THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

A Statistical Analysis of Influential Factors in UTMB Ultra Marathon Racing

Kyla Mazzotta

DATS 4001: Data Science Capstone

Dr. Sushovan Majhi

December 12, 2025

1. Introduction

Ultra marathons are generally defined as races in which a distance greater than 26.2 miles (mi) (or 42.2 kilometers) has been completed. While typically a set distance, races may also be measured by distance accomplished in a set amount of time. The Ultra Trail du Mont Blanc (UTMB) World Series is a series of trail races ranging in length from 21 to 170+ kilometers (km) and taking place around the world. These races are completed with the hopes of earning tokens to qualify for the UTMB World Series Finals. While marathons tend to be completed primarily on paved courses with a standardized distance and a focus on speed, ultra marathons are completed with the primary goal of finishing the event and a secondary focus on speed.

Previous research in the field of ultra marathon running primarily focused on races of specific distances and looked at influential factors on finish times at an athlete-level (Turnwald et al., 2025). Specific factors looked at included but were not limited to athlete nationality, age, gender, experience, and training volume (Knechtle & Nikolaidis, 2018). In addition to athlete-specific factors, course variables such as location, distance, duration, and terrain were a main focus as well (Costa et al., 2019). The majority of ultra marathon finishers were found to be middle-aged, well educated white men (Cejka et al., 2013). While marathons tend to be dominated by African racers due to there being large economic incentives, such incentives tend to not be as present in ultra marathons, therefore resulting in less cultural diversity in ultra racing compared to marathons. Due to this lack of monetary incentive to perform well in ultra marathons, personal and social motivations tend to be the main reason for participation (Cejka et al., 2013).

Since the majority of research published regarding ultra marathons focuses on main effects on finish times for races of set distances, this study seeks to quantify how distance, elevation gain, and participant demographics such as age and gender interact to influence average finish times of US UTMB races using multiple linear regression with interaction and higher order terms.

2. Data Description

The data used for this report came from a webscraped dataset made publicly available containing results and demographics from UTMB races between 2014 and 2023. With data for 38,460 races, it was decided that only races in the United States would be looked at, thus bringing the race total to 3,927 races. Since the UTMB World Series contains races varying in distance, to account for only races in the ultra marathon category, only races with distances greater than or equal to 50 km were included in the analysis.

Distances and elevation gains for each race were mean-centered to allow for them to be used as continuous variables in regressions. While initially distance was to be treated as a categorical variable, due to the variability of distance in ultra marathons, clean race distance distinctions could not be made, therefore resulting in its use as a continuous variable. A mean-centering approach was used for both distance and elevation to reduce multicollinearity, increase model stability, and make interpretations of coefficients meaningful to quantify deviations from the mean. The baselines for distance and elevation gain were therefore calculated to be 85.06 km and 2564.22 meters (m) respectively. To complete the mean-centering process, the average distance and elevation gain were subtracted from the distance and elevation gain of each race, thus making the coefficients interpreted at the average race distance and elevation gain in future regressions.

To further narrow the scope of races in attempts to reduce variability and control for smaller races that may be more sensitive to extreme times, a cutoff of 85 participants was determined, meaning that only races with 85 or more participants were included in the analysis. This cutoff was determined by looking at the 50th percentile of the number of participants in US races. Since the 50th percentile was determined to be 84, a race size considered to be sufficiently large in the world of racing, it was used as a cutoff value for determining the minimum number of participants for races used in this analysis.
To account for the varying number of participants in each race when quantifying the demographics of participants, percentages were used to measure the proportion of individuals in each age group and women. The original data broke up age groups into five-year increments. Since there is such little variability between age groups so close together and small in size, age groups were consolidated into three larger categories: under 35 years, 35-59 years, and 60 and older. These represented significantly different life stages while also ensuring that there were enough individuals in each age group to be able to make statistically significant comparisons. Percentages were then calculated by dividing the number of athletes in each age group by the total number of participants of each race. The same calculation was done to obtain the percentage of women participants in each race without the age group component.

