MSc in Statistical Science
Statistical methods for functional data

Analysis of biometric data during a run

Lorenzo Cifelli
Marta Gaio
Giulia Pacchetti

May 31 2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Index

Index

Cardiorespiratory data for 10 athletes during a 6.1km long run. The route includes four stages with different gradients:

- **Flat initial** section, 1.3 km long with a 0% gradient;
- **Uphill** section, 1.3 km long with a +6.8% gradient;
- **Downhill** section, 1 km long with a -7.2% gradient;
- **Flat final** section, 2.6 km long with a 0% gradient;

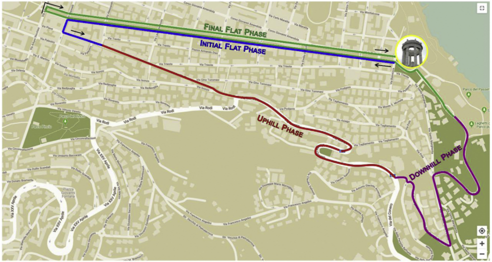


Figure: "Around Ancona" route

- Weekly training rate;
- Smoking habit;

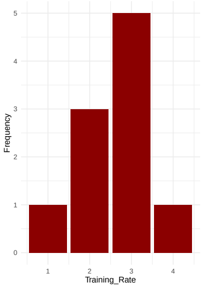


Figure: Weekly training rate

- Finish times;

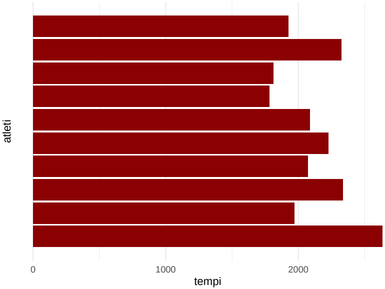


Figure: Finish times

Features

Functional features

- Heart rates `hr`;
- Respiratory rate `br`.

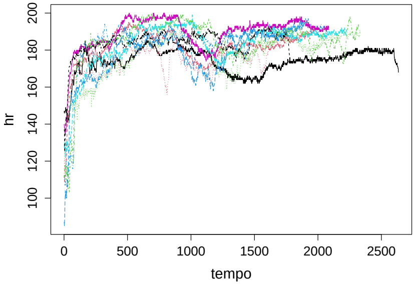


Figure: `hr`

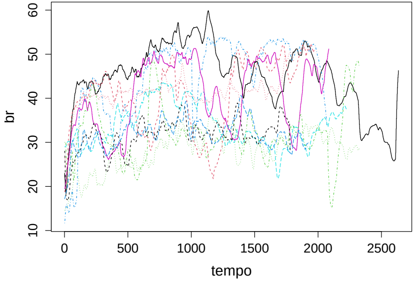


Figure: `br`

Index

Università degli Studi di Padova

- The data series oresent **different lengths**.
- It is appropriate to modify the series by changing the **domain** from temporal to **spatial**.

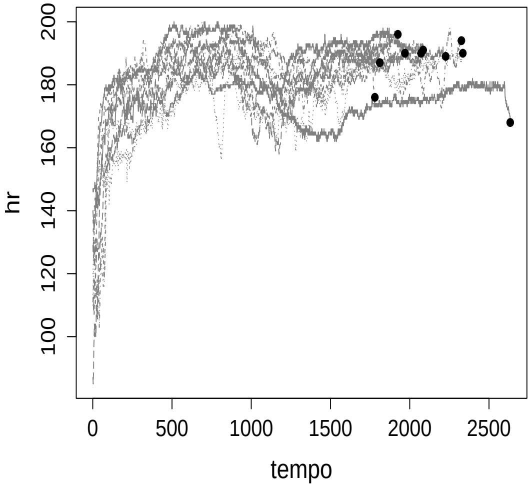
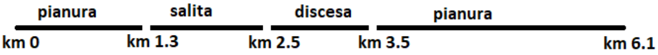


Figure: `hr` features on temporal domain

The following **information** were used for this purpose.:

- points at which the slope of the route changes:



- instants of time when each subject arrives at a change in slope;

Asunction: the speed is constant within the 4 different sections of the route.

$$\Downarrow$$

It is possible to calculate for each subject at each time instant his or her position relative to the initial point of the path.

