answering this question.

Among many mountain ranges in the world, the Himalayas are famous for their unique geographical location and extremely high altitude. This study focuses on Himalayan mountaineering activities in Nepal, including bilateral climbing records of famous frontier peaks such as Mount Everest, Cho Oyu, Makalu and Kangchenjunga.

My research objective (estimand) is to explore the relationship between multiple key factors and mountaineering success rate. These factors include altitude, season selection, team member role, oxygen use, and team size. By analyzing this data, we aim to identify the key factors that affect mountaineering success and quantify the extent of these factors.

This study uses a dataset from Cookson (2020), which is derived from the Himalayan database (Salisbury 2023). We use a Bayesian logistic regression model to predict the probability of mountaineering success based on various factors. This approach allows us to not only quantify the impact of each factor, but also assess the uncertainty of these impacts.

The study found that the most significant factor affecting the success rate of mountaineering is the use of oxygen, followed by the role category of the climber and the choice of season. Specifically, climbers who use oxygen have a significantly higher success rate than those who do not use oxygen; professional climbers (such as Sherpas) have a higher success rate than ordinary climbers; and mountaineering activities in spring and autumn are more likely to succeed than in summer and winter. At the same time, the success rate of mountaineering shows a clear downward trend with increasing altitude.

These findings have important practical significance for the planning and risk management of mountaineering activities. They can help climbers and organizers make more informed decisions and improve the safety and success rate of mountaineering activities. At the same time, these findings also provide data support for standard setting and safety management in the field of high-altitude mountaineering.

The structure of this paper is as follows: Section 2 introduces the source and processing of the data; Section 3 details the Bayesian logistic regression model used; Section 4 presents the analysis results; Section 5 discusses the implications of the main findings in depth and points out the limitations of the study and future research directions. This structural arrangement enables readers to fully understand the research process and results.

2 Data

For this analysis, I have combined three datasets into one, which is used for analysis. The datasets were cleaned and analysed using the statistical programming software `R` (R Core Team 2023) along with the help of `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2014), `ggplot2` (Wickham 2016), `here` (Müller 2020), `dplyr` (Wickham et al. 2023), `rstanarm` (Goodrich et al. 2024), `arrow` (Richardson et al. 2024), `broom.mixed` (Bolker and Robinson 2022) and `kableExtra` (Zhu 2024).

2.1 Analysis Dataset

The raw datasets were obtained from Cookson (2020), who sourced them from The Himalayan Database (Salisbury 2023).

The Himalayan Database is a compilation of records for all expeditions that have climbed in the Nepal Himalaya. The database is based on the expedition archives of Elizabeth Hawley, a long-time journalist based in Kathmandu, and it is supplemented by information gathered from books, alpine journals and correspondence with Himalayan climbers.

The original database currently covers all expeditions from 1905 through Spring-Summer 2023 to the most significant mountaineering peaks in Nepal. Also included are expeditions to both sides of border peaks such as Everest, Cho Oyu, Makalu and Kangchenjunga as well as to some smaller border peaks. Data on expeditions to trekking peaks are included for early attempts, first ascents and major accidents. The updates to this database are published bi-annually.

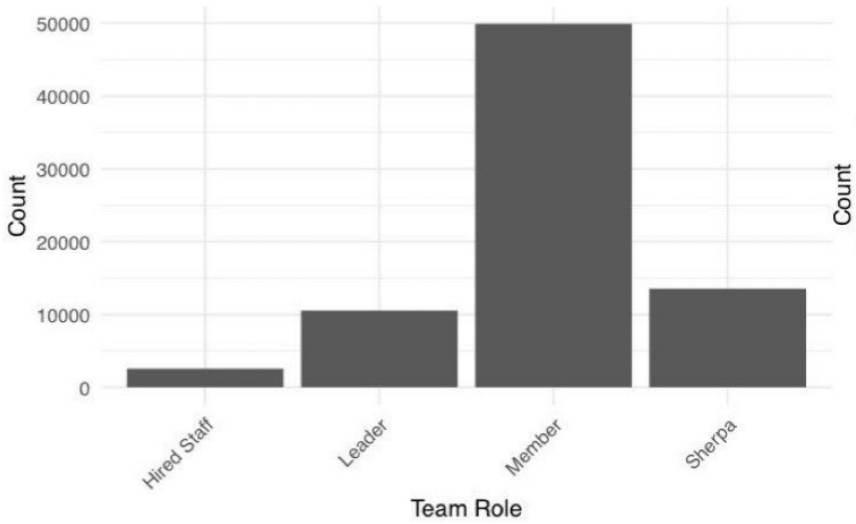
My dataset, derived from Cookson's, contains the entries from 1905 through Spring 2019.

The three datasets I considered included information about all peaks in Nepal, all expeditions on those peaks and all members on those expeditions. I then combine the data from these three datasets to form the main analysis dataset.

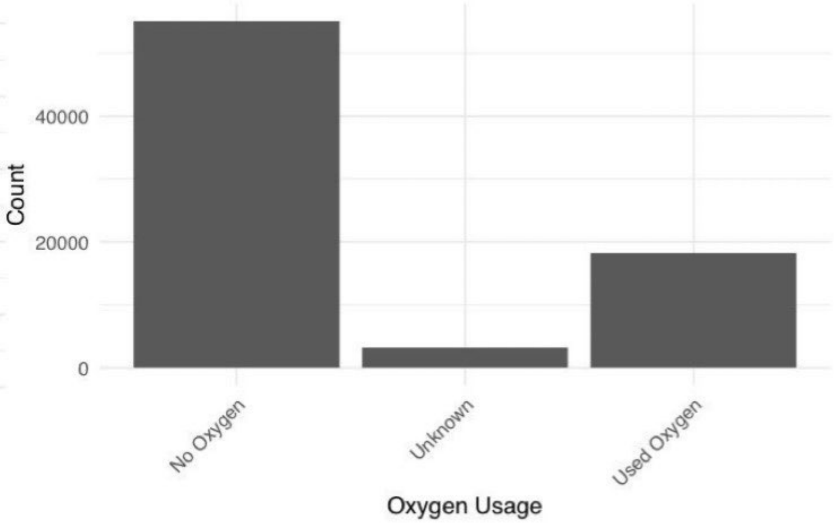
A person becomes an entry in my analysis dataset if, between 1905 and Spring 2019, they attempted to climb any one of the many Himalayan peaks in Nepal. It also included expeditions to both sides of border peaks as mentioned before.

