Marathon Pacing Analysis

Ryan Long
Bellevue University
DSC-530 Winter 2019-2020 - Term Project

Impact of pacing on Marathon results

Running negative splits in a race is commonly thought of as the optimal method for achieving performance for a given distance.

Negative splitting is the practice of finishing a race faster than the initial pace [1]. There is no defined method for the practice in terms of how a distance can be segmented. A distance can be divided equally, with the second half completed faster than the first. For example the first half a mile can be run in 2:55 (175 seconds) and the second half in 2:35 (155), for a total time of 5:30 (330). Alternatively, the distance can be quartered, with quarters subsequent the first completed faster than the previous. Example quarter miles could be covered in 1:30 (90), 1:25 (85), 1:20 (80), and 1:15 (75), again for a total time of 5:30 (330).

Do most runners negative split and does negative splits result in a faster marathon pace? Results from the 2019 IMT Des Moines Marathon were used for analysis [2].

[1] Negative Split. (n.d.) Retrieved November 27, 2019 from Wikipedia https://en.wikipedia.org/wiki/Negative_split

[2] http://onlineraceresults.com/race/view_race.php?race_id=70421&submit_action=select_result&order_by=default&group_by=default#results

Variables Used in the Dataset

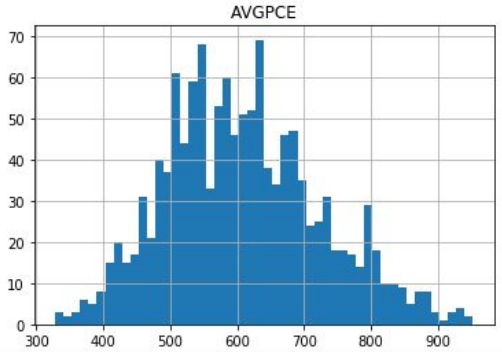
The dataset contained a wide variety of variables including 10k, Half Marathon, 20 mile, and Last 10K times along with Total time and Average Pace. Additional variables were calculated to refine the analysis. In addition to calculating additional variables, pacing metrics were converted from HH:MM:SS format to a seconds per mile integer value. The following variables were created or converted:

Variable	Calculation	Purpose
'SHALF'	df['TIME']-df['HALF']	calculates time for second half of race
'AVGPCE'	df['PACE'] / np.timedelta64(1, 's')	converts average pace for entire race into seconds
'FSTPCE'	(df['HALF']/13.1) / np.timedelta64(1, 's')	converts average pace for first half into seconds
'SNDPCE'	(df['SHALF']/13.1) / np.timedelta64(1, 's')	calculates average pace the second half and converts to seconds
'DIFF'	df['SNDPCE']-df['FSTPCE']	difference between 1st half pace and 2nd half pace, negative meaning faster 2nd half
'FSTQ'	(df['10K']/6.21371) / np.timedelta64(1, 's')	converts average pace first 10K into seconds
'SNDQ'	((df['HALF']-df['10K'])/(13.1-6.21371)) / np.timedelta64(1, 's')	calculates pace for 6.88629 miles after 10k
'TRDQ'	((df['20MI']-df['HALF'])/(20-13.1)) / np.timedelta64(1, 's')	calculates pace for 6.9 miles after half
'FTHQ'	((df['TIME']-df['20MI'])/(6.21371)) / np.timedelta64(1, 's')	calculates pace for last 10K
'PACING'	np.where(df['FSTPCE']>=df['SNDPCE'],'Neg','Pos')	adds a category based on pacing execution

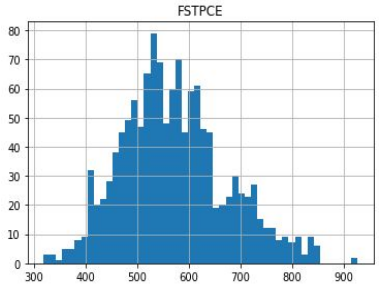
Histograms

Histograms were created using the average pace for each the entire race [1], first half pace [2], and second half pace [3]. Additionally, histograms were created for average pace for the negative split [4] and positive split groups [5]. Outliers were maintained in the data as pacing strategies for the entire field could reveal significant insights. None of the data appeared to be erroneous.

