

# Negative splits

(sources and last version at [GitHub](#))

This post is about negative splits. Which is a fancy way to say that you start running slower and then increase the speed during the race. The name comes from the fact that when you look the splits, the list of time it took you to run each kilometer (or mile), the numbers decreases.

The splits strategy is somewhat disputed, but many coaches recommend using negative splits (look [here](#)).

The question of this post is whether negative or positive splits running has any relation with the results.

## The race(s)

This time I had to change the races used. The thing with the races in Santiago is that they must have uphill and downhill segments, which could affect the results. So I went for a flat race, and flat Berlin flat is. So the chosen race was the [Berlin Halb Marathon](#).

The specific races where from years 2011, 2013 and 2015 (see the results [site](#)).

This races had some runners:

```
##  
##      BerlinHb2011 BerlinHb2013 BerlinHb2015  
##      M          13726     14229     14636  
##      W          6330      7786      8658
```

that ran at certain speeds (km/h):

```
## $BerlinHb2011  
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.  
##      6.154  9.061 10.150 10.320 11.310 20.880  
##  
## $BerlinHb2013  
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.  
##      6.745  9.448 10.550 10.700 11.730 21.240  
##  
## $BerlinHb2015  
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.  
##      5.915  9.481 10.590 10.740 11.770 21.190
```

## The splits

The races results contains partial times for the first 10 km, which allows to calculate the average for the two segments. Splits are negative if the speed on the second segment is faster than the first. Two attributes were added to the data, *dsign*, which is -1 (for negative splits), 0 or +1. And *delta\_pct* which is the change of speed expressed in percent.

This is how the sign of the splits is distributed:

```
##      event  
##      BerlinHb2011 BerlinHb2013 BerlinHb2015  
##      -1          2316       6399       6483  
##      1          17740      15616      16811
```

And the total percentage of negative splits is **23.25%**.

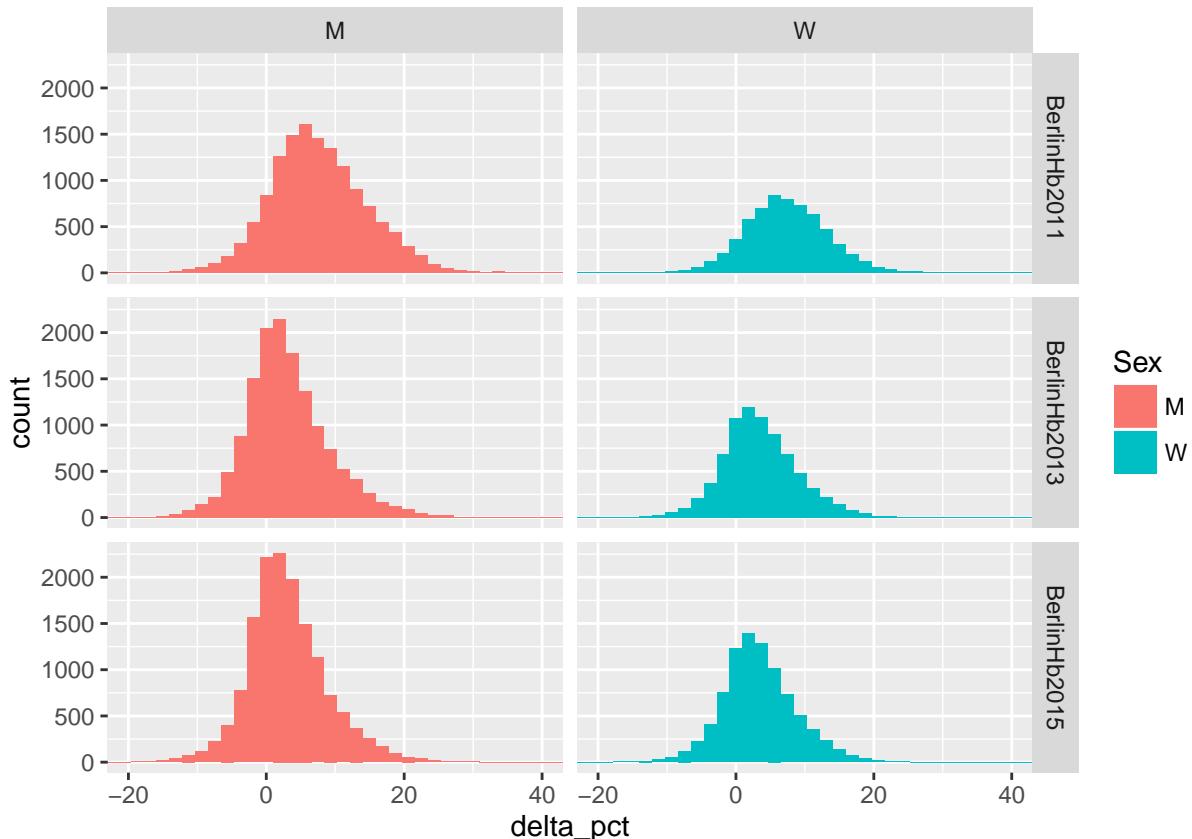
The change in speed (delta\_pct) range is (overall):

```
##      Min. 1st Qu. Median     Mean 3rd Qu. Max.  
## -125.000    0.276   3.807    4.515   8.271  60.510
```