Among the overall range of variables available, I chose the following to be included in the analysis dataset.

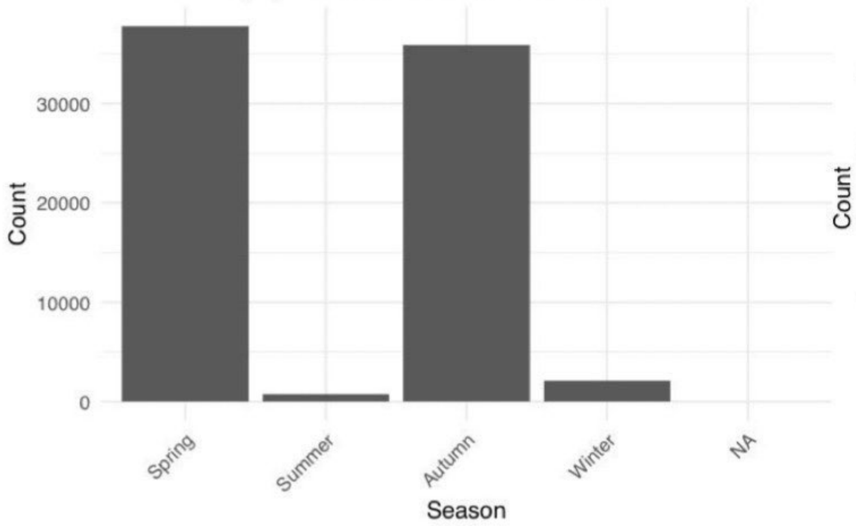
- `Height` is the height range in which the peak's height in metres falls. This is for the peak the person in the current entry is on an expedition for. The categories for this are 5400-6000m, 6000-7000m, 7000-8000m, and 8000m+.
- `Role` indicates the role of the person in the expedition. The categories are Leader, Member, Sherpa, and Hired Staff.
- `Oxygen` indicates the oxygen usage status during the expedition. This can be Used Oxygen, No Oxygen, or Unknown.
- `Season` is the season the expedition is embarked on. This takes on either of the four values: Spring, Summer, Autumn, Winter.
- `Success` indicates if the person's expedition resulted in a successful summit.
- `Death` indicates if the person died during the expedition.
- `TOTMEMBERS` indicates the total number of members in the expedition team.