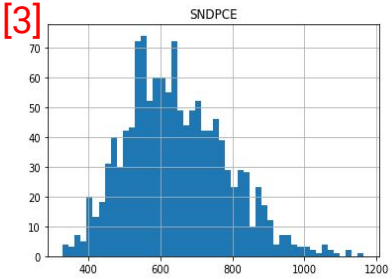
[1]



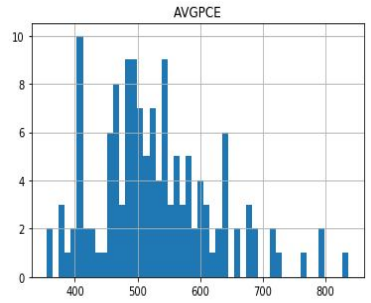
[2]



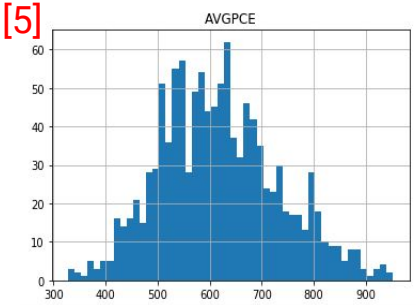
[3]



[4]



[5]



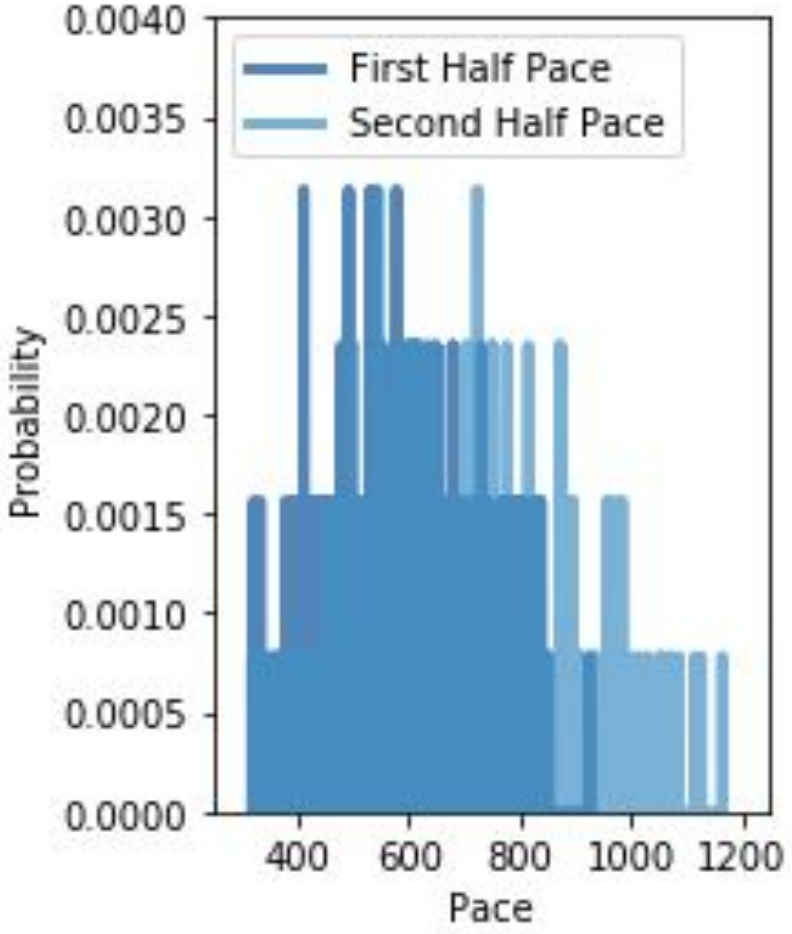
Descriptive Characteristics of Variables

The data set contained 1,276 finishers with complete data (split times). The following are the primary characteristics of the data

Variable	Mean	Variance	Standard Deviation	Mode
AVGPCE	608.9858934	13489.85078	116.1458169	0 543.0 1 546.0
FSTPCE	572.001364	10359.36204	101.7809513	0 412.213740 1 412.366412 2 412.442748 3 493.282443 4 532.519084 5 533.664122 6 540.381679 7 579.160305
SNDPCE	645.0005982	19327.78076	139.0243891	0 545.496183 1 724.580153
Negative Split Group (**139** of 1,276)	528.7697842	8835.091544	93.99516766	0 408.0 1 412.0 2 464.0 3 483.0 4 524.0 5 527.0 6 541.0
Positive Split Group (**1,137** of 1,276)	618.7924362	13183.59244	114.819826	0 646.0