The last major element of the UTMB dataset is the way in which race finish times were measured and analyzed. Since this analysis looks at ultra marathons on a race-level as opposed to an athlete-level, the main outcome being studied is the mean race finish time measured in hours (hrs).

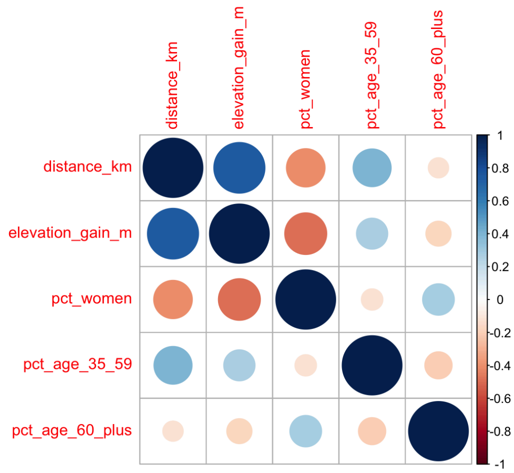
Sample of UTMB Race-Level Data

First 5 Rows — Centered Predictors Included

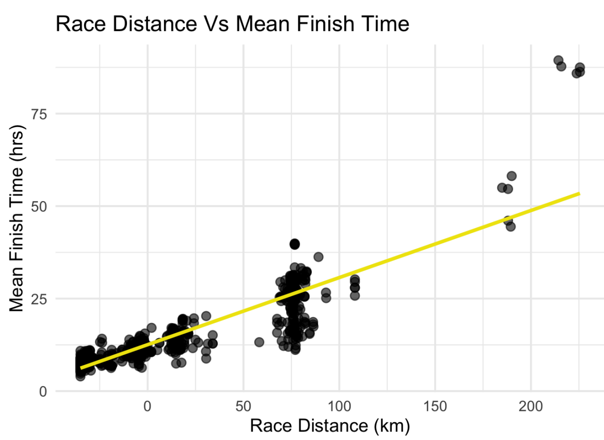
		Centered Predictors		Participant Demographics (%)			
Race ID	Mean Finish Time (hrs)	Distance (km, centered)	Elevation Gain (m, centered)	% Women	Under 35	35–59	60+
10129.2018	8.258698	−35.0581	185.7766	0.3333	0.3750	0.6250	0.0000
10129.2019	8.553171	−35.0581	185.7766	0.4161	0.3758	0.5839	0.0403
10129.2020	7.849280	−35.0581	185.7766	0.2845	0.4310	0.5431	0.0259
1014.2014	10.187414	−4.0581	−964.2234	0.2664	0.2587	0.5792	0.1622
1014.2015	10.186978	−4.0581	−964.2234	0.2316	0.2316	0.7158	0.0526

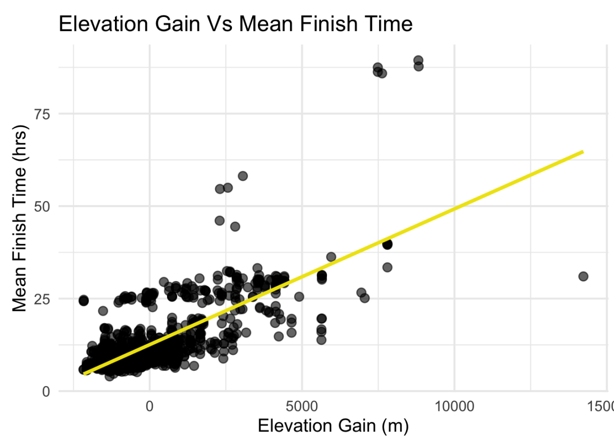
3. Methods

Before preliminary regressions could be run to determine which factors are most important when determining the mean time to finish a race, a correlation plot was created to visualize potential relationships between factors and determine which interactions would be worth including when building models for regression analysis. From the correlation matrix, with the Under 35 age group removed to avoid perfect multicollinearity, it appears that the terms with the highest correlations were Elevation Gain with Distance, Distance with Percentage of Women, and Elevation Gain with Percentage of Women, therefore when creating a model with interaction terms, these interactions will be included to test their significance.



