Bring the Heat: Exploring the Relationship between Temperature and Short & Long-Distance Running Performance

Abstract:

Race-day weather substantially influences performance in outdoor running. However, a knowledge gap exists on how temperature is differentially associated with short- and long-distance running performance. Using a dataset provided by Konstantinos Mantzios and colleagues that examines finishing times and weather conditions from the world's largest running events, we fit a linear regression model with interactions to investigate whether the association with temperature and running performance depends on the race being a 5k or a marathon. To establish a quantifiable measure of running performance comparable across races, we created a novel metric *world record deviation* ($\frac{\text{world record time} - \text{first place time}}{\text{world record time}}$). Our analysis revealed a significant association between increasing air temperature and greater declines in running performance (as evaluated by *world record deviation*) for marathon runners compared to 5K runners. This analysis introduces a novel methodological approach to comparing running performance and highlights the need for race-specific training strategies.

Background

Environmental conditions play a crucial role in athletic performance, particularly in outdoor running events. For instance, several studies have found that increased temperature is associated with slower marathon times (Ely et al., 2007; Trubee et al., 2014; Vihma 2009). Furthermore, temperature effects are exacerbated by high humidity, as humidity reduces evaporative heat loss and may lead to dehydration (Bongers et al., 2017). Wind speed also plays a role in the relationship between temperature and running performance, as strong headwinds may amplify the negative effects of high temperature, whereas tailwinds may provide a performance benefit (Galloway et al., 1997).

While there is abundant research on temperature effects on long distance running performance, there is a knowledge gap regarding how temperature is differentially associated with short and long-distance running performance. Indeed, very few studies have examined temperature in relation to shorter running distances, such as the 5k or the mile. Further investigation into the associations between temperature and running performance, and how these relationships differ based on race distance, could provide valuable insights for athletes, coaches, and race organizers. These insights could inform the development of better race-specific training to optimize performance and reduce the risk of injury. Thus, this study aims to address the question: **Does the relationship between air temperature and running performance depend on race length for elite runners?**

To explore this question, we use a dataset provided by Konstantinos Mantzios, Leonidas G. Ioannou, Andreas Flouris that examines the finishing times and weather conditions during some of the world's largest running events from 1952-2019. We decided to focus on data from 5ks and marathons as these races are very popular, recognized as official Olympic events, and take place outdoors. Furthermore, these distances are useful in examining the physical demands of different lengths of running, with 5Ks requiring on average 84% aerobic and 16% anaerobic energy contributions and marathons requiring an average 97.5% aerobic and 2.5% anaerobic energy contributions ("Aerobic", n.d.). We hypothesize the existence of a statistically significant interaction between air temperature and race length in predicting running performance, in particular expecting that higher air temperature will be associated with worse running performances in marathon races as opposed to 5k races.

Methodology

To fairly analyze running performance across both 5k and marathon times, we defined the *world record deviation* as the difference between the first place time and the world record time (in minutes), divided by the world record time (in minutes): $\frac{\text{world record time} - \text{first place time}}{\text{world record time}}$. This variable is designed such that increases correspond to improved running performance, and decreases correspond to worsening running performance. Using this as our outcome variable, we fit a multiple linear regression model with an interaction term between *air temperature* and *race type*. We additionally controlled for the continuous variables *air temperature, wind speed, relative humidity*, and *year,* and categorical variables *sex, race type,* and *competition*.

Our choice of weather-based controls (*air temperature, wind speed,* and *relative humidity)* was informed by existing literature on ideal outdoor running conditions. Since we were wary of collinearity when adding weather variables, we chose not to add weather data such as solar radiation or time of day. Outside of weather variables, we considered other factors that could lead to different race times. In the end, we chose to control for sex, year, race type, and competition. We decided to control for competition since the proficiency and effort of the runners may change across the competition, sex since men often run faster than women, and year since running ability has improved over time.

We removed race results missing first-place time data and those that did not disaggregate results by gender. To test whether there was sufficient statistical evidence against the null hypothesis that regression slopes equaled 0, we used a t-test with 202 degrees of freedom and at a significance level of $\alpha = 0.05$.

