Master degree in Mathematical Engineering

Master Thesis, Politecnico di Torino

Academic Year 2018/2019

Sensory evaluation and rheological measurement of bread samples:

Comparing subjective perceptions of trained tasters against objective measurements by a texture analyzer.

Thesis Supervisor: Paolo Brandimarte

Author: Lucia Baratella

Contents

Thanks

Thanks to my family, my boyfriend and my friends for always having faith in me; I wish I believed in myself as much as you do in me.

Thanks to my thesis supervisor for helping me to face a not standard real problem with valuable advices.

Thanks to my company supervisor, to the team of the Laboratory, to the team of the Rheological Laboratory and to the Sensory team of Soremartec s.r.l for the support received for this project.

Finally, thanks to Politecnico that let me grow professionally.

Introduction

The thesis deals with my personal internship experience at Soremartec s.p.a GROUPE Ferrero. My company supervisor is Antonucci Marco, he works at Raw Materials R&D and he is the manager of the Rehological Laboratory and of the Sensory team of Soremartec.

My internship lasted five months, from February to June 2019, and it deals with cooperate with both Rheological Laboratory and Sensory team in order to find a correlation between sensory and rheological data.

The first two months were useful for the bibliographic research about bread, sensory evaluation of bread and so on, for the study of food sensory evaluation and for the usage of Texture Analyzer instrument.

Next, with the help of the Laboratory of Soremartec, I saw and helped to make the bread recipe and the bread samples in order to understand better the problem relate with bread.

Figure 1 Me during bread preparation

In fact, bread is a sensitive matter both during the preparation and during the taste or the measurements. For example, if a bread in a baking pan remains in the oven some seconds more than another one, it could be very different in hardness, elasticity and so on. Another element that influences the bread preparation is the temperature of the oven: indeed, if a dough is put in the oven as first, the oven is at a higher temperature compared to the same, in case it remains open for some seconds.

Another example is the following one: if someone opens the wrapper of a bread sample for a long time, such as half an hour or even less before the taste, the bread starts to become drier and drier: this effect is due to the retrogradation phenomena.

Retrogradation phenomena is linked to starch; starch consists of two types of molecules: the linear and helical amylose and the branched amylopectin.

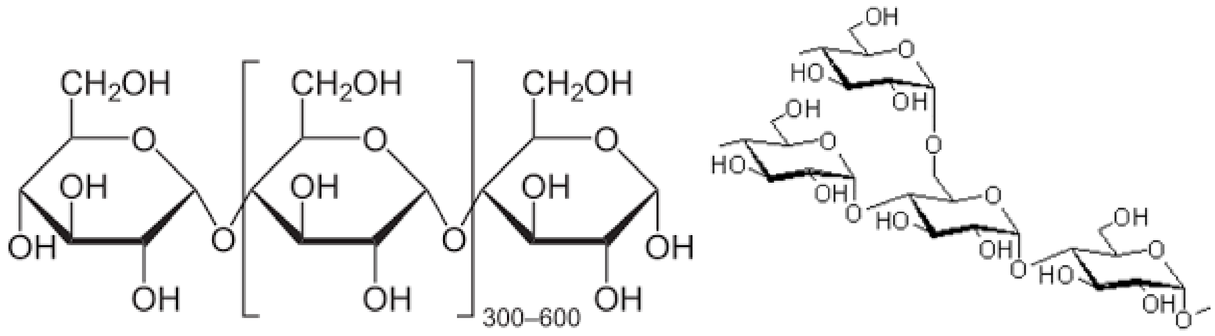


Figure 2 Amylose and amylopectin molecules (Wikipedia figures)

"Retrogradation is a reaction that takes place when the amylose and amylopectin chains in cooked, gelatinized starch realign themselves as the cooked starch cools.