Domain changing

- the **observation interval** (0, 6,1) km is the same for all subjects;
- a common observation grid was chosen for all subjects consisting of 200 points equi-spaced between 0 km and 6.1 km;
- the same operation was carried out for the functional variable related to respiratory rate `br`.

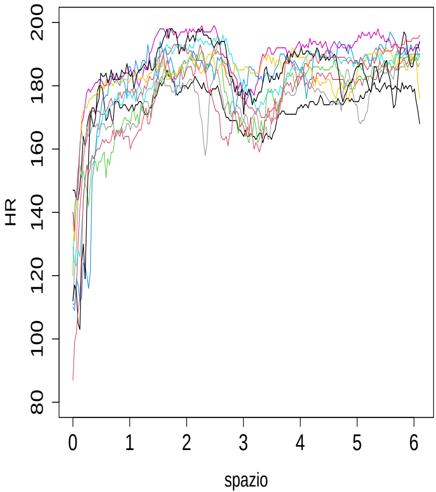


Figure: `hr` feature on spatial domain

Speed feature

The domain change results in a loss of information related to arrival times ⇒ correct for speed.

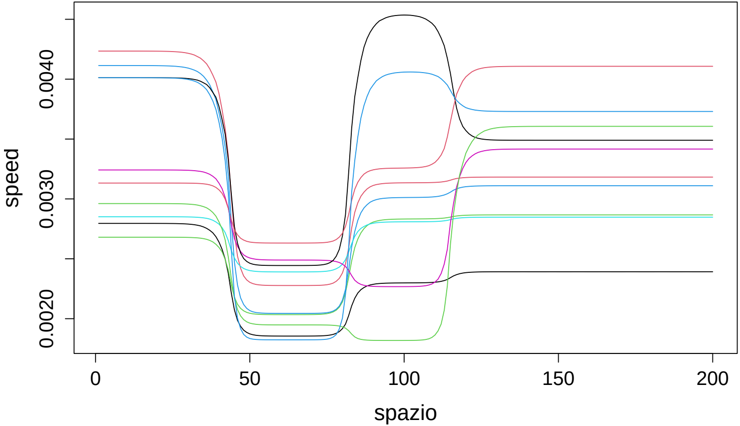


Figure: Speed along the route

We smoothed the function `hr` for each subject using smoothing splines of order 6 by penalizing the fourth derivative. The adjustment parameter was set equal to $\lambda_{opt} = 0.00001$.

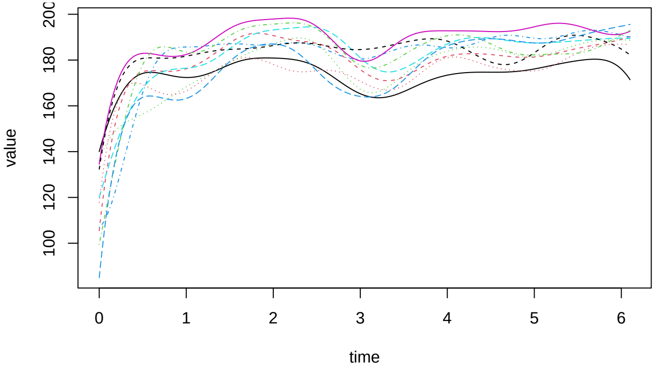


Figure: `hr` feature after smoothing

Smotthing

We next smoothed the function `br` for each subject using smoothing splines of order 6 by penalizing the fourth derivative. The adjustment parameter was set equal to $\lambda_{opt} = 0.001$.

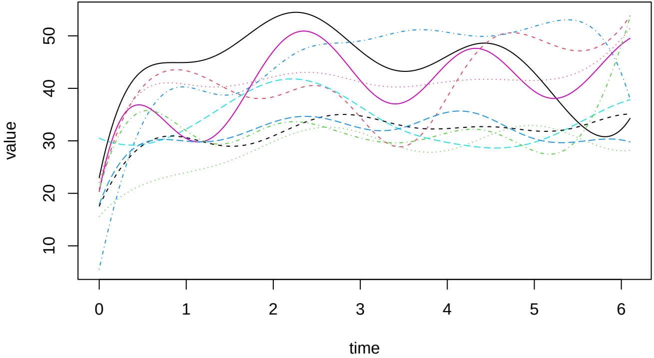


Figure: `br` feature after smoothing

Index

Exploratory analysis

Media

- Common trend among the subjects.
- Growth in the initial phase until the middle of the climb where it stabilizes.
- Decrease in the downhill section.
- Growth at the beginning of the plain and then a stabilization.