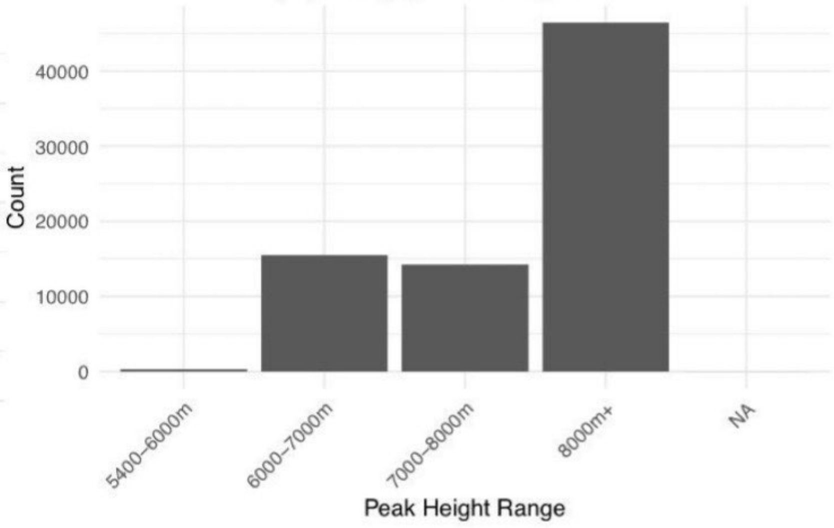
(a) Team Composition



(b) Oxygen Usage



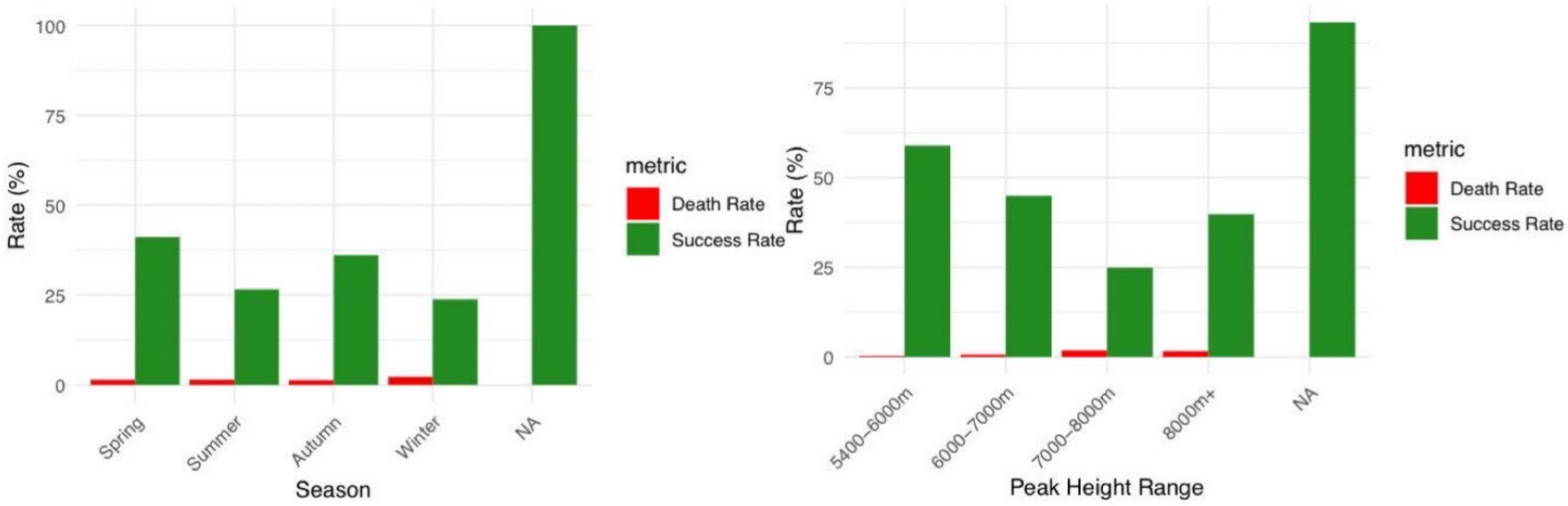
(c) Seasonal Distribution



(d) Peak Heights

Figure 1: Distribution of Key Variables in Himalayan Expeditions

From the distribution results in Figure 1, we can observe several salient features of Himalayan mountaineering activities. In terms of team composition, the number of ordinary members is the largest (about 50,000), followed by Sherpa (about 12,000), while the number of captains (Leader) and hired staff (Hired Staff) is relatively small. In terms of oxygen use, the number of climbers who choose to use oxygen is the largest, about 20,000, which shows that oxygen support is a common choice in high-altitude mountaineering; the number of people who do not use oxygen and whose usage is unknown is relatively small. Seasonal distribution shows that spring and autumn are the most popular climbing seasons, with the number of climbers in both seasons being around 30,000, while summer and winter have significantly fewer climbing activities, which may be related to the harsh weather conditions in these two seasons. . In terms of altitude distribution, most mountaineering activities are concentrated in the range of 7000-8000 meters and 6000-7000 meters, with about 15,000 and 18,000 people respectively, which shows that mid-altitude peaks are the most popular mountaineering targets, while those above 8000 meters There are relatively few mountaineering activities at extremely high altitudes and lower altitude peaks below 6,000 meters. This distribution may reflect the balancing considerations of mountaineering difficulty and risk.



(a) Success and Death Rates by Season

(b) Success and Death Rates by Height

Figure 2: Distribution of Key Variables in Himalayan Expeditions

Figure 2 shows the comparison of success rates and mortality rates of mountaineering activities in different dimensions. From a seasonal perspective, spring and autumn are not only the most popular climbing seasons, but also show the highest success rates, both close to 50%, while the mortality rate remains low (about 5%); in comparison, Although there is less mountaineering activity in winter, the mortality rate doubles and the success rate drops below about 25%, which may be directly related to the harsh winter weather conditions. Analyzed from the perspective of altitude, the data shows an obvious trend: as the altitude increases, the success rate of mountain climbing shows a decreasing trend, while the mortality rate gradually increases. Specifically, at the lower altitude of 5,400-6,000 meters, the success rate is as high as about 60%, and the mortality rate is the lowest among the four groups; when the altitude rises to about 8,000 meters, the success rate drops significantly to about 25%. At the same time, the mortality rate has increased many times compared to 5400-6000 meters. What is more interesting is that the success rate increases significantly above 8,000 meters, which reflects people's great enthusiasm for climbing the mountain. At the same time, the mortality rate is the same as that at 7,000-8,000 meters.

3 Model

I used a Bayesian Logistic Regression model to determine the likelihood of an individual successfully reaching the summit of the Himalayan peak they are attempting to climb during the expedition. Logistic regression is a statistical technique used for binary classification to predict the probability of a categorical dependent variable.

My model will be based on five independent variables: height, season, role_category, oxygen_status, and TOTMEMBERS, and the dependent variable will be success.

The logistic regression model I will be using is:

$$log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 \times \text{height} + \beta_2 \times \text{season} + \beta_3 \times \text{roleCategory} + \beta_4 \times \text{oxygenStatus} + \beta_5 \times \text{TOTMEMBERS}$$

$$\begin{aligned}
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5) \\
\beta_4 &\sim \text{Normal}(0, 2.5) \\
\beta_5 &\sim \text{Normal}(0, 2.5)
\end{aligned}$$

where,

- $\hat{p}$ represents the probability that someone will successfully summit the peak they are on the expedition for.
- β_0 represents the intercept term of this logistical regression. It is the probability that someone will successfully summit the peak they are on the expedition for if the predictors' values are zero
- β_1 is the coefficient corresponding to height of the peak
- β_2 is the coefficient corresponding to season of the expedition
- β_3 is the coefficients corresponding to role category of the person
- β_4 is the coefficients corresponding to oxygen usage status
- β_5 is the coefficients corresponding to total number of team members

In this model, both coefficients and intercept use normal prior distributions with zero mean and 2.5 standard deviation. I chose a zero mean for the priors because I wanted to start without any assumptions about the direction or size of the effects. The choice of 2.5 for standard deviation allows for considerable uncertainty in our prior beliefs, while still providing some regularization to prevent extreme estimates.

These weakly informative priors strike a balance between completely uninformative priors and strongly informative ones. They primarily let the data shape the posterior distributions while offering some protection against overfitting, which is particularly valuable given the complexity of mountaineering data. This approach is especially useful for our analysis because it helps stabilize estimates while still capturing the true patterns in expedition outcomes.

The technical details of model coefficients, including standard errors and diagnostic checks, are presented in detail in Section A. These diagnostics help validate our modeling choices and ensure reliable conclusions from our analysis.