When native starch is heated and dissolved in water, the crystalline structure of amylose and amylopectin molecules is lost, and they hydrate to form a viscous solution. If the viscous solution is cooled or left at lower temperature for a long enough period, the linear molecules, amylose, and linear parts of amylopectin molecules retrograde and rearrange themselves again to a more crystalline structure. " Wikipedia - Retrogradation starch

Moreover, after two or three compressions, the structure of a bread sample changes because the compression breaks the chemical bonds of the starch; therefore, using the same sample to do several tests is not recommended.

The recipes of bread were decided by the Laboratory team and tasting sessions were planned by Sensory team (this structure was used to plan each of the three experiments).

After each taste and measurement, I took one or two weeks in order to see the data, to discuss them with my thesis supervisor and to decide if something needs to be changed or if the next the ordtaste should have been different: for example, on the advice of my thesis supervisor and my company supervisor, the third experiment is different than the first two, as explained in the next chapters.

So, in the following sections, I will explain about the Texture analyzer instrument, the panelists, the tastes and the data analysis.

1. Texture profile analysis

TPA is all about the recognition of texture as a multi-parameter attribute; the test consists of compressing a bite-size piece of food two times in a reciprocating motion that imitates the action of the jaw and extracting from resulting force-time curve number of textural parameters.

Figure 3 Instrument for TPA

The instrument measures the mechanical textural characteristic of foods that can be divided into the primary parameters of **hardness**, **cohesiveness**, **springiness** and **adhesiveness**, and into the secondary (or derived) parameters of **fracturability**, **chewiness** and **gumminess**.

The instrument measures the force that the probe detects every 0,004 seconds and builds a time-force curve.

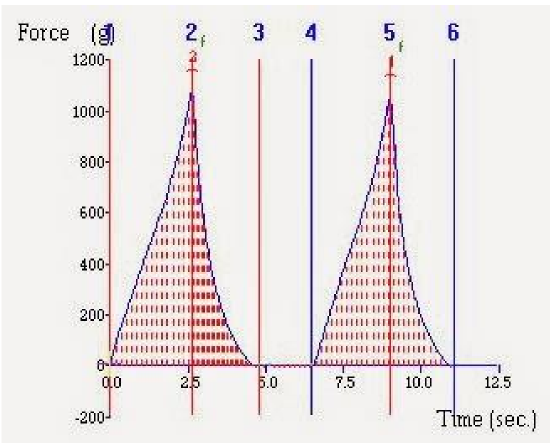


Figure 4 TPA force curve

Let's define the parameters that are used to analyze bread pieces:

- **HARDNESS**

  Hardness is defined as the maximum peak force during the first compression cycle; units are g.

- **SPRINGINESS (Elasticity)**

  Springiness relates to the height that the food recovers during the time that el apses between the end of the first bite and the start of the second bite.

  There are no units for this parameter and the value is calculated as below (look at the picture above) because the velocity during the compression is constant, so the ratio between heights is a ratio only between time values:

$$\frac{Time\ Difference\ 4:5}{Time\ Difference\ 1:2}$$

- **COHESIVENESS**

  Cohesiveness may be measured as the rate at which material disintegrates under mechanical action.

  There are no units for this parameter and the value is calculated as (look at the picture above):

$$\frac{Area\ 4:6}{Area\ 1:3}$$

2. Planning the first taste

First of all, the planning of the experiment (rheological and sensory) needs to be done.

The planning consists of:

- ✓ Choice of the bread recipes
- ✓ Choice of sensory attributes

Bread sample will be tested by the instrument and by the tasters, so the attributes of the TPA curve need to be logically linked to the sensory attributes.

2.1 Choice of the bread recipes

For the first taste, four different recipes are used; they have been created in order to mark differences between breads:

- **A**
  This recipe is the standard one of the Ferrero Kinder Brioss.

Figure 5 Ferrero Kinder Brioss

- **B**
  This recipe is obtained by increasing the rate to 25% of grease (sunflower oil) of the standard recipe, which consists of an increase to 2% in absolute term on the dough; by adding grease, the bread is expected to be softer.
- **C**
  This recipe is obtained by adding a rate to 0,2% of a specific enzyme in the mixture of powders (flour, gluten, premix, ...) of the standard recipe; by adding the enzyme, the bread is expected to be chewier.

