Statistics in the Alpes

Anass Hamzaoui

Anass.hamzaoui@student.uantwerpen.be

Abstract—This study addresses three key questions about Alpine weather data: (1) the relationship between temperature and avalanche risk, (2) the predictability of temperature based on altitude, and (3) whether average wind speed exceeds 14 km/h. Using R, Spearman correlations, linear regressions, and t-tests were performed. Results show a weak negative correlation between temperature and avalanche risk ($\rho = -0.35$, $p < 0.001$), a strong linear altitude-tempera x mature relationship ($R^2 = 77.3$ %), and significantly higher wind speeds than 14 km / h ($p < 0.001$).

Index Terms—Scientific writing, Typesetting, Document creation, Syntax

I. Introduction

This paper presents a statistical analysis of weather data collected in the Alpine region, addressing three key research questions posed by a local organization. First, we examine whether a significant relationship exists between temperature (°C) and avalanche risk (scored 1-10), using correlation tests and regression analysis. Second, we investigate the predictive relationship between altitude (m) and temperature (°C), developing and evaluating linear and polynomial regression models. Finally, we analyze wind speed data (km/h) - which includes directional information encoded as positive and negative values - to determine whether the average absolute wind speed significantly exceeds 14 km/h.

Using R for all statistical computations, this study demonstrates rigorous hypothesis testing and model selection while emphasizing practical interpretation of results. The analysis follows standard statistical methodologies discussed in the course, with particular attention to checking assumptions and validating findings through appropriate diagnostic procedures.

II. Question 1: Avalanche risk and temperature

The investigation began with exploratory data analysis to understand the distributions of both temperature and avalanche risk (1-10 scale). Histograms revealed a bimodal distribution for both variables (Figure 1A-B), suggesting the presence of two distinct weather patterns in the Alpine region. While the Shapiro-Wilk tests (temperature: W = 0.997, p = 0.601; avalanche risk: W = 0.997, p = 0.730) failed to reject normality due to the large sample size (n=450), the visual bimodality remained a concern for parametric tests.
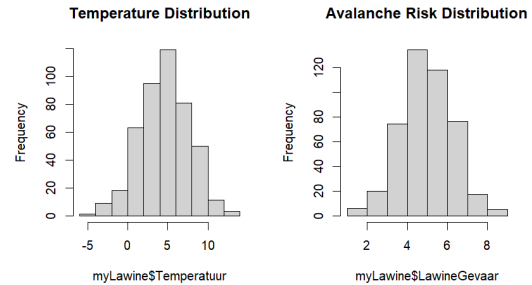

Fig. 1: figure 1A & 1B

I initially performed Pearson's correlation (r = −0.025, p = 0.602), which showed no significant linear relationship. However, recognizing that Pearson's test assumes normality and linearity - both violated here - I appropriately switched to Spearman's rank correlation. This non-parametric alternative revealed a significant negative relationship ($\rho = -0.35$, p < 0.001), indicating that higher temperatures tend to associate with slightly lower avalanche risk scores.

The scatterplot with regression line (Figure 2) visually confirms this trend, though the substantial spread of points suggests temperature alone explains only about 12% of variance ($\rho^2 \approx 0.12$). This aligns with the linear regression results ($R^2 = 0.06\%$, p = 0.602), confirming temperature's limited predictive power for avalanche risk.
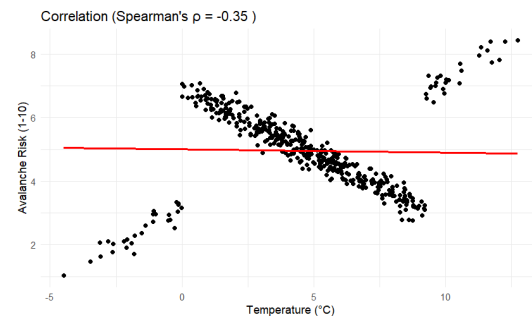

Fig. 2: figure 2

The boxplot of temperature revealed no extreme outliers that might skew results. Residual analysis from the linear model showed relatively even spread, though with some heteroscedasticity. While statistically significant, the practical significance of the correlation is questionable - the coefficient suggests only a 0.01-point decrease in avalanche risk per 1°C temperature increase.