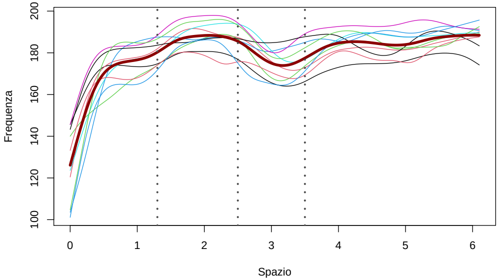


Figure: Heart rate

- At the beginning of the activity there is a lot of variability because subjects start from different frequencies;
- It decreases until shortly after the start of the climb where the frequencies stabilize and remains fairly constant for the rest of the run.

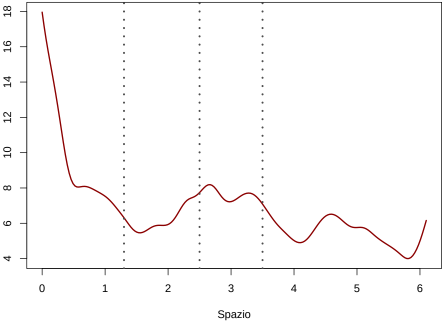


Figure: Functional variance

Università degli Studi di Padova

- Derivatives change sign when a runner moves from one phase of the race to the next.
- For example, at the beginning of the climb both become positive to detect the fact that beats are increasing and accelerating.

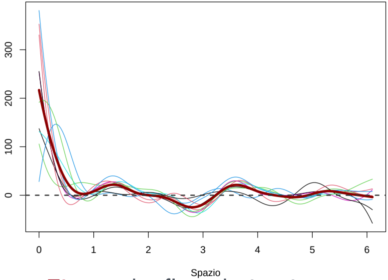


Figure: hr first derivative

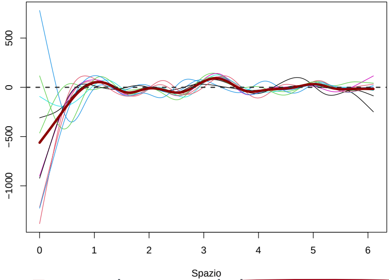


Figure: h

- he first 3 principal components capture about 86% of the variability;
- The first component represents the overall trend.

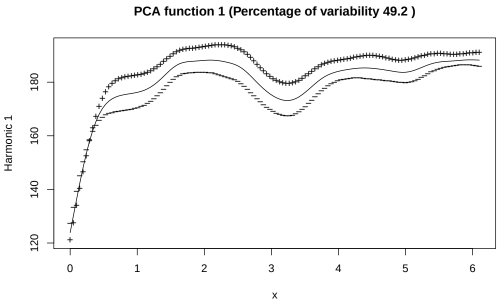


Figure: First functional principal component

Functional pricipal components

- The second component captures the variability in the two flat section, note the change of sign;
- The third component captures variability in segments where frequency is more stable;

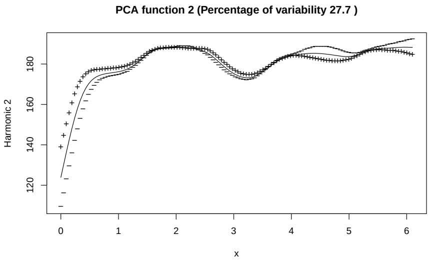


Figure: Second functional principal component

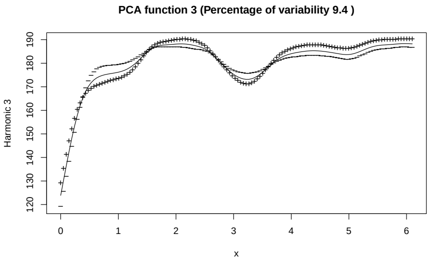


Figure: Third functional principal component

Università degli Studi di Padova

- Arrival times are modeled as a function of heart rate trends along the run:

$$\texttt{tempi}_i = \alpha + \int \beta(t)\texttt{hr}_i(t)dt + \varepsilon_i \quad i = 1, ..., 10$$

- The functional coefficient $\beta(t)$ is represented in bases via *splines* with 25 basis functions and λ chosen via (OCV).