In addition to a correlation matrix, scatter plots were created to model the relationship between Average Finish Time and Distance and Elevation Gain to gain a preliminary understanding of the relationship between the most influential factors according to previous research.





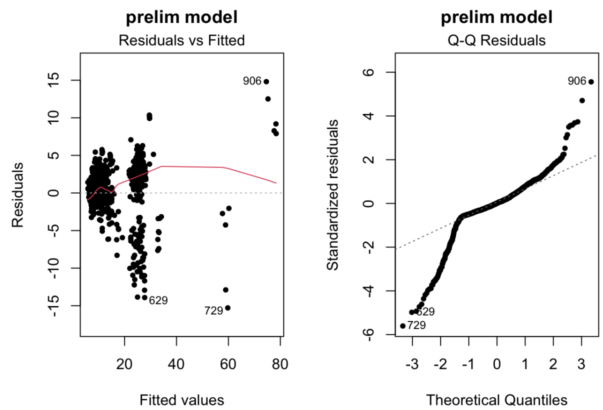
The plot of Race Distance against Mean Finish Time displayed a positive relationship that indicated a slightly nonlinear relationship as distance increased. Such a relationship led to a higher order term of $Distance^2$ being considered as being meaningful in later regressions. The Elevation Gain versus Mean Finish Time plot, on the other hand, aside from showing some potential outliers, communicated a positive relationship appearing to be roughly linear, therefore not leading to the inclusion of a higher order elevation coefficient.

To establish which factors are most significant in explaining the average time to complete an ultra marathon, multiple linear regression was conducted on several models to determine the model of best fit. Starting with a preliminary parsimonious model including only main effects and eventually running stepwise regression from both sides, using AIC-based selection, on a full model containing higher order and interaction terms for Distance^2, Elevation × Distance, Distance × Percentage of Women, and Elevation × Percentage of Women, regressions were run and analyzed to determine which models best explained variations in average race finish times.

Models were chosen based on improvement in adjusted R^2 values and model significance using ANOVA F-tests; with the final selected model being used to interpret how predictors relate to mean finish time, rather than to generate predictions.

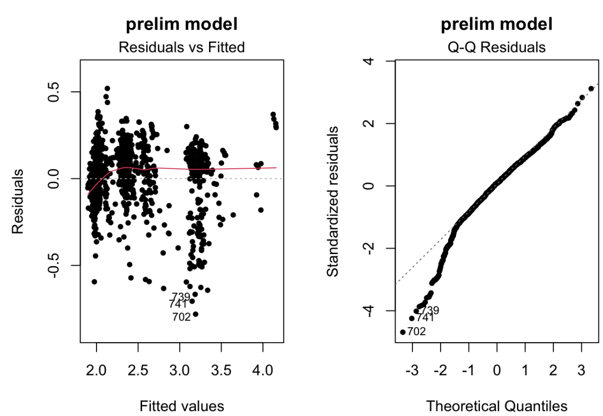
4. Results

4.1 Initial Model (Untransformed)



The first regression model created included main effects for Distance, Elevation Gain, Percentage of Women, Percentage of Ages 35 - 59, Percentage of Ages 60+, and the higher order term Distance^2. After running the regression, while the model had an adjusted R^2 of 0.8978, the only significant predictors at a 5% significance level were Distance, Distance^2, and Elevation Gain, therefore indicating that the height adjusted R^2 may be a result of overfitting. A residual analysis was then performed to determine whether the necessary assumptions were met and determine overall model adequacy. Residual plots showed slight signs of nonnormality. While some variations were expected due to the high variability of ultra marathon times in general, moving forward it was decided that a model using log(Mean Finish Time) as the response variable would help to improve the model and better satisfy the necessary model assumptions.

4.2 Log-Transformed Model



Since the residuals for the previous model showed signs of nonnormality, the model was run with the log of Mean Finish Time as the response variable, and included all factors in the previous model with the addition of interaction terms for Elevation × Distance, Distance × Percentage of Women, and Elevation × Percentage of Women. This new regression model showed that the interaction terms for Distance × Percentage of Women and Elevation × Percentage of Women were significant at the 5% significance level, and the percentage of participants 60 and older was significant at the 10% significance level.