and by race:

```
## $BerlinHb2011  
##      Min. 1st Qu. Median     Mean 3rd Qu. Max.  
## -125.000    3.006   7.044    7.443  11.630  60.510  
##  
## $BerlinHb2013  
##      Min. 1st Qu. Median     Mean 3rd Qu. Max.  
## -100.6000  -0.5542   2.5780   3.2540   6.5550  50.2200  
##  
## $BerlinHb2015  
##      Min. 1st Qu. Median     Mean 3rd Qu. Max.  
## -86.0400  -0.3607   2.6520   3.1850   6.3650  42.2600
```

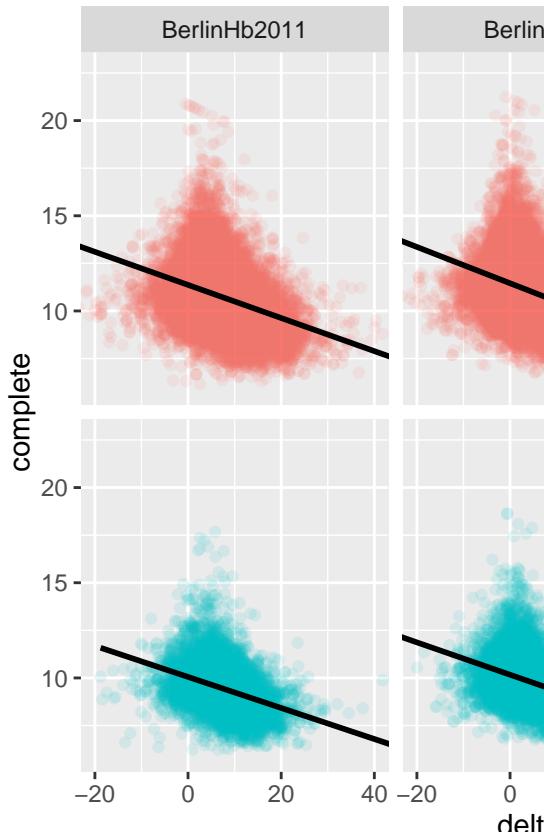
and looks like this:



2011 race looks different than the other two years in shape and center, for reasons unknown to me.

Most people runs positive splits, ranging from 0% to around 8% slower in the second half.

## Splits and speed



This plot illustrates the relation of speed with the change in speed (delta\_pct):

From this charts its clear that there is a relation between the speed difference and the total speed. Those with lower splits tend to run faster. But at the same time splits don't seem to explain the best times, fastest runners tend to have small splits, around zero.

Let's do the numbers, comparing negative splits speeds with no-negative, for women:

```
##  
## Welch Two Sample t-test  
##  
## data: vNeg and vPos  
## t = 29.93, df = 8198, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.5788349 0.6599688  
## sample estimates:  
## mean of x mean of y  
## 10.244138 9.624736
```

and men:

```
##  
## Welch Two Sample t-test  
##  
## data: vNeg and vPos
```

```

## t = 33.954, df = 19185, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.6164811 0.6920167
## sample estimates:
## mean of x mean of y
## 11.54288 10.88863

```

So in both cases we have a significant difference (p-values almost 0).

That was using just a negative/non-negative attribute. What about the size in speed change?