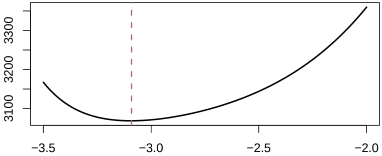


Figure: Cross-validation error, $\lambda_{opt} = 0.003$

Estimated coefficient

The effect of heart rate on finish times appears to be different depending on the stages of the course:

- **Climb**: as `hr` increases, the arrival time increases;
- **Downhill**: as `hr` increases, the arrival time decreases;
- **Final section**: as `hr` increases, the arrival time decreases considerably.

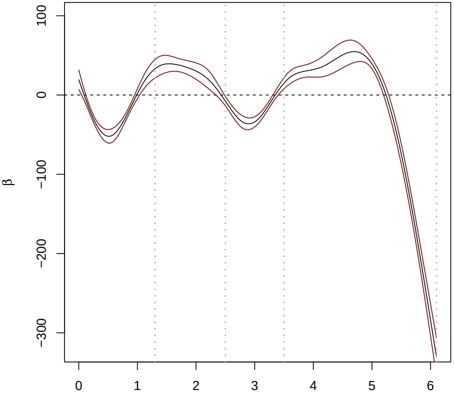


Figure: Estimated functional coefficient for `hr`

Permutation test

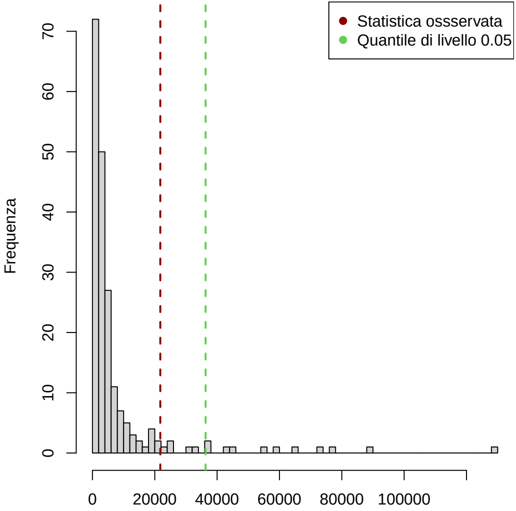


Figure: F permutation test

pvalue of permutation test: 0.08

Model on principal components

- *fpca* with the first 8 principal components;
- linear model on the principal components with *stepwise* selection of covariates;

fpca	Estimate	*pvalue*
`Intercept`	2116.40	***
`Pc4`	31.15	*
`Pc5`	-31.14	*
`Pc6`	60.21	**

Table: linear model output

pvalue of F test: 0.008735

Estimated coefficients

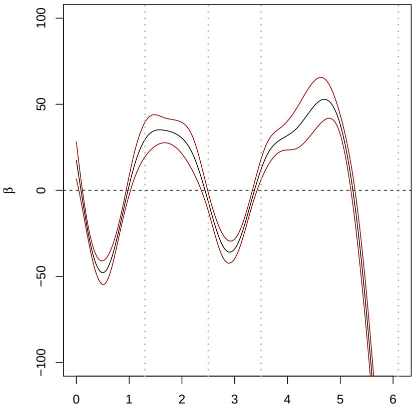


Figure: Coefficient estimated for the model with functional covariates

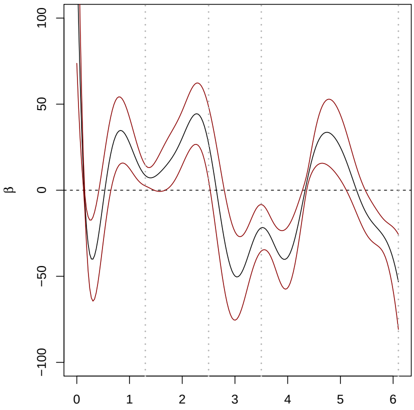


Figure: Estimated coefficient for the model on principal components

Do changes in heart rate have an effect on finishing time throughout the course? $\Rightarrow$ FLiRTI approach