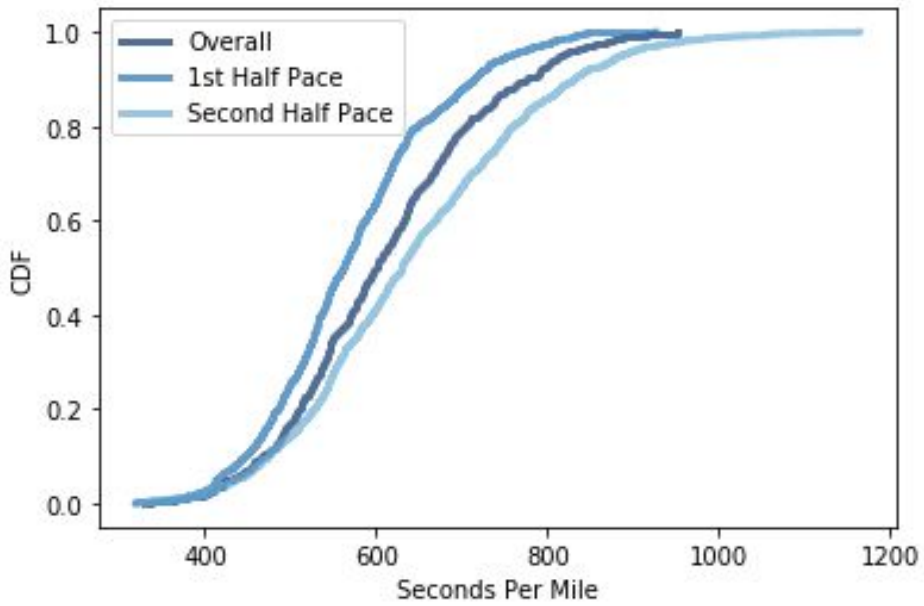
Probability Mass Function (PMF)

To the right, a PMF of the first half pace and second half pace is displayed. This shows runners are more likely to run slower in the second half of the race.



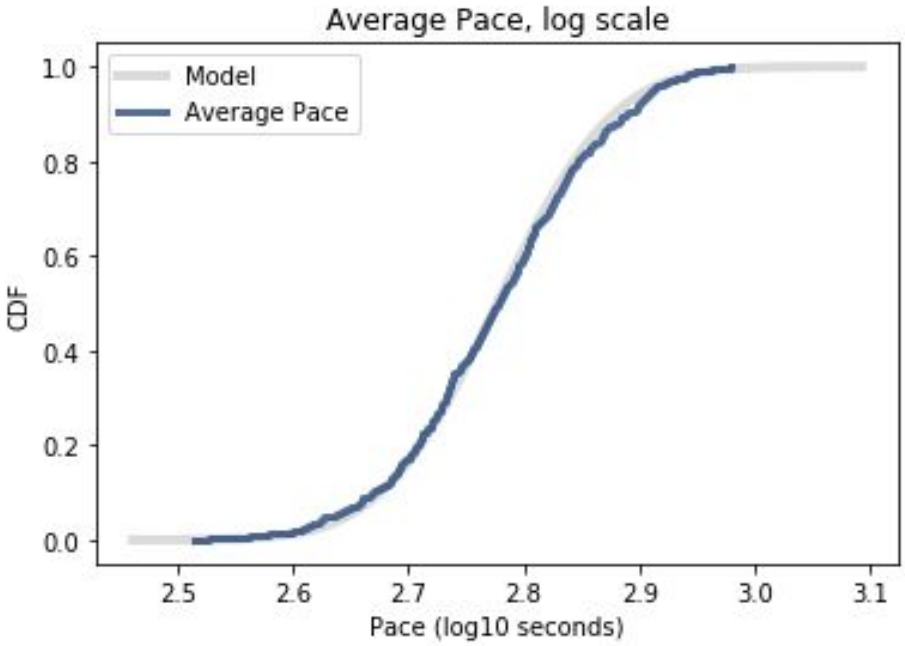
Cumulative Distribution Function (CDF)

To the right, a CDF is plotted showing the first half pace, second half pace, and overall average pace. This shows the fastest runners and slowest runners are more likely to run near their average paces.