Results

Our final sample contained 214 races after removing 31 observations for missing first-place time data and 714 observations for results not disaggregated by gender. There were 148 5k race results and 66 marathon race results. The mean air temperature across all races was 13.9°C, with the hottest temperature being 33.4°C, and the coldest temperature being -5°C. All linear model assumptions appear to be reasonably satisfied (**Appendix A**). Regression results are presented in the table below:

Table 1: Estimated regression coefficients for linear model

Model Term	**Estimate**	**Standard Error**	**P-Value**
Intercept	1.5238	0.2845	<0.0001
Sex			
Men	(Ref.)		
Women	0.0017	0.0029	0.5670
Year	-0.0008	0.0001	<0.0001
Temperature	-0.0001	0.0004	0.7297
Wind Speed	0.0011	0.0017	0.5228
Relative Humidity	-0.0001	0.0001	0.4725
Race Type			
5k	(Ref.)		
Marathon	0.0271	0.0126	0.0332
Competition			
Commonwealth	(Ref.)		
Diamond League	0.0321	0.0053	0.0000
Olympics	0.0173	0.0053	0.0013
World Championships	0.0153	0.0049	0.0019
World Cup	-0.0242	0.0081	0.0030
Temperature*Race Type *(Marathon)*	-0.0019	0.0006	0.0009

Given the p-value of 0.0009 for our interaction term, which is less than our 0.05 significance level, there is sufficient evidence to suggest that the association between temperature and deviation from the world record depends on the race type. For 5ks, for each 1°C increase in air temperature, the first-place percentage of the world record time is expected to decrease by 0.0001397%, controlling for all other variables in the model. For example, if the 5k world record was 12 minutes and 30 seconds, then each 1°C increase in air temperature would be associated the first-place runner being 0.001 seconds slower than the world record. On the other hand, for marathons, for each 1°C increase in air temperature, the first-place percentage of the world record time is expected to decrease by 0.002008%. For example, if the marathon world record was 2 hours and 10 minutes, then 1°C increase in air temperature would be associated with the first-place runner being 0.157 seconds slower than the world record.

Discussion

Our analysis suggests evidence that the associations between temperature and running performance differ between marathon and 5K runners. Specifically, we found that increasing air temperature was associated with greater decreases in running ability (quantified by *world record deviation*) for marathon runners compared to 5K runners. Though the differences in performance may initially appear small, these results still have real-world significance considering that world records are often beat by milliseconds for both races. This indicates that a 1°C increase in temperature may be associated with breaking or not breaking a world record. This finding contributes to the body of literature that has shown that longer distance runners are vulnerable to the negative effects of high temperatures through providing a comparison point against shorter distance running. Furthermore, this analysis introduces a novel statistic based on individual performance relative to race-specific record-setting paces that allows researchers to directly compare running performance across different types of race lengths.

Our results suggest that race organizers and coaches should pay close attention to the weather conditions and provide appropriate cooling strategies for marathon runners to prevent excessive heat stress and improve performance. Moreover, the findings of this study have broader implications for understanding the complex relationship between environmental factors and athletic performance and highlight the need for further research in this area to develop evidence-based guidelines for athletes competing in a range of weather conditions.

Limitations

One major limitation in our methodology is that the dataset does not provide runner IDs for first place winners, making it possible that the independence assumption is betrayed: if a runner won first place across multiple races, that data would no longer be independent. However, considering our relatively large sample size and the low chance that the same person won a very large number of the races across these competitions, we assume that this assumption is reasonably satisfied.

Another limitation in our methodology is that we use a complete case analysis to deal with missingness. Given the administrative reasons for removing observations, we believe that missingness completely at random is reasonable to assume, leading only to a loss of power without biasing our regression results. An additional limitation in this study is that it is observational, and thus we cannot directly conclude that hotter temperatures result in differing race outcomes. Future studies using randomized control trials would be useful to determine causality.

Conclusion

In conclusion, we find that the relationship between temperature and running performance does depend on the race being a marathon or a 5k for elite runners.

Works Cited

Aerobic Vs Anaerobic Training, Compared. (n.d.). Retrieved April 19, 2023, from https://marathonhandbook.com/aerobic-vs-anaerobic-training/

Bongers, C. C. W. G., Hopman, M. T. E., & Eijsvogels, T. M. H. (2017). Cooling interventions for athletes: An overview of effectiveness, physiological mechanisms, and practical considerations. *Temperature (Austin, Tex.)*, *4*(1), 60–78. https://doi.org/10.1080/23328940.2016.1277003

Ely, M. R., Cheuvront, S. N., Roberts, W. O., & Montain, S. J. (2007). Impact of weather on marathon-running performance. *Medicine and Science in Sports and Exercise*, *39*(3), 487–493. https://doi.org/10.1249/mss.0b013e31802d3aba