- **D**
  This recipe is obtained by adding a rate to 22% gluten, which consists of an increase to 1,1% in absolute term on the mixture of powders of the standard recipe; by adding gluten, the bread is expected to be harder.

Bread samples are not stuffed with milk cream, because it could make less effective the measurement of the instrument and of the tasters.

2.2 Choice of the sensorial attributes

The tasters need appropriate sensory attributes to sensory evaluate the different bread samples. In order to do that, a meeting is organized whit the tasters: there are two groups of people who come to taste the sample, one in the morning and the other in the afternoon. Each group is made up of twenty people, females and males, and their age range is between eighteen and sixty years.

First, some bread samples are handed out to the tasters, who have to write some characteristics that come to their mind. After doing that, they are asked to focus on the attributes that could be logically linked with the TPA ones. Some food examples, like Goleador (taffy), biscuits, wafer, marshmallow, were used to help panelists understand physically the concepts of hardness, elasticity and so on by the touch and the mouth. The definition of an attribute is not very simple, because it needs to find an agreement among the tasters: everyone must share it and understand it in order to make the experiment significant.

The selection of sensory attributes and their definition should be closely related to the real chemical and physical properties of bread samples. Next, they are asked to compare the different samples for each attribute, indeed the experiment goal is to compare the different samples and not to evaluate in an absolute way.

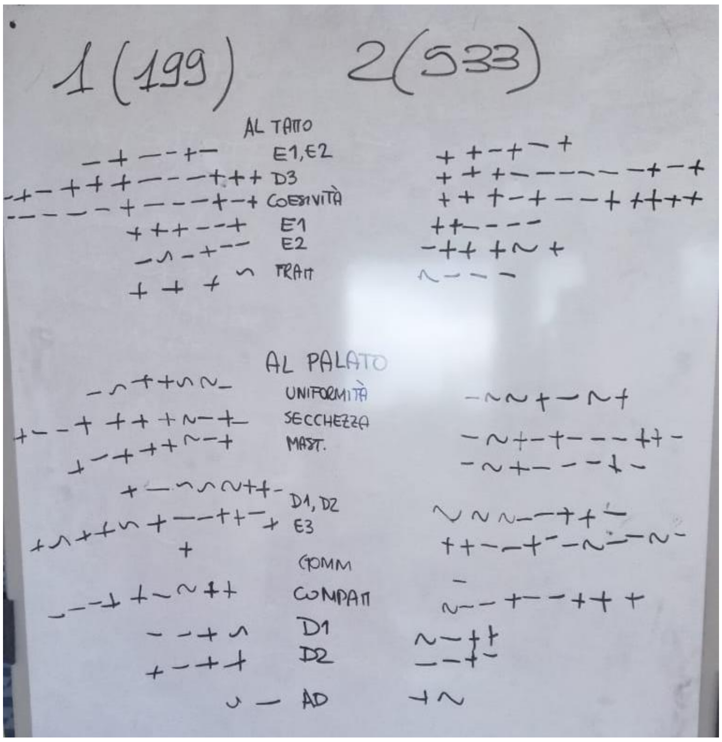


Figure 6 Comparison between the different sample

Finally, after collecting all the information about the attributes, a sensory vocabulary is drawn up (it is written in Italian because the tasters are all Italian people).