### Relation with other factors

Let's compare with **Sex**:

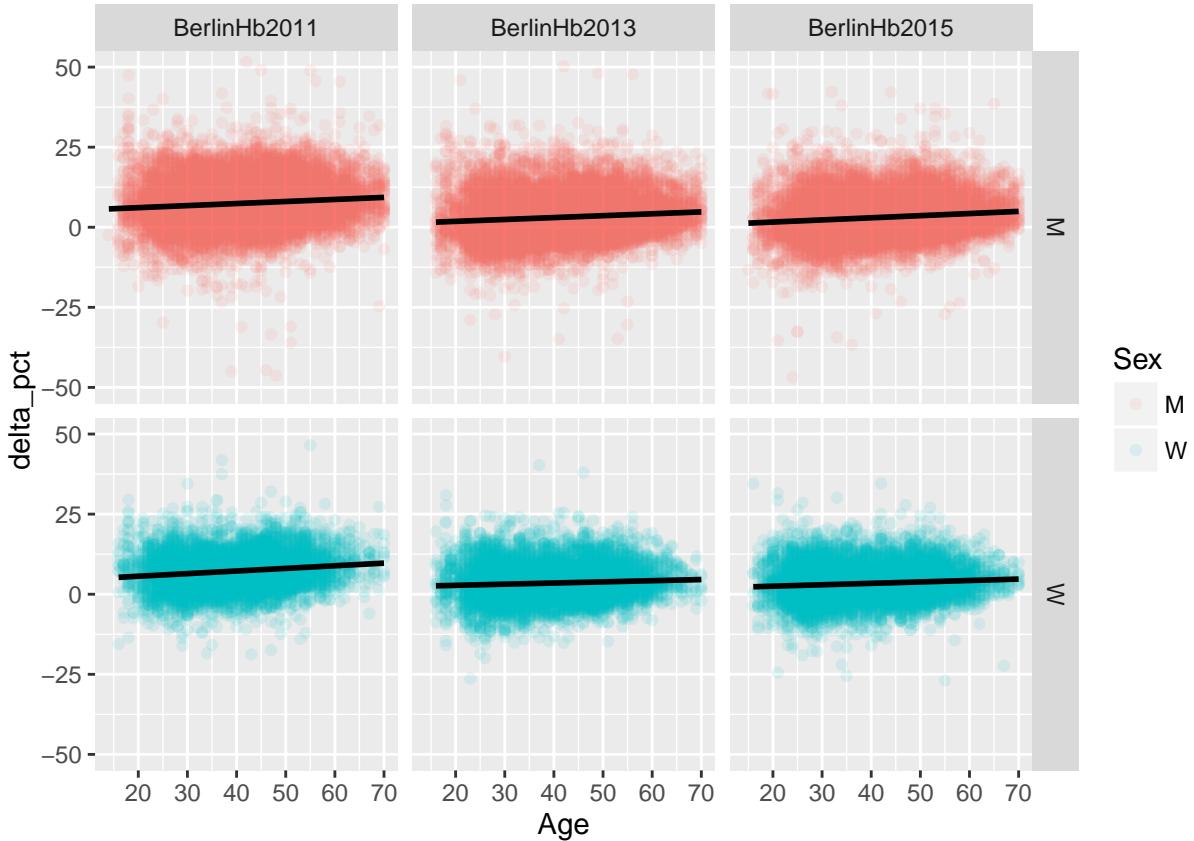
```

##
## Welch Two Sample t-test
##
## data: vM and vW
## t = 0.80468, df = 52630, p-value = 0.421
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05981364 0.14313333
## sample estimates:
## mean of x mean of y
## 4.529098 4.487438

```

So it doesn't seem to be an association between **Sex** and the speed change.

Then the relation with **Age**:



```
## lm(formula = delta_pct ~ Age, data = df)

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.99655813 0.099077172 20.15155 4.903596e-90
## Age         0.06117036 0.002324882 26.31117 8.867464e-152
```

And there is a relation with Age, **0.6%** of change in speed for each decade of age.

### Comparing association with speed

First look at the relation between speed and speed change on splits:

```
## lm(formula = complete ~ delta_pct, data = df)

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.01116582 0.0079639887 1382.61947      0
## delta_pct   -0.09155948 0.0009969797  -91.83686      0
```

So the split delta accounts for **11.4%** of the total speed variability among runners.

Comparing with **Age**:

```
## lm(formula = complete ~ Age, data = df)
```

```

##             Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 11.39360173 0.0267821824 425.41723 0.000000e+00
## Age         -0.01933209 0.0006284537 -30.76136 2.580757e-206

```

Which explains just **1.4%** of the total variability.