4 Results

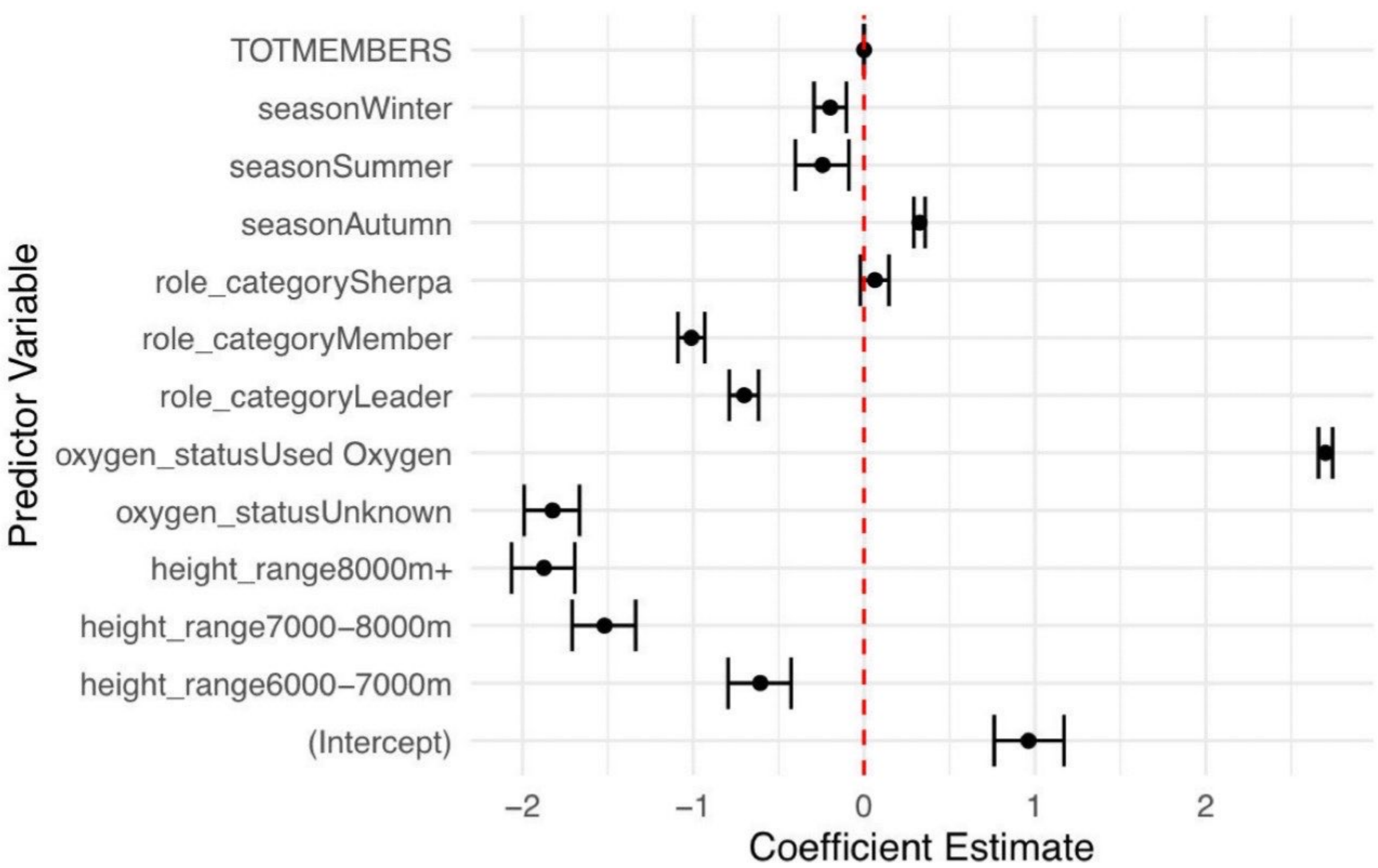


Figure 3: Coefficients of the Model

From the model coefficient analysis results shown in Figure 3, we can clearly observe the degree and direction of the impact of each factor on the success rate of mountain climbing.

The most significant positive effect comes from the use of oxygen (oxygen_statusUsed Oxygen), with a coefficient of about 2.70, indicating that using oxygen can significantly increase the probability of mountain climbing success; conversely, the unknown oxygen usage status (oxygen_statusUnknown) shows a significant negative effect , the coefficient is about -1.82.

In terms of altitude, as the altitude increases, the negative effect gradually strengthens, from -0.61 at 6000-7000 meters to -1.87 above 8000 meters, which reflects the significant inhibitory effect of high altitude on the success rate of mountaineering.

In terms of climbing seasons, autumn (seasonAutumn) shows a positive effect (coefficient 0.32), while summer and winter show a negative effect (-0.24 and -0.20, respectively).

In terms of player roles, compared to the baseline category, both the captain (role_categoryLeader) and the ordinary player (role_categoryMember) showed significant negative effects (-0.70 and -1.01 respectively), while the impact of Sherpa (role_categorySherpa) was relatively Smaller (0.06).

The effect of team size (TOTMEMBERS) is close to zero, indicating that team size may not be a critical factor in determining mountaineering success. The confidence intervals (error bars) of these results are generally narrow, indicating a high degree of precision in the estimates.

