Exploring Strava's Suffer Score

Saryu Onishi*

April, 2023

*Department of Health and Exercise Sciences, University of Oklahoma. E-mail address: saryu@ou.edu

1 Introduction

Strava is a popular social media platform many athletes use to share their most recent athletic endeavours. These can range from walks around the block to snowboarding down a mountain. Even activities like pickleball can be uploaded onto Strava. In addition to the sharing, liking, and commenting on fellow users' uploads, athletes can use Strava to analyze their workouts. This is why Strava is particularly popular amongst endurance athletes. The analysis goes beyond the typical activity summaries GPS watches provide at the end of a workout. Although many of the workout analysis features can be accessed on the free version of Strava, some other features require users to be a paid member.

Strava's "suffer score" one of these feature that is only accessible to paid members. The suffer score is built to be a performance metric that represents how tough an activity was, in a single number. In exercise physiology, metrics like these are called measures of training load Bourdon et al. (2017). According to Strava (2016), their suffer score is calculated based on time spent in various heart rate (HR) 'zones'. The time spent in a higher HR zone is weighed heavier, compared to the time spent in a lower HR zone.

This project aims to estimate the Strava Suffer Score training load metric using simpler activity summary metrics to cater to new users who are curious or existing users who prefer not to pay a subscription fee.

2 Literature Review

The goal of a training load measure is to capture the physiological impact of a workout session. More lit on topics related to the different types of measures (internal, external, TRIMPS, etc).

- Bourdon et al. (2017)
- Lambert and Borresen (2010)
- Roos et al. (2013)

3 Data

3.1 Data Collection

Data for this project was scraped from the activity log of a single Strava user, using the Strava API. The data was loaded into R studio software for further analysis using the rStrava R package.

Data was only collected from one user for a few reasons:

- Limitations of API - authentication required to access user data.
- Variability of heart rate responses across individuals.
- Ample data can be obtained from a single user.

3.2 Data Cleaning

The activity log was filtered to only include metrics such as time, distance, average heart rate and average speed. This was done because the goal of the project was to estimate the Suffer Score based on simple summary metrics such as these. Table 1 contains a descriptive statistics of the selected metrics. The data was also filtered to only include runs, as physiological responses to cycling and running (the two main types of activities in this data set) can differ significantly Hassmén (1990).

4 Methods

The primary model can be depicted in the following equation:

$$Y_r = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \varepsilon, \tag{1}$$

where Y is Suffer Score for recorded runs (r), and X are summary statistics. The parameter of interest are the β_n.

The backwards stepwise regression method was used to include only the significant independent variables. This was done in R studio by using ...

The models were then tested for (assumptions) and (fit).

- figures for plot(model) - linearity, homoscedasticity etc

5 Research Findings

The main results are reported in Table 2.

6 Conclusion

The results showed:

- the limitations of using simple statistics.
- non-linear pattern.

For future research, machine learning could be integrated to generate a model that may estimate Suffer Score better.

References

Bourdon, Pitre C, Marco Cardinale, Andrew Murray, Paul Gastin, Michael Kellmann, Matthew C Varley, Tim J Gabbett, Aaron J Coutts, Darren J Burgess, Warren Gregson et al. 2017. "Monitoring athlete training loads: consensus statement." *International journal of sports physiology and performance* 12 (s2):S2–161.

Hassmén, Peter. 1990. "Perceptual and physiological responses to cycling and running in groups of trained and untrained subjects." *European journal of applied physiology and occupational physiology* 60:445–451.

Lambert, Michael Ian and Jill Borresen. 2010. "Measuring training load in sports." *International journal of sports physiology and performance* 5 (3):406–411.

Roos, Lilian, Wolfgang Taube, Monika Brandt, Louis Heyer, and Thomas Wyss. 2013. "Monitoring of daily training load and training load responses in endurance sports: what do coaches want?" *Schweizerische Zeitschrift für Sportmedizin und Sporttraumatologie* 61 (4):30–36.

Strava. 2016. "Suffer Score: How Hard is Hard?" `https://blog.strava.com/suffer-score-how-hard-is-hard-11775/`. Accessed: 2023-04-24.

Figures and Tables

Table 1: Statistics for Strava Run Data

Metric (units)	Mean	Std. Dev.	Min	Max
Suffer Score	4.127	31.38	7.00	175.00
Distance (km)	12.19	3.76	5.01	23.51
Elapsed Time (min)	63.64	20.83	17.95	144.72
Elevation Gain (m)	59.43	60.74	0.00	303.10
Average Heart Rate (bpm)	147.4	11.30	116.60	184.00
Max Heart Rate (bpm)	170.2	170.18	136.00	204.00
Average Cadence (spm)	88.01	1.78	83.70	95.20
Average Speed (kph)	12.31	1.37	8.35	16.74

Notes: Sample size for all variables is $N = 306$.

Table 2: Model Summaries

	All Variables	Backwards Stepwise Regression
Distance	-1.977***	-0.536**
	(0.219)	(0.214)
Time	xx.xx	xx.xxx
	(0.219)	(0.214)
Average Heart Rate	xx.xxx	xx.xxx
	(0.219)	(0.214)
...	xx.xxx	xx.xxx
	(x.xxx)	(x.xxx)
R value	0.71	0.73

Notes: actual values will be inserted later..

Figure 1: Performance of Regression Model