Considering also **Sex**, with **delta\_pct**:

```
## lm(formula = complete ~ Sex + delta_pct, data = df)
```

```

##             Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 11.46368706 0.0086571097 1324.19335 0
## SexW        -1.29515316 0.0128251465 -100.98545 0
## delta_pct   -0.09184154 0.0009272741 -99.04465 0

```

it's **23.4%**.

And with **Sex** and **Age**:

```
## lm(formula = complete ~ Sex + Age, data = df)
```

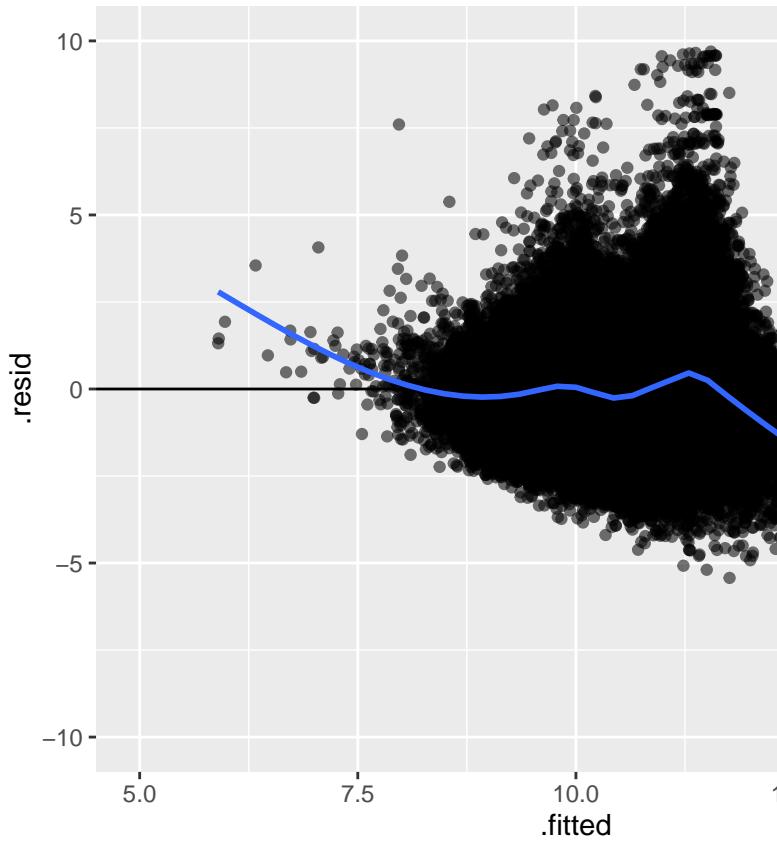
```

##             Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 12.09752538 0.0259742116 465.75140 0
## SexW        -1.34793835 0.0136330736 -98.87267 0
## Age         -0.02502357 0.0005889715 -42.48689 0

```

it's **14.3%**.

## How well the splits predict the speed?



This is the residuals plot for fit with **Sex** and **delta\_pct**:

As was seen in the second figure, the speed change between first and second half seem to be a better predictor of slower speeds than of higher.

Let's see what is the range for **delta\_pct** for the *top 5% of male* runners:

```
## $BerlinHb2011
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
## -125.000   1.825   3.831   3.734   5.965  18.040
##
## $BerlinHb2013
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
## -7.7820 -0.7309  1.0590  1.1630  2.9570 19.0100
##
## $BerlinHb2015
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
## -17.2200 -0.7451  0.7912  1.0450  2.4950 38.0200
```

The center of this ranges is closer to 0 (and more negative) than the range for all runners.

And if we test for difference between negative and no-negative splits:

```
##
## Welch Two Sample t-test
##
```

```
## data: vNeg and vPos
## t = 2.0686, df = 1013.3, p-value = 0.03883
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.006702763 0.254118218
## sample estimates:
## mean of x mean of y
## 15.50940 15.37898
```

We find the in this group the split sign still have a significant association with the speed, being positive splits slower (with a p-value of **0.039**).

## Conclusion

Negative splits are a good thing for running. But as most good things should not be overdone. Best results are associated with just a slight improvement of speed on second half.

One notable thing is that having negative, or lower splits, have a stronger association with the overall speed than age. Meaning that much of the age disadvantage can be neutralized with pace management.

regards