Estimated model:

$$\texttt{time}_i = \beta_0 + \int \beta_1(t)\texttt{hr}_i(t)dt + \varepsilon_i(t) \quad i = 1, \dots, 10$$

Target function to be minimized:

$$PENSSE(\lambda) = \sum_{i=1}^{10} \Big(\texttt{time}_i - \beta_0 - \int \beta_1(t)\texttt{hr}_i(t)dt\Big)^2 +$$

$$\lambda_1 \int |\beta_1(t)|dt + \lambda_2 \int |\beta_1^{(2)}(t)|dt$$

Estimated coefficient

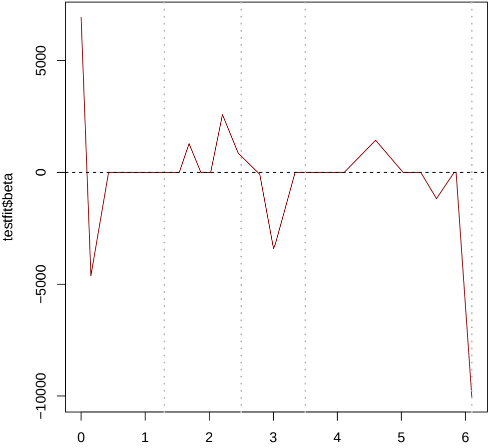


Figure: FliRTI regression functional coefficient

Index

We transformed the categorical weekly training rate variable as follows:

- Less than 3 workouts per week $\Rightarrow$ "poorly trained".
- At least 3 workouts per week $\Rightarrow$ "very trained".

The poorly trained always have higher beats than the very trained, except for the last stretch.

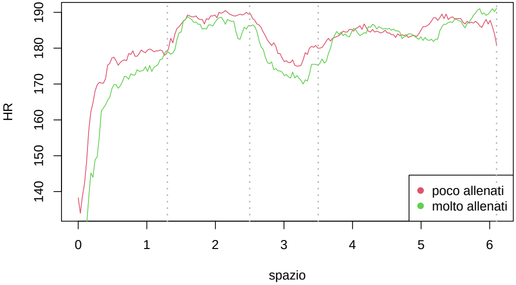


Figure: Trend of `hr` in the two groups

Model for HR - allenamento

t-test

There seems to be little evidence of significance of the difference between the two groups, except for the very first initial section where there is evidence at the 10% level.

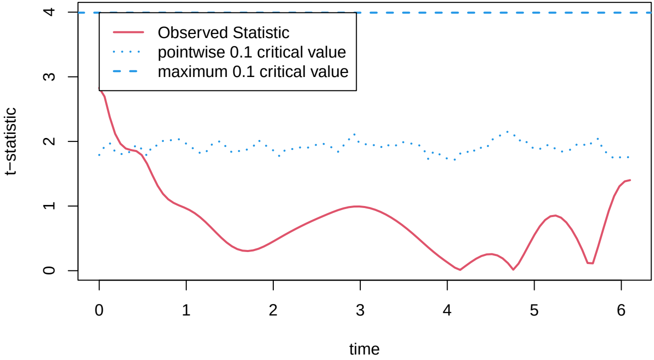


Figure: T-test

Model for HR - allenamento

Functional ANOVA

heart rate is modeled according to weekly training rate and adjusted for finishing times:

Estimated model:

$$\mathtt{hr}_i(t) = \alpha(t) + \beta_1(t)\mathtt{training}_i + \beta_2(t)\mathtt{time}_i + \varepsilon_i(t) \quad i = 1, \ldots, 10$$

The smoothing parameters for the functional coefficients are chosen via generalized cross-validation.

Model for HR - allenamento

Estimated coefficient

- Finish time decreases as beats increase.
- Compared to the less trained, the most trained athletes have lower beats at the start;