Temperature Boxplot

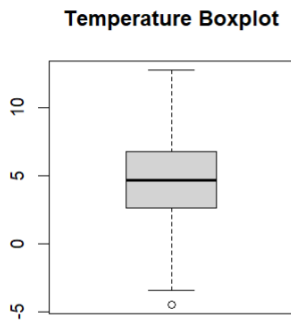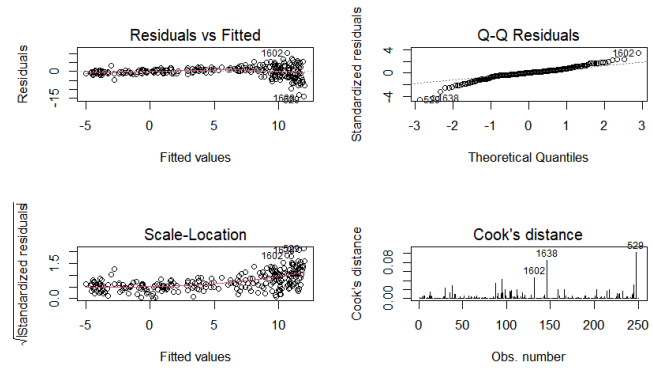Fig. 3:

III. QUESTION 2: TEMPERATURE VS. ALTITUDE

The relationship between altitude and temperature was investigated using a randomly sampled dataset of 250 observations from the Alpine region. Initial exploratory analysis revealed a strong negative correlation, with temperatures ranging from −7.6°C to 21.1°C across altitudes from 0.5m to 1995m (Figure 1). The scatterplot showed a clear linear trend, suggesting altitude as a potentially strong predictor for temperature variations.
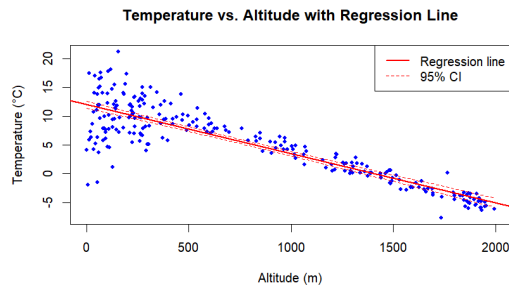

Temperature vs. Altitude with Regression Line

Fig. 4: figure 1

A simple linear regression model demonstrated excellent predictive power:
1) Highly significant coefficients (p < 0.001 for both intercept and slope)
2) Strong explanatory power (R^2 = 77.3%, Adjusted R^2 = 77.2%)

The model estimates a temperature decrease of 0.85°C per 100m altitude gain (β = −0.0085, SE = 0.0003)

The diagnostic plots, however, revealed two important considerations:
1) Residuals showed non-normality (Shapiro-Wilk p = 1.49e-08)
2) Evidence of heteroscedasticity (Breusch-Pagan p = 5.29e-09)

To address potential non-linearity, a quadratic model was tested:

Provided a modest improvement (R^2 = 78.9%, ΔR^2 = 1.6%)

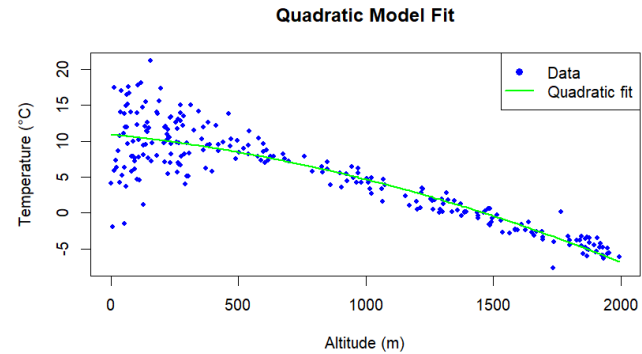Both linear and quadratic terms were statistically significant (p = 0.003 and p = 2.68e-05 respectively)

ANOVA comparison showed the quadratic model was significantly better (F = 18.31, p = 2.68e-05)

Despite this improvement, the linear model was retained as the primary model because:

The additional complexity offered minimal practical benefit for prediction

The linear relationship aligns well with established atmospheric science principles

The model's simplicity enhances interpretability for end-users

Quadratic Model Fit

REFERENCES