A Look into Ultra Running

Introduction

- Ultra marathons are considered to be running races that, distance-wise, are longer than a marathon (26.2 miles or 42.2 km)
- Ultra marathons can be measured in either time or distance
- Typical distances are 50km, 50 miles, 100km, and 100 miles

- There has been extensive research done regarding the impacts of age, elevation, and gender on finish times, however this study looks at whether there are significant interactions between such factors

Impacts

- Provide athletes with realistic finish-time predictions
- Provide race coordinators with accurate cut-off times and insight into aid-station needs
- Provide coaches with expectations regarding different elevations, ages, and genders to tailor plans
- Provide researchers with regression and predictive models with only a small set of key factors

Objectives

Sought to answer the following questions:

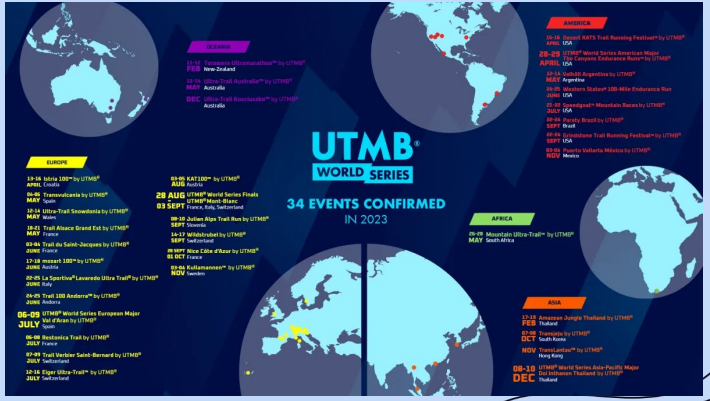
- Which factors contribute the most to average race finish times?
- Is elevation gain a significant factor? Does it interact with any others?
- As elevation increases, does the effect of gender percentage on race time change?
- Average race times, especially with ultra marathons exhibit large amounts of variability, this study aims to determine whether a large portion of variability can be explained by a few key factors

Web-Scraped UTMB Data

The UTMB (Ultra-Trail du Mont-Blanc) World Series is a series of ultra trail running races held around the world with results posted publicly

After some minor difficulties with datasets, eventually I happened upon a dataset containing web scraped data from the UTMB website containing information on:

- Age (percentage of finishers in each age group)
- Gender (measured in percentage of women finishers)
- Elevation gain (measured in meters)



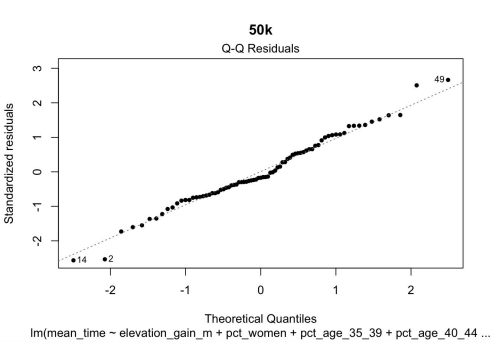
50k Regression - the basic model (Adj. R^2 =0.24)

The basic model

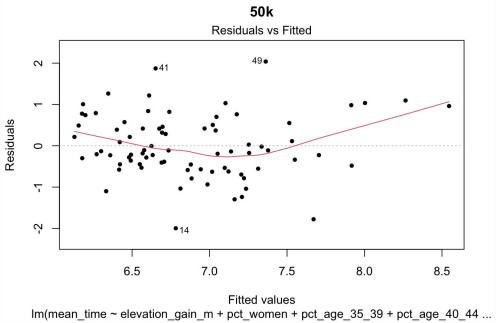
Average race finish time = elevation gain + percentage of women finishers + each age group

Elevation gain was found to be the only significant factor

Note: due to singularities some age groups returned coefficients of NA



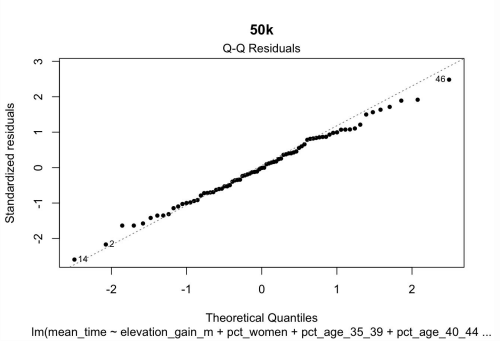
Looking at main effects of elevation, gender, and age group