Galloway, S. D., & Maughan, R. J. (1997). Effects of ambient temperature on the capacity to perform prolonged cycle exercise in man. *Medicine and Science in Sports and Exercise*, *29*(9), 1240–1249. https://doi.org/10.1097/00005768-199709000-00018

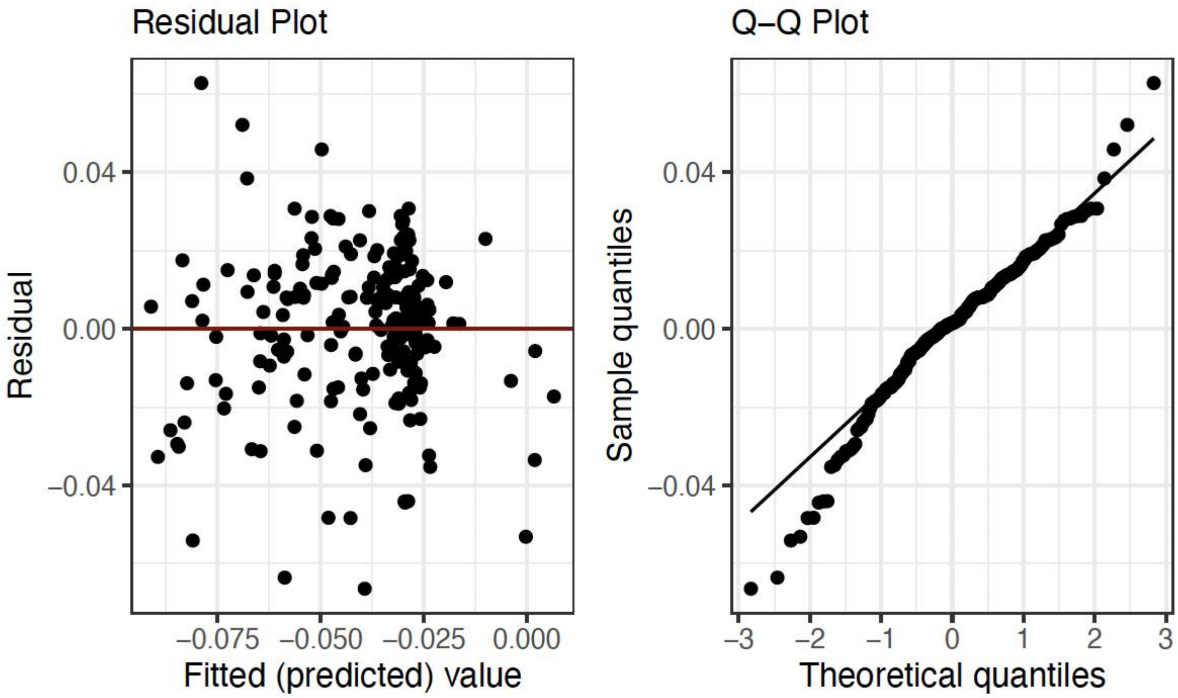
Mantzios, Konstantinos, Ioannou, Leonidas, & Flouris, Andreas. (2021). *Effects of weather parameters on endurance running performance* (p. 713587 Bytes) [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.14753565.V1

Trubee, N. W., Vanderburgh, P. M., Diestelkamp, W. S., & Jackson, K. J. (2014). Effects of Heat Stress and Sex on Pacing in Marathon Runners. *The Journal of Strength & Conditioning Research*, *28*(6), 1673. https://doi.org/10.1519/JSC.0000000000000295

Vihma, T. (2010). Effects of weather on the performance of marathon runners. *International Journal of Biometeorology*, *54*(3), 297–306. https://doi.org/10.1007/s00484-009-0280-x

Appendix A. Linear Model Assumptions

Figure A1. Residual and Q-Q Plot for Linear Model



Independence: *satisfied*

- The assumption of independence in this study was met as each race in the dataset were not related to one another. However, see the limitations section for potential violations of this assumption.

Linearity: *satisfied*

- In our residual plot, the observations are symmetrically distributed around the horizontal axis. Thus, we find that this regression model is linear in the parameters.

Constant Variance: *satisfied*

- To satisfy the constant variance condition, the observations should be evenly spaced along the y-axis in our residual plot. There is a clump of observations around fitted values -0.025 and a slight fan out shape. However, the plot is relatively consistent otherwise. Thus, we find that the variance of the errors is constant.

Normality: *satisfied*

- There is minimal deviation in our Q-Q plot. Thus, we conclude that the normality condition is satisfied.