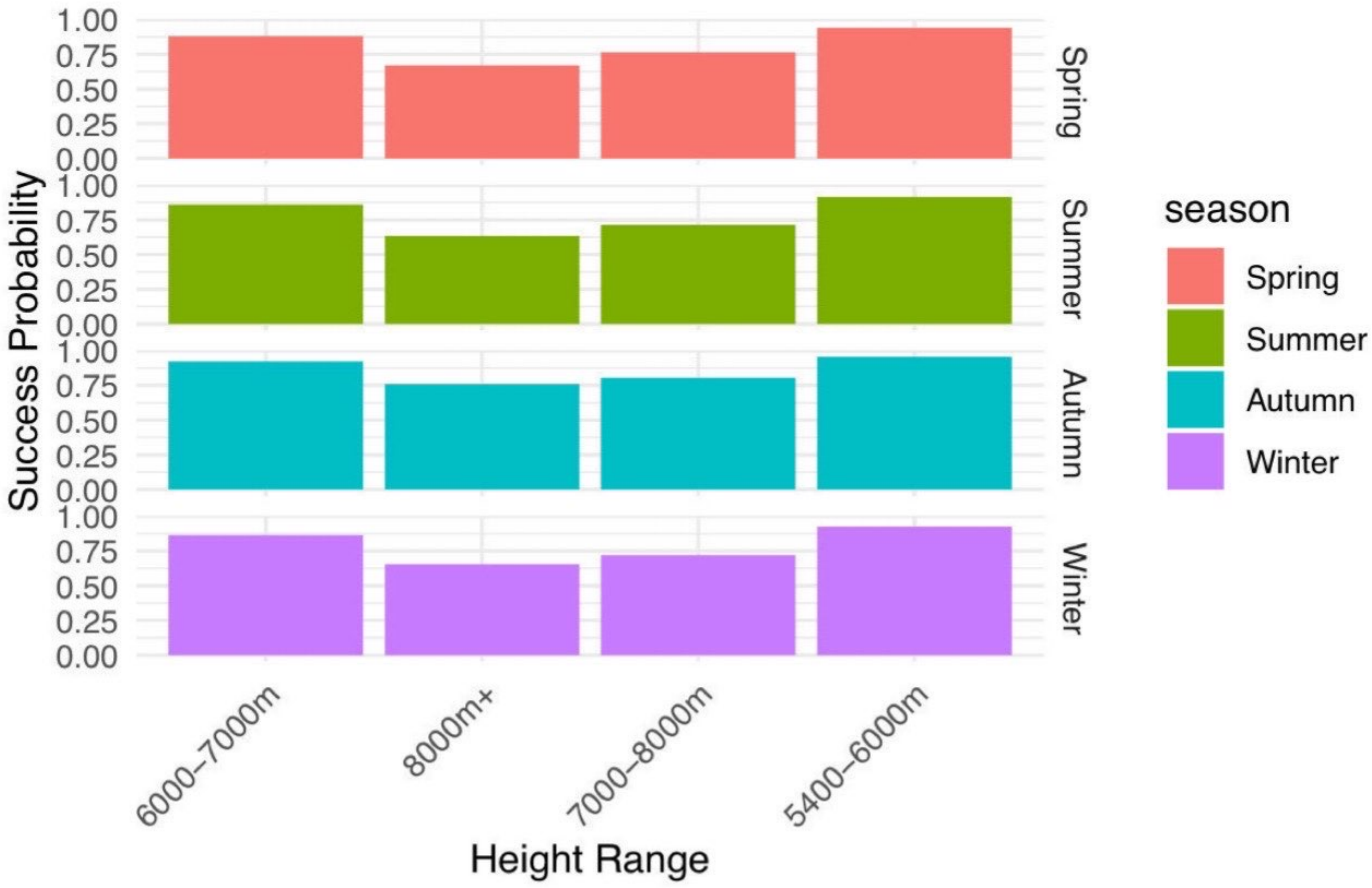


Figure 4: Predicted Success Probability by Height and Season

Figure 4 shows the prediction of mountaineering success probability under different altitude and season combinations. There is a general downward trend in success probability with increasing altitude in all seasons. Spring and autumn show a higher probability of success at all altitudes, especially on peaks below 6,000 meters, where the probability of success is close to 0.75. In contrast, the success probability at high altitudes in winter is significantly lower, only about 0.6. Summer performance is intermediate between spring and autumn and winter, but also shows a significant decrease in success at high altitudes. This pattern clearly demonstrates the interaction of season and altitude on mountaineering success.

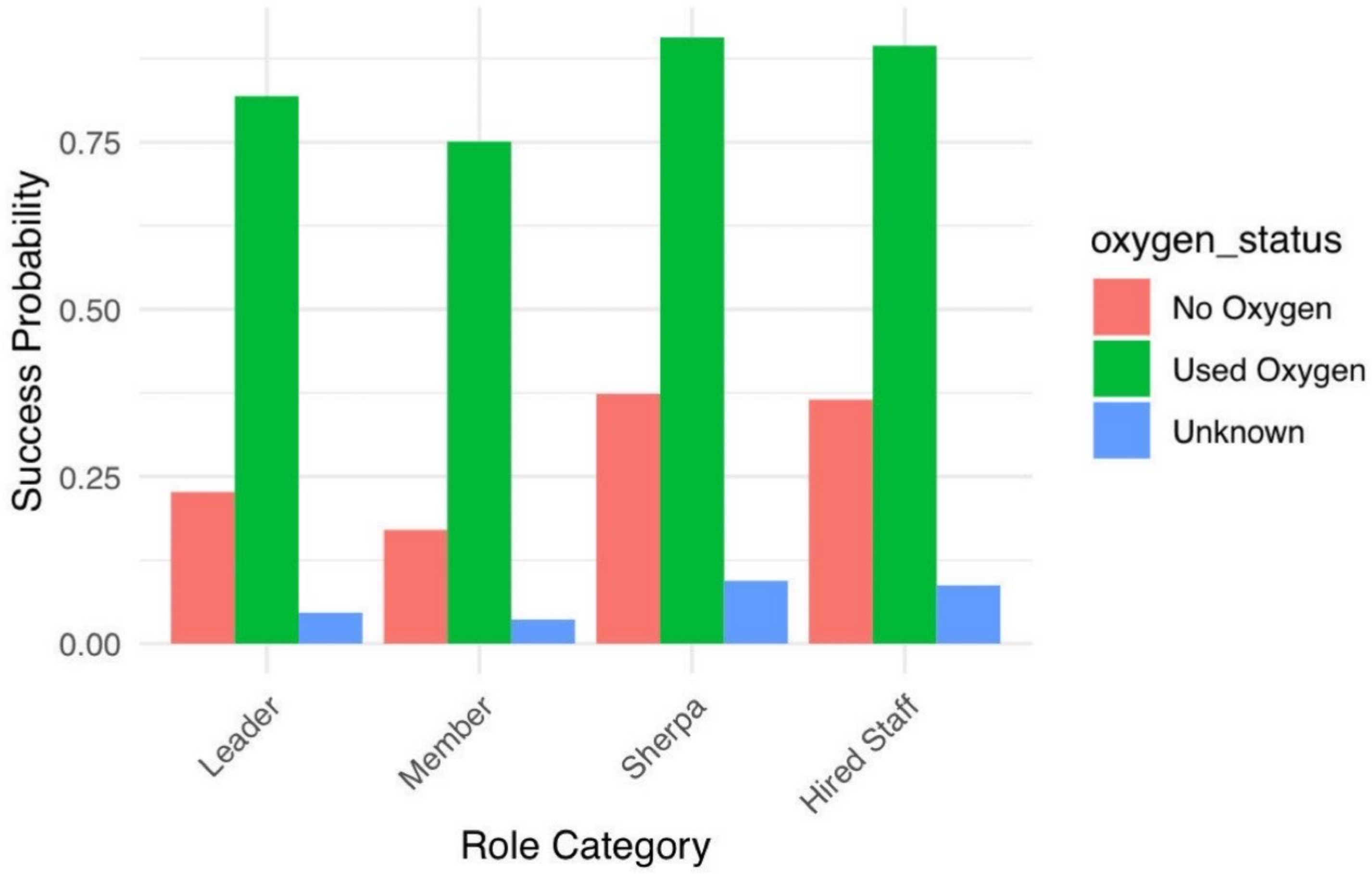


Figure 5: Predicted Success Probability by Role and Oxygen Usage

Figure 5 compares and analyzes the impact of different roles and oxygen usage on the probability of mountain climbing success. The data shows that across all character classes, using oxygen significantly increases the probability of success, by approximately 40-50 percentage points. Sherpas and hired personnel showed the highest probability of success (0.85) when using oxygen, while captains and ordinary players had a relatively low probability of success (0.75) even when using oxygen. For the case without oxygen, the probability of success is significantly lower for all characters, but the Sherpa and the Mercenary still maintain a relatively high success rate (0.40). These results highlight that oxygen use is a key factor in improving mountaineering success, but that its effect varies depending on the climber's role.