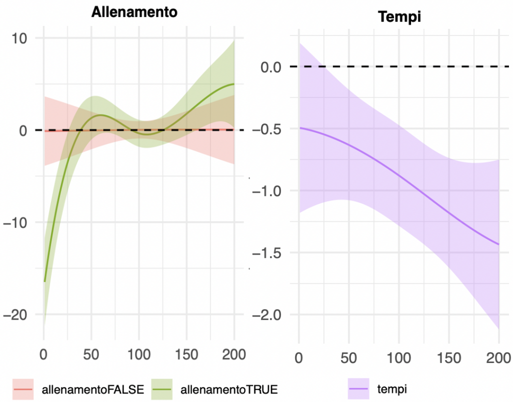


Figure: Estimated coefficients by ANOVA

Model for HR - allenamento

Resiaduals analysis

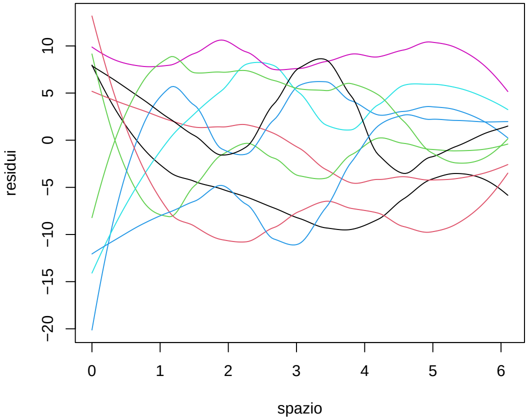


Figure: Functional residuals of ANOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

There seems to be a difference only in the upward phase between the rates of smokers and nonsmokers.

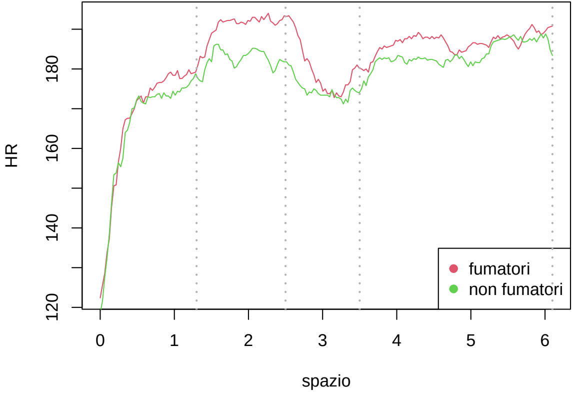


Figure: Trend of `hr` in the two groups

Model for HR - smoker

t-test

The t-test confirms the significant difference only in the climb phase.

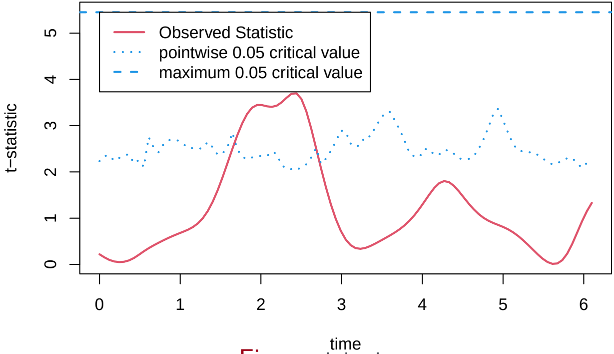


Figure: t-test

Model for HR - smoker

Funzional ANOVA

the heart rate trend along the run is modeled according to the smoking habit of the athlete taking into account the time of arrival:

Modello stimato:

$$\mathtt{hr}_i(t) = \alpha(t) + \beta_1(t)\mathtt{smoker}_i + \beta_2(t)\mathtt{time}_i + \varepsilon_i(t) \quad i = 1, \ldots, 10$$

The smoothing parameters for the functional coefficients are chosen via generalized cross-validation.

- In the climb section smokers have higher rate than non smokers.