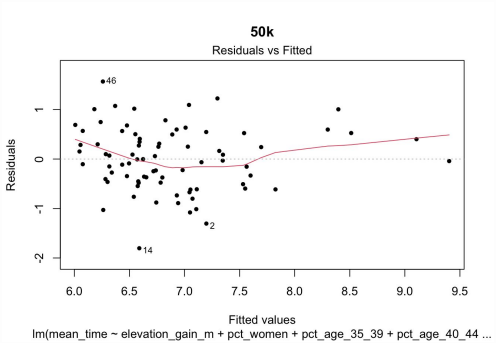
50k Regression – the full model (Adj. R^2 =0.38)

The full model

Average race finish time = elevation gain + percentage of women finishers + elevation gain*percentage of women + elevation gain*each age group



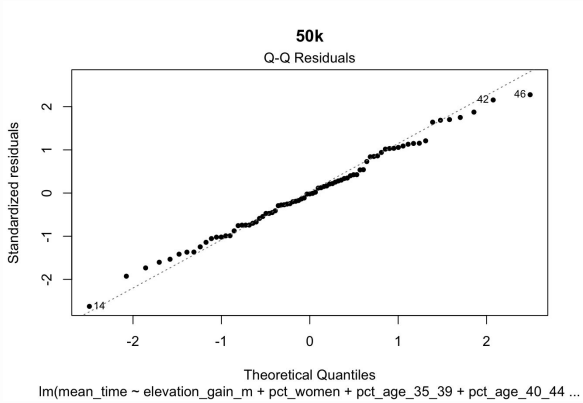
Looking at main effects of elevation, gender, and age group and including interactions with elevation gain

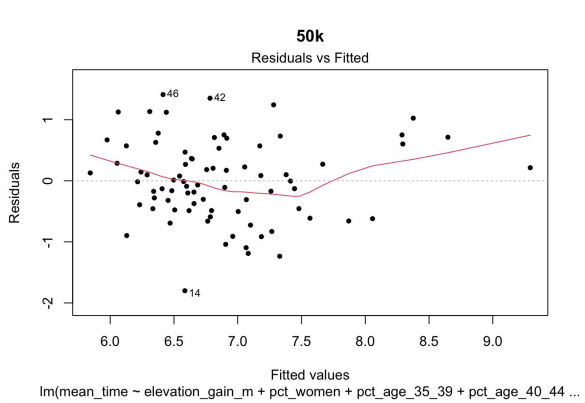


Stepwise Regression - Backwards Elimination (Adj. R^2 = 0.42)

The reduced model

Average race finish time = elevation gain + percentage of women finishers + percentage age 35-39 + percentage age 40-44 + percentage age 55-59 + elevation gain*percentage of women + elevation gain*percentage age 40-44 + percentage of women* percentage age 35-39





What this means

Main effects:

- Races with higher % women have longer finish times
- Races with more participants ages 40-44 are faster
- Races with higher elevation gains are slower on average, but only slightly as an isolated factor

Interactions:

- Elevation gain slows runners less in races where there are higher % women (faster average finish times)
- Elevation gain slows runners more when there is a higher % of 40-44 year olds
- Interaction between women and 35-39 year olds should be looked into more

	GVIF	Df	GVIF^(1/(2*Df))
elevation_gain_m	43.303998	5	1.457648
pct_women	47.727067	5	1.471894
pct_age_35_39	16.341762	3	1.593003
pct_age_40_44	27.646219	3	1.738892
pct_age_55_59	1.357457	1	1.165100

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.874e+00	2.167e+00	0.865	0.390014	
elevation_gain_m	1.356e-03	8.538e-04	1.588	0.116829	
pct_women	2.340e+01	5.729e+00	4.084	0.000116	***
pct_age_35_39	1.921e+01	1.131e+01	1.698	0.093893	.
pct_age_40_44	-2.050e+01	8.575e+00	-2.390	0.019520	*
pct_age_55_59	-8.522e+00	3.572e+00	-2.385	0.019767	*
elevation_gain_m:pct_women	-7.583e-03	1.912e-03	-3.965	0.000175	***
elevation_gain_m:pct_age_40_44	1.087e-02	4.756e-03	2.286	0.025289	*
pct_women:pct_age_35_39	-5.891e+01	3.266e+01	-1.804	0.075535	.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Thanks!