VOCABOLARIO SENSORIALE

Attributi al tatto:

- elasticità (indice e pollice): capacità di ritornare nella forma iniziale dopo una compressione fra l'indice e il pollice (marshmallow molto elastico)
- elasticità (indice e medio): capacità di ritornare nella forma iniziale dopo una compressione con l'indice e il medio
- durezza (indice): resistenza alla compressione usando l'indice
- coesività: capacità del campione di essere deformato prima della rottura (goleador molto coesiva, wafer per nulla coesivo)

Attributi al palato:

- elasticità (labbra): capacità di ritornare nella forma iniziale dopo una compressione tra le labbra
- durezza (incisivi): forza richiesta per mordere il campione con gli incisivi e spezzarne un pezzo (marshmallow poco duri, biscotti abbastanza duri)
- durezza (masticazione): forza richiesta per masticare il campione
- secchezza: misurazione della quantità di saliva assorbita dal campione
- compattezza: capacità del campione di formare un bolo in bocca che fatica ad essere deglutito

2.3 Software Compusense

In order to prepare sensory test and to collect sensory data in a simple way, the sensory team of Soremartec uses the software Compusense and its cloud to share data among different companies.

The software offers sensory services:

- screening, training, monitoring and maintaining of the panel (the group of the tasters)
- training programs for lexicon development
- consulting service option to leverage the experience of Compusense with: statistical analysis and interpretation, consumer guidance and concept testing, creative consulting and ideation

In order to achieve the thesis goal, only the graphical package is used to help the panelist to express his feedback.

The software gives the possibility to choose the layout that the panelists are going to use during the experiment and the layout of the method, thanks to which they vote the different attributes.

Moreover, it can be required to the panel to vote all the attributes. This option is used for all the tastes, because it helps to avoid having incomplete tasting data. The software assigns a sample code to the different samples, and randomize the order of the samples for each panelist.

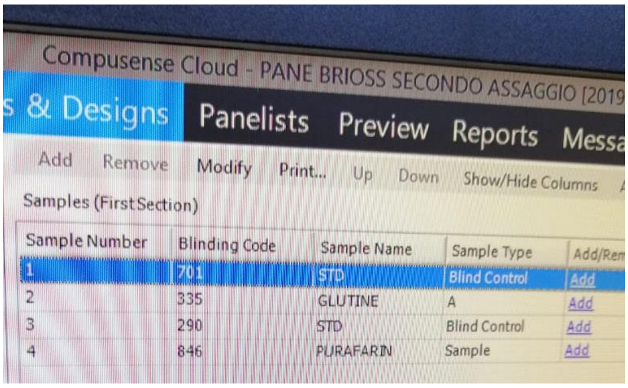
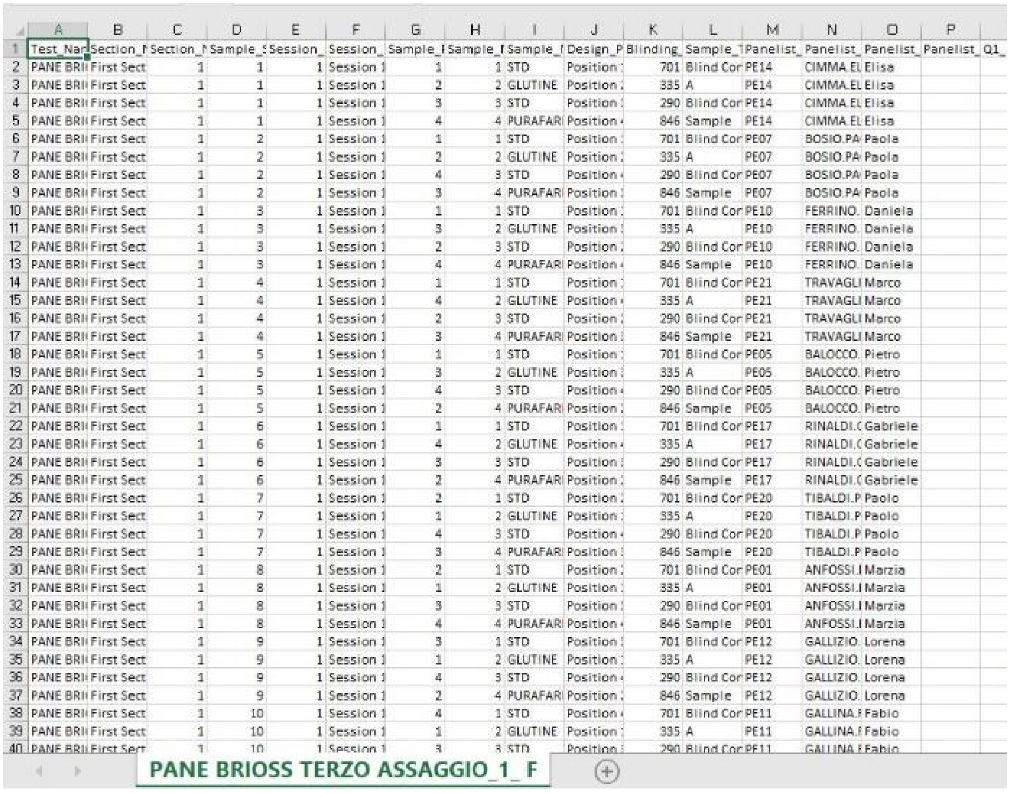