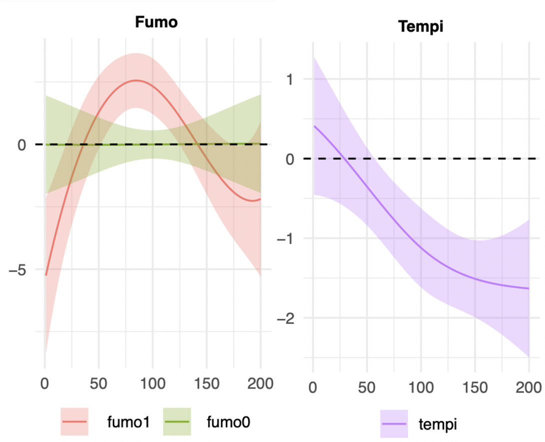


Figure: Estimated coefficient for ANOVA

Model for HR - smoker

Residuals analysis

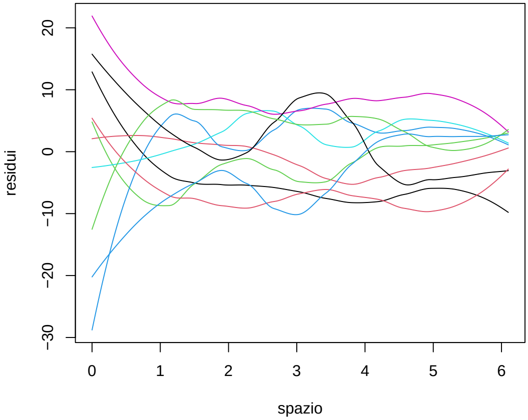


Figure: Functional residual of anova ANOVA

Simultaneous model

The influence of respiratory rate on heart rate at the simultaneous level is investigated, using speed as a control variable.

$$\texttt{hr}_i(t) = \alpha(t) + \beta_1(t)\texttt{br}_i(t) + \beta_2(t)\texttt{speed}_i(t) + \varepsilon_i(t).$$

The smoothing parameter for the functional coefficient is set through generalized cross-validation.

- The functional coefficient is positive and significantly different from 0 only in the first part of the run.
- This implies that at this stage there is an increase in both breaths and heartbeats.
- In the rest of the activity at the simultaneous level, there does not seem to be an influence between the two functions.

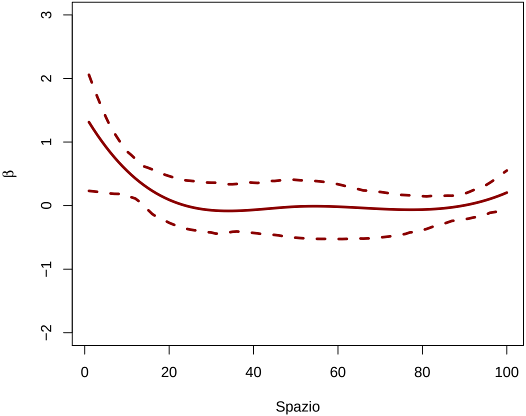


Figure: Function coefficient for `br`

Non-simultaneous model

one can think of a delayed effect of breathing on beats at successive km;

$$\Downarrow$$

heart rate is modeled as a function of respiratory rate, adjusting for speed:

$$hr_i(t) = \alpha(t) + \int_0^t \beta_1(s,t)\texttt{br}_i(t)ds + \int_0^t \beta_2(s,t)\texttt{speed}_i(t)ds + \varepsilon_i(t)$$

The functional coefficients $\beta_1(s,t)$ and $\beta_2(s,t)$ are represented in bases by the product between two sets of penalized *B-splines*.

Non-simultaneous model

There is a significant effect of breaths in the first kilometers on all subsequent beats.

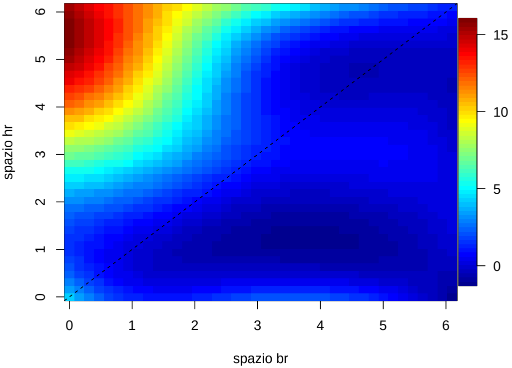


Figure: Bivariate functional coefficient associated with the variable `br`

Index

Conclusion

- In order to complete the route in short **time**, it is desirable for the heart rate to remain low on the climb and then increase later on the descent and especially in the final stretch;
- Performing at least three **workouts** per week seems to cause a decrease in beats in the initial section;
- subjects with aptitude for **smoking** seem to have higher beats than those who do not smoke during the uphill phase;
- **breathing frequency** at the beginning seems to positively affect the heart rate in the last km.

Possible future developments

Limits:

- Low sample size.
- Assumption of constant speed within route sections.

Improvements:

- In a future study, also record speed in a fine grid of space.
- Considering models for heart rate acceleration.
- Considering bivariate functions;
- Compare with athletic trainers and sports physicians to be able to interpret and implement the results obtained.

Thank you for your attention!