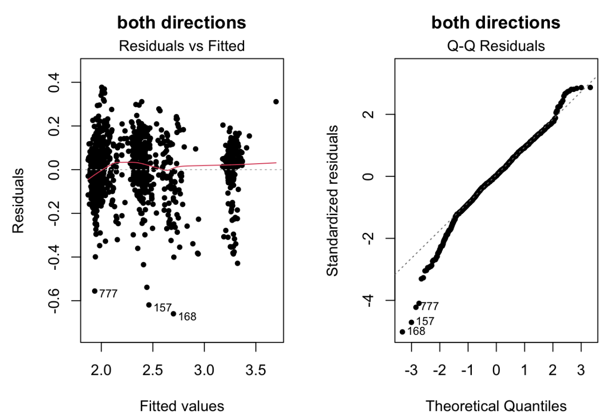
While the interaction terms involving the percentage of women were significant to the model, the percentage of women by itself was not. This model produced an adjusted R^2 value of 0.8896 which is lower than the previous, less complicated model, indicating that this model is not significantly better than the previous at explaining mean finish times. While not a significant improvement, a residual analysis shows that log-transforming the response variable did help to meet model assumptions.

4.3 Stepwise-Selected Model

Due to the full model not being a significant improvement from the simple model, stepwise regression from both directions using AIC-based selection was used to eliminate insignificant factors from the model. The resulting model was one containing Distance, Distance^2, Elevation Gain, Percentage of Women, Percentage Ages 60+, Elevation × Distance, Distance × Percentage of Women, and Elevation × Percentage of Women. This model has an adjusted R^2 of 0.8896 which is very similar to the one produced by the previous model. Even though the adjusted R^2 are similar, this model is considered to be an improvement as the same percentage of variation in log(Mean Finish Time) is being explained but with fewer predictors.

Before selecting this model as the model of best fit, a check for outliers was performed using a Cook's distance criteria of 0.0033, calculated by dividing 4 by the number of observations (1195). This method flagged 73 observations as potential outliers. These outliers were removed and the stepwise regression was run again to check for significant model improvements.

The stepwise regression with outliers removed yielded a model with an adjusted R^2 of 0.9221 which included the predictors: Distance, Distance^2, Elevation, Percentage of Women, Elevation × Distance, and Elevation × Percentage of Women. The higher adjusted R^2 along with the removal of two factors indicates model improvement with the removal of outliers, therefore making this model the final model. Residual analysis on the model showed that the residuals approximate normality with minor deviation which can be accounted for due to the variable nature of ultra marathons as discussed previously.



In addition to a residual analysis performed on the final model, a multicollinearity check was also conducted by checking variance inflation factors (VIFs). All main effects had VIFs below 1, suggesting minimal multicollinearity. Mean-centering Distance and Elevation helped achieve this by reducing correlations among predictors. Thus, the final stepwise-selected model is robust and interpretable, capturing meaningful variation in log(Mean Finish Time) while accounting for key predictors and interactions.

5. Discussion

This statistical analysis of US UTMB races indicates that Distance and Elevation are the strongest predictors of mean finish time, a finding that is consistent with prior research regarding course difficulty in ultra marathons. The inclusion of the quadratic term, $Distance^2$, in the final model is indicative of a slight nonlinear relationship between Distance and Mean Finish Time, meaning that longer races increase mean finish time at faster rate. Interaction effects for Elevation × Distance and Elevation × Percentage of Women were also significant in the final model. The interaction Elevation × Distance was associated with a negative coefficient, showing that elevation becomes less impactful in longer races and has a greater impact on mean finish time in ultra marathons on the shorter end of the distance spectrum. While the Percentage of Women participants was not significant as a main effect, combining it with Elevation in the form of an interaction term changed that. The negative coefficient associated with the Elevation × Percentage of Women term indicates that elevation gain has less of an effect on Mean Finish Time in races with more women, potentially signifying that women may handle tackling large elevations better than men. While demographic compositions alone were not as significant predictions of Mean Finish Time as course characteristics such as elevation and distance these findings suggest that they do have an influence through interactions.