Figure 7 Compusense

Finally, photos or descriptions can be added for each attribute in order to help the panelist to remember the correct attribute definition.

At the end of the taste, Compusense gives the possibility to obtain an output on Excel where each line corresponds to a panelist and his votes for a fixed bread sample, accompanied by other information like the name of the attributes, the code of the sample, the name assigned by the author of the test to the bread sample and the timestamp for each vote.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Test_Nar	Section_I	Section_I	Sample_S	Session_	Session_	Sample_I	Sample_I	Sample_I	Design_P	Blinding_	Sample_T	Panelist_	Panelist_	Panelist_	Panelist_	Q1_
2	PANE BRI	First Sect	1	1	1	Session 1	1	1	STD	Position	701	Blind Cor	PE14	CIMMA.EL	Elisa		
3	PANE BRI	First Sect	1	1	1	Session 1	2	2	GLUTINE	Position	335	A	PE14	CIMMA.EL	Elisa		
4	PANE BRI	First Sect	1	1	1	Session 1	3	3	STD	Position	290	Blind Cor	PE14	CIMMA.EL	Elisa		
5	PANE BRI	First Sect	1	1	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE14	CIMMA.EL	Elisa		
6	PANE BRI	First Sect	1	2	1	Session 1	1	1	STD	Position	701	Blind Cor	PE07	BOSIO.PA	Paola		
7	PANE BRI	First Sect	1	2	1	Session 1	2	2	GLUTINE	Position	335	A	PE07	BOSIO.PA	Paola		
8	PANE BRI	First Sect	1	2	1	Session 1	4	3	STD	Position	290	Blind Cor	PE07	BOSIO.PA	Paola		
9	PANE BRI	First Sect	1	2	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE07	BOSIO.PA	Paola		
10	PANE BRI	First Sect	1	3	1	Session 1	1	1	STD	Position	701	Blind Cor	PE10	FERRINO.	Daniela		
11	PANE BRI	First Sect	1	3	1	Session 1	3	2	GLUTINE	Position	335	A	PE10	FERRINO.	Daniela		
12	PANE BRI	First Sect	1	3	1	Session 1	2	3	STD	Position	290	Blind Cor	PE10	FERRINO.	Daniela		
13	PANE BRI	First Sect	1	3	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE10	FERRINO.	Daniela		
14	PANE BRI	First Sect	1	4	1	Session 1	1	1	STD	Position	701	Blind Cor	PE21	TRAVAGLI	Marco		
15	PANE BRI	First Sect	1	4	1	Session 1	4	2	GLUTINE	Position	335	A	PE21	TRAVAGLI	Marco		
16	PANE BRI	First Sect	1	4	1	Session 1	2	3	STD	Position	290	Blind Cor	PE21	TRAVAGLI	Marco		
17	PANE BRI	First Sect	1	4	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE21	TRAVAGLI	Marco		
18	PANE BRI	First Sect	1	5	1	Session 1	1	1	STD	Position	701	Blind Cor	PE05	BALOCCO.	Pietro		
19	PANE BRI	First Sect	1	5	1	Session 1	3	2	GLUTINE	Position	335	A	PE05	BALOCCO.	Pietro		
20	PANE BRI	First Sect	1	5	1	Session 1	4	3	STD	Position	290	Blind Cor	PE05	BALOCCO.	Pietro		
21	PANE BRI	First Sect	1	5	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE05	BALOCCO.	Pietro		