5 Discussion

5.1 The critical role of oxygen use

Of all the factors that influence success in Himalayan climbing, oxygen use exhibits the most significant positive impact. Data shows that climbers who use oxygen have a 40-50 percent higher success rate than climbers who don't. This significant difference is evident across all character classes, but is particularly pronounced in the Sherpa and Mercenary groups, who have an 85% success rate when using oxygen (Figure 5). This finding highlights the key role of oxygen supply in high-altitude mountaineering, which can not only improve the success rate of summiting, but also reduce the risk of climbing. This inspires us to consider the use of oxygen as an important safety measure when planning high-altitude mountaineering activities.

5.2 Role differences and professionalism

The analysis showed significant differences in climbing success rates between characters. Sherpas and hired personnel showed the highest success rates, reflecting the importance of professional climbers' experience and adaptability. In contrast, the success rates of captains and ordinary team members are relatively low. This difference may be due to several factors: the degree of professional training, ability to adapt to high altitudes, and familiarity with the local terrain. This reminds us that we should rely more on and value the experience and advice of professionals when organizing mountaineering activities (Figure 5).

5.3 Seasonal effects

Season selection has an important impact on the success of mountaineering activities. The data clearly shows that spring and autumn are the most suitable seasons for mountaineering. These two seasons not only have the largest number of climbers, but also have the highest success rate (nearly 60% in Figure 4). In contrast, although there is less mountaineering activity in winter, the mortality rate increases significantly to about 15%. This seasonal difference is mainly due to changes in weather conditions. The milder climate conditions in spring and autumn provide a better environment for mountaineering activities. This discovery has important guiding significance for the planning of mountaineering activities.

5.4 Altitude Challenge

Research results show that there is a significant negative correlation between altitude and mountaineering success rate. As the altitude increases, the success rate gradually drops from 70% at 5,400-6,000 meters to about 50% above 8,000 meters (Figure 4). At the same time, the mortality rate also increases with altitude, from 2% at low altitudes to nearly 10% at high altitudes. This trend reflects the tremendous challenges high-altitude environments pose to the human body, including the effects of thin oxygen, extreme weather conditions and other factors.

5.5 Research limitations

Although this study provides valuable insights, there are several limitations. First, there may be reporting bias in the data, particularly in earlier records. Second, our analysis failed to fully capture certain important factors, such as individual experience levels, specific weather conditions, etc. In addition, although team size (TOTMEMBERS) was included in the analysis, its impact was not significant, which may require more detailed group analysis.

5.6 Future research directions

Future research can be deepened in several aspects: 1) Explore the relationship between climbers' personal experience and success rate; 2) Study the impact of specific weather conditions on mountaineering activities; 3) Analyze the impact of choosing routes of different difficulty on the success rate; 4) Investigate the impact of advancements in equipment technology on mountaineering safety.

A Appendix

A.1 Cleaning

For the analysis data, the cleaning steps I took were:

1. Initial merging: The raw data from the expeditions and members datasets is merged based on common identifiers, `PEAKID` and `EXPID`, consolidating information about expedition participants.
2. Secondary merging: The combined dataset is merged with the peaks dataset based on `PEAKID`, to incorporate information about the height of each peak climbed during expeditions.
3. Height range categorization: Created height range categories based on the `HEIGHTM` variable: 5400-6000m, 6000-7000m, 7000-8000m, 8000m+.
4. Role categorization: Classified expedition members into four categories based on their roles: Leader (where `LEADER` = TRUE), Sherpa (where `SHERPA` = TRUE), Hired Staff (where `HIRED` = TRUE), Member (all others).
5. Oxygen usage classification: Created oxygen usage categories based on `MO2USED` and `MO2NONE`: Used Oxygen (where MO2USED = TRUE), No Oxygen (where MO2NONE = TRUE), Unknown (all other cases).
6. Season standardization: Mapped `season` values from numeric codes (1,2,3,4) to descriptive categories: Spring, Summer, Autumn, Winter.
7. Data cleaning:

- Removed rows with missing values in key variables
- Filtered out inconsistent or impossible values
- Standardized variable formats

8. Final dataset creation: Selected and renamed relevant columns for the analysis:

- `height_range`
- `role_category`
- `oxygen_status`
- `season`
- `success` (from `MSUCCESS`)
- `death` (from `DEATH`)
- `TOTMEMBERS`

A.2 Analysis dataset

Table 1: Analysis dataset

MSUCCESS	height_range	season	role_category	oxygen_status	TOTMEMBERS
FALSE	6000-7000m	Autumn	Leader	No Oxygen	8
FALSE	6000-7000m	Autumn	Member	No Oxygen	8
FALSE	6000-7000m	Autumn	Member	No Oxygen	8
FALSE	6000-7000m	Autumn	Member	No Oxygen	8
FALSE	6000-7000m	Autumn	Member	No Oxygen	8
FALSE	6000-7000m	Autumn	Member	No Oxygen	8

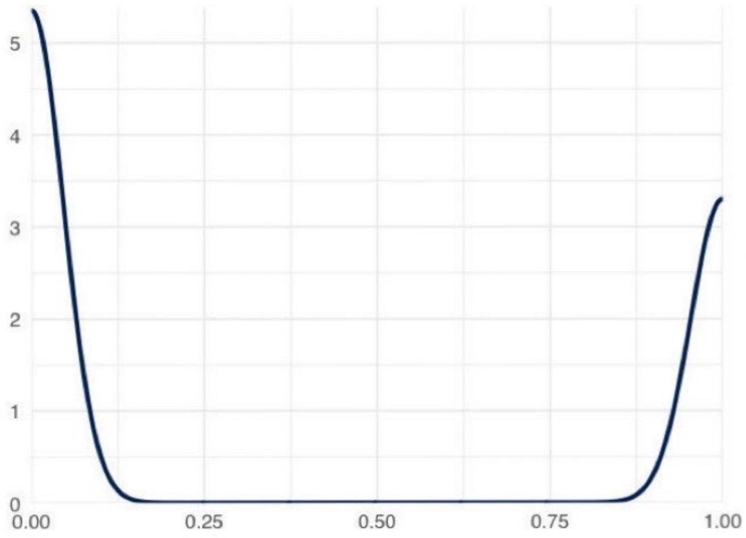
Table 1 provides a glimpse of the dataset used for analysis.