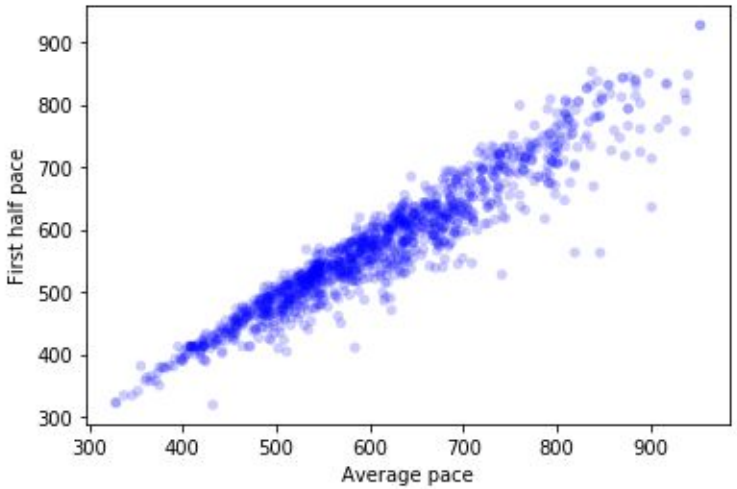
Analytical Distribution

Here's the distribution of average pace and a lognormal model, plotted on a log-x scale. The model is a better fit for the data, although the slower runners are slower than the model expects.

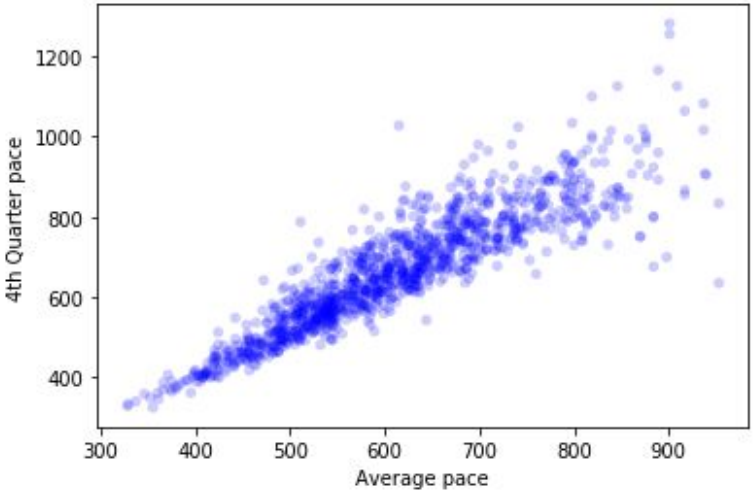


Scatter Plots

The first plot shows first half pace compared to average pace of runners. The plot shows relationship of first half pace is usually faster than the average pace. Calculated correlation (0.9515) and Spearman's Correlation (0.9517) indicate a strong relationship between the first half pace and the average pace of runners.



The second plot show 4th quarter pace compared to average pace of runners. The plot shows relationship of the last portion of the race is usually slower than the average pace. Calculated correlation (0.9032) and Spearman's Correlation (0.9214) indicate a relationship between the 4th quarter and the average pace of runners, but not as strong as the first half.



Hypothesis Testing & Regression Analysis

Hypothesis Testing: To model the null hypothesis, average pace of the negative and positive split groups are the same, a permutation or difference of means test was performed. The resulting p-value from the test was 0, thus the conclusion that the average pace of the two groups is significant.

Regression Analysis: Regression analysis was performed on the two groups, Negative Splitters (NS) and Positive Splitters (PS). Average Pace for both groups was used as the dependent variable, and first and second half paces were chosen as the independent variables.

Formula	R-squared
NS : 'AVGPCE ~ FSTPCE'	0.994
NS : 'AVGPCE ~ SNDPCE'	0.993
PS : 'AVGPCE ~ FSTPCE'	0.910
PS : 'AVGPCE ~ SNDPCE'	0.950

Conclusion: First and Second half pace provide a strong relationship to the average pace of negative splitters. A positive relationship between positive splitters first half pace exists, but the second half pace is stronger. This information combined with the previously covered exploration indicate though fewer runners negative split, it results in a significantly faster average race pace.