Age group compositions were not found to be significant once Distance and Elevation were accounted for. While the Percentage of 60 and Older factor was included in the stepwise model, once potential outliers were removed, it was no longer significant to the model.

Model diagnostics indicate that the log transformation of Mean Finish Time improved residual normality, while the removal of outliers enhanced model stability and adjusted R^2. Multicollinearity was minimal, aided by the mean-centering of Distance and Elevation, and residual plots suggest that model assumptions are reasonably satisfied despite minor deviations expected due to the inherent variability of race finish times in ultra-running.

When it comes to applying these results in the world of ultra-running and running in general, these findings can be used by runners, coaches, and race directors, providing insight into how race characteristics and athlete demographics interact to influence mean finish times.

Limitations of this analysis include the use of race-level aggregated data, which prevents causal inference about individual athlete performance, and the reliance on percentages for demographic variables, which may mask individual variability. Future research could incorporate athlete-level data or explore additional course-specific factors such as terrain type, weather conditions, or altitude profiles. It would also be beneficial to follow an athlete across different races to be able to compare results in a more standardized way.

6. Conclusion

The final model for explaining Mean Finish Time was found to be:

$$log(Mean\ Finish\ Time) = Distance + Distance2 + Elevation + \%\ Women + (Distance \times Elevation) + (Elevation \times \%\ Women)$$

The interaction terms reveal that the impact of elevation depends on race distance and demographics of participants with regard to gender. The final model highlights that average race finish time is not a simple function of distance and elevation, but is rather a complex equation with many intricacies involving both course and participant demographics.

This study determined Distance and Elevation Gain, along with their interactions, to be the primary determinants of mean finish time in US UTMB races, while participant demographics such as age and gender play a smaller role, primarily through interaction effects. The final stepwise-selected, outlier-removed model is robust, interpretable, and explains meaningful variation in log(Mean Finish Time), providing clear insight into how race characteristics and participant composition influence performance.

These findings can inform ultra runners, coaches, and race organizers about expected race difficulty and performance trends, and offer a foundation for future studies incorporating more detailed course-level variables such as weather patterns and course conditions.

References

About the UTMB World Series. UTMB World Series - Meet your extraordinary! (n.d.). https://utmb.world/sports-system

Cejka, N., Rüst, C. A., Lepers, R., Onywera, V., Rosemann, T., & Knechtle, B. (2013). Participation and performance trends in 100-km ultra-marathons worldwide. *Journal of Sports Sciences*, *32*(4), 354–366. https://doi.org/10.1080/02640414.2013.825729

Costa, R. J. S., Knechtle, B., Tarnopolsky, M., & Hoffman, M. D. (2019). Nutrition for ultramarathon running: Trail, track, and road. *International Journal of Sport Nutrition and Exercise Metabolism*, *29*(2), 130–140. https://doi.org/10.1123/ijsnem.2018-0255

Knechtle, B., & Nikolaidis, P. T. (2018). Physiology and pathophysiology in Ultra-Marathon running. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.00634

Partyka, A., & Waśkiewicz, Z. (2024). Motivation of marathon and ultra-marathon runners. A narrative review. *Psychology Research and Behavior Management*, *Volume 17*, 2519–2531. https://doi.org/10.2147/prbm.s464053

Poirot. (2024, May 29). *UTMB World Race Data*. Kaggle. https://www.kaggle.com/datasets/mgpoirot/utmb-world-race-daa/data

Spittler, J., & Oberle, L. (2019). Current trends in ultramarathon running. *Current Sports Medicine Reports*, *18*(11), 387–393. https://doi.org/10.1249/jsr.0000000000000654

Turnwald, J., Valero, D., Forte, P., Weiss, K., Villiger, E., Thuany, M., Scheer, V., Wilhelm, M., Andrade, M., Cuk, I., Nikolaidis, P. T., & Knechtle, B. (2025). Analysis of the 50-mile ultramarathon distance using a predictive XGBoost model. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-92581-w