A.3 Model summary

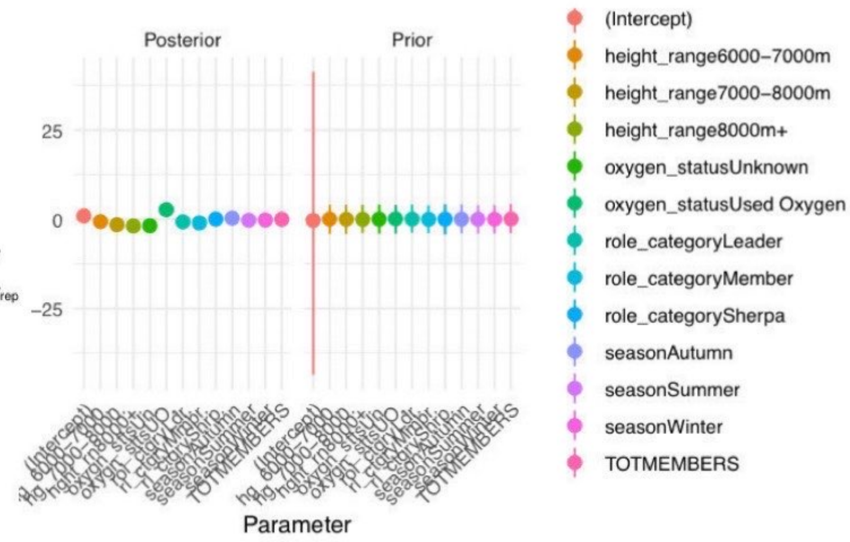
Table 2: Summary of the model

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.96	0.13	0.76	1.17
height_range6000-7000m	-0.61	0.12	-0.80	-0.43
height_range7000-8000m	-1.52	0.12	-1.71	-1.34
height_range8000m+	-1.87	0.12	-2.06	-1.69
seasonSummer	-0.24	0.09	-0.40	-0.09
seasonAutumn	0.32	0.02	0.29	0.36
seasonWinter	-0.20	0.06	-0.29	-0.10
role_categoryLeader	-0.70	0.05	-0.79	-0.62
role_categoryMember	-1.01	0.05	-1.09	-0.93
role_categorySherpa	0.06	0.05	-0.02	0.14
oxygen_statusUnknown	-1.82	0.10	-1.99	-1.67
oxygen_statusUsed Oxygen	2.70	0.02	2.66	2.74
TOTMEMBERS	0.00	0.00	0.00	0.00

Table 2 shows the coefficients for my Bayesian model along with the standard error and the 95% credible interval. The standard error (SE) is a measure of the precision with which a sample statistic estimates a population parameter. It quantifies the variability of sample statistics around the population parameter.



(a) Posterior Predictive Check



(b) Prior vs Posterior

Figure 6: Model Diagnostics

A.4 Posterior predictive check

Figure 6 are used to understand assess the quality of my model.

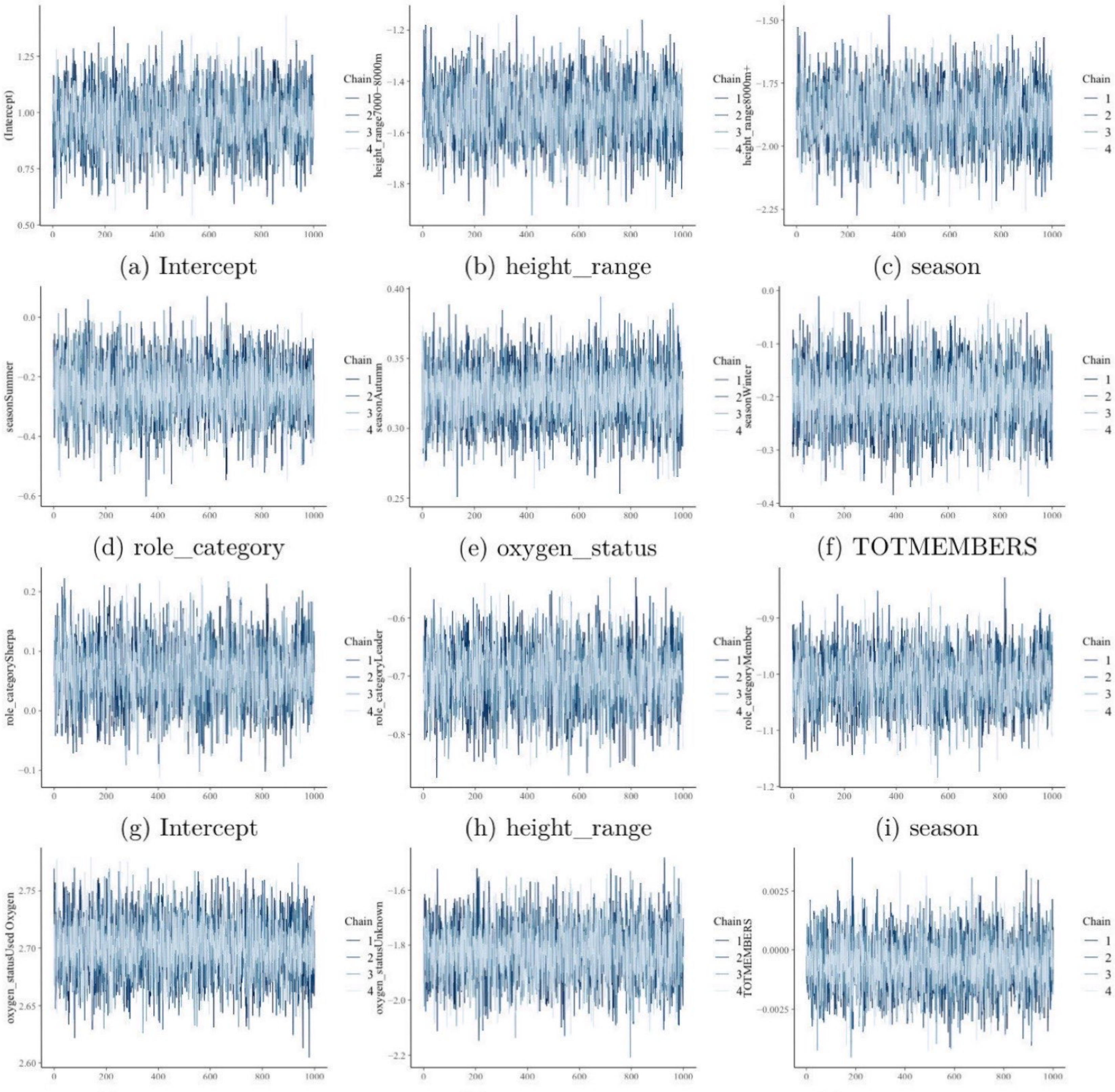
`Posterior Predictive Check` is the result of a posterior predictive check and is used to compare the actual outcome variable with simulations from the posterior distribution. `Prior vs Posterior` is the result of comparing the posterior with the prior to see how much the estimates change once data are taken into account. Both these results support the model doing a good job of fitting the data.

A.5 Markov chain Monte Carlo (MCMC) Convergence Check

The model was fitted using the rstanarm (Goodrich et al. 2024) package, using the MCMC algorithm to obtain samples from the posterior distribution of interest.

Figure 7 shows the chain tracing plots for different parameters, including the intercept term, the coefficient of altitude range, the seasonal effect, the role category, the oxygen usage status, and the group size. It can be observed from the figure that the four chains for all parameters show good mixing, indicating that the MCMC sampling fully explored the parameter space. The chain trajectories show appropriate random fluctuations without obvious trends or mutations, which indicates that the sampling has reached a stable state.

Figure 8 shows the Rhat statistic, and the Rhat values of all parameters are close to 1.00 and do not exceed 1.05, further confirming the good convergence of the model. Together, these diagnostic results show that our MCMC sampling is reliable and the model estimates have good stability. It is particularly noteworthy that the convergence of the chains is well maintained even for more complex parameters, which strengthens our confidence in the reliability of the model results.



(a) Intercept (b) height_range (c) season

(d) role_category (e) oxygen_status (f) TOTMEMBERS

(g) Intercept (h) height_range (i) season

(j) role_category (k) oxygen_status (l) TOTMEMBERS

Figure 7: Trace plot

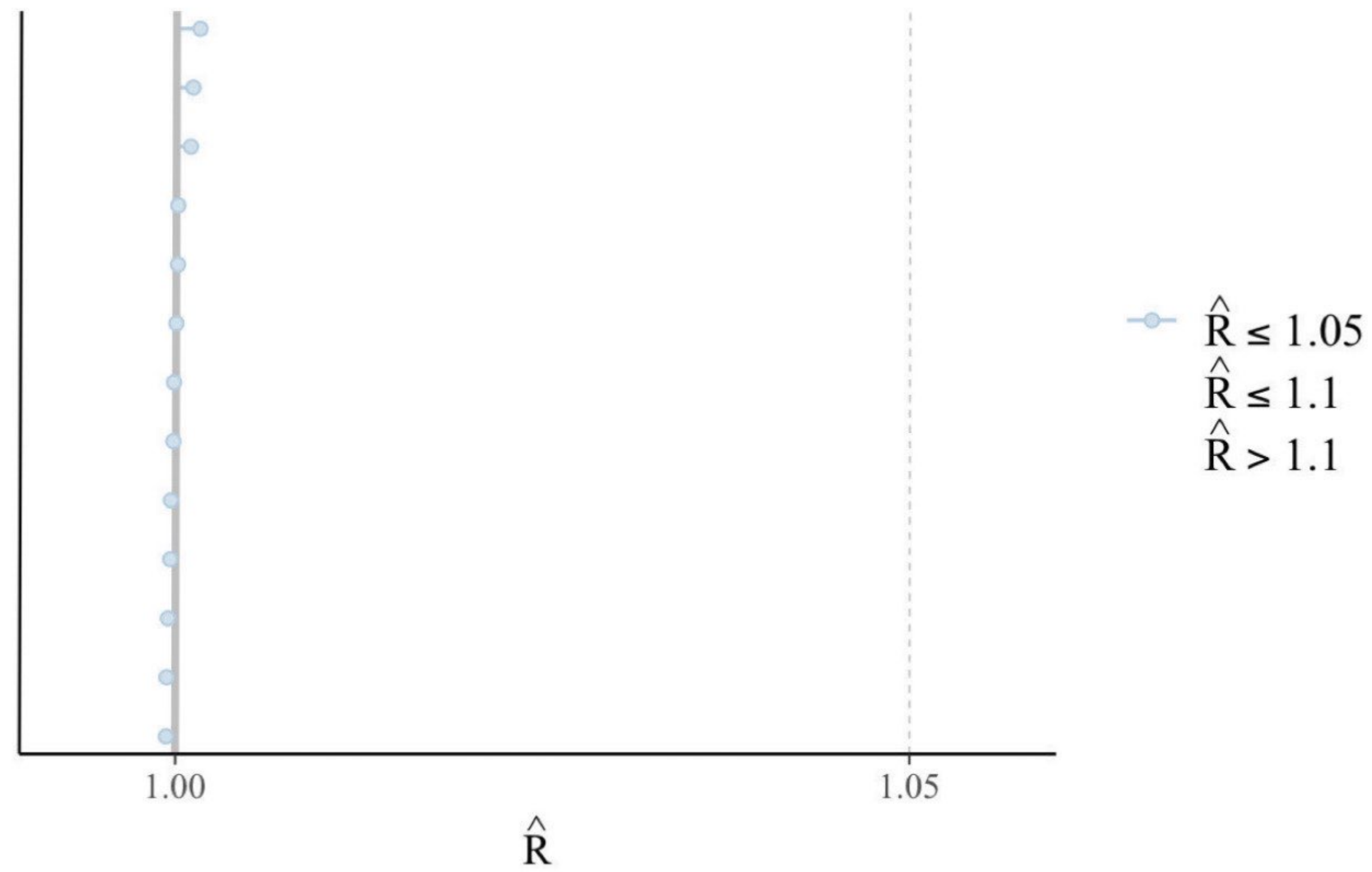


Figure 8: Rhat plot

References

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models.* https://CRAN.R-project.org/package=broom.mixed.

Cookson, Alex. 2020. "Data/Himalayan-Expeditions at Master · Tacookson/Data." *GitHub.* https://github.com/tacookson/data/tree/master/himalayan-expeditions.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://here.r-lib.org/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Salisbury, Richard. 2023. "The Himalayan Database, the Expedition Archives of Elizabeth Hawley." *The Himalayan Database, The Expedition Archives of Elizabeth Hawley.* https://www.himalayandatabase.com/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.