22	PANE BRI	First Sect	1	6	1	Session 1	1	1	STD	Position	701	Blind Cor	PE17	RINALDI.(	Gabriele		
23	PANE BRI	First Sect	1	6	1	Session 1	4	2	GLUTINE	Position	335	A	PE17	RINALDI.(	Gabriele		
24	PANE BRI	First Sect	1	6	1	Session 1	3	3	STD	Position	290	Blind Cor	PE17	RINALDI.(	Gabriele		
25	PANE BRI	First Sect	1	6	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE17	RINALDI.(	Gabriele		
26	PANE BRI	First Sect	1	7	1	Session 1	2	1	STD	Position	701	Blind Cor	PE20	TIBALDI.P	Paolo		
27	PANE BRI	First Sect	1	7	1	Session 1	1	2	GLUTINE	Position	335	A	PE20	TIBALDI.P	Paolo		
28	PANE BRI	First Sect	1	7	1	Session 1	4	3	STD	Position	290	Blind Cor	PE20	TIBALDI.P	Paolo		
29	PANE BRI	First Sect	1	7	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE20	TIBALDI.P	Paolo		
30	PANE BRI	First Sect	1	8	1	Session 1	2	1	STD	Position	701	Blind Cor	PE01	ANFOSSI.I	Marzia		
31	PANE BRI	First Sect	1	8	1	Session 1	1	2	GLUTINE	Position	335	A	PE01	ANFOSSI.I	Marzia		
32	PANE BRI	First Sect	1	8	1	Session 1	3	3	STD	Position	290	Blind Cor	PE01	ANFOSSI.I	Marzia		
33	PANE BRI	First Sect	1	8	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE01	ANFOSSI.I	Marzia		
34	PANE BRI	First Sect	1	9	1	Session 1	3	1	STD	Position	701	Blind Cor	PE12	GALLIZIO.	Lorena		
35	PANE BRI	First Sect	1	9	1	Session 1	1	2	GLUTINE	Position	335	A	PE12	GALLIZIO.	Lorena		
36	PANE BRI	First Sect	1	9	1	Session 1	4	3	STD	Position	290	Blind Cor	PE12	GALLIZIO.	Lorena		
37	PANE BRI	First Sect	1	9	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE12	GALLIZIO.	Lorena		
38	PANE BRI	First Sect	1	10	1	Session 1	4	1	STD	Position	701	Blind Cor	PE11	GALLINA.F	Fabio		
39	PANE BRI	First Sect	1	10	1	Session 1	1	2	GLUTINE	Position	335	A	PE11	GALLINA.F	Fabio		
40	PANE BRI	First Sect	1	10	1	Session 1	3	3	STD	Position	290	Blind Cor	PE11	GALLINA.F	Fabio		

PANE BRIOSS TERZO ASSAGGIO_1_ F

Figure 8 Part of the Compusense output of a taste

2.4 Choice of the software layout for the first taste

In order to help the panelist, the software offers the global vision of the attributes on a wheel with categories and subcategories. Indeed, as the sensory vocabulary shows, there are touch attributes and palate attributes. Moreover, the wheel identifies the order of the attributes that each panelist must use for each bread sample. Panelist should start from touch attribute **Elasticità indice pollice** and continue clockwise.

The order of the attributes on the wheel is important: if the panelist has the opportunity to choose the order by himself, if he tests the hardness or the elasticity that need the sample to be entire, he could ruin or break the sample to try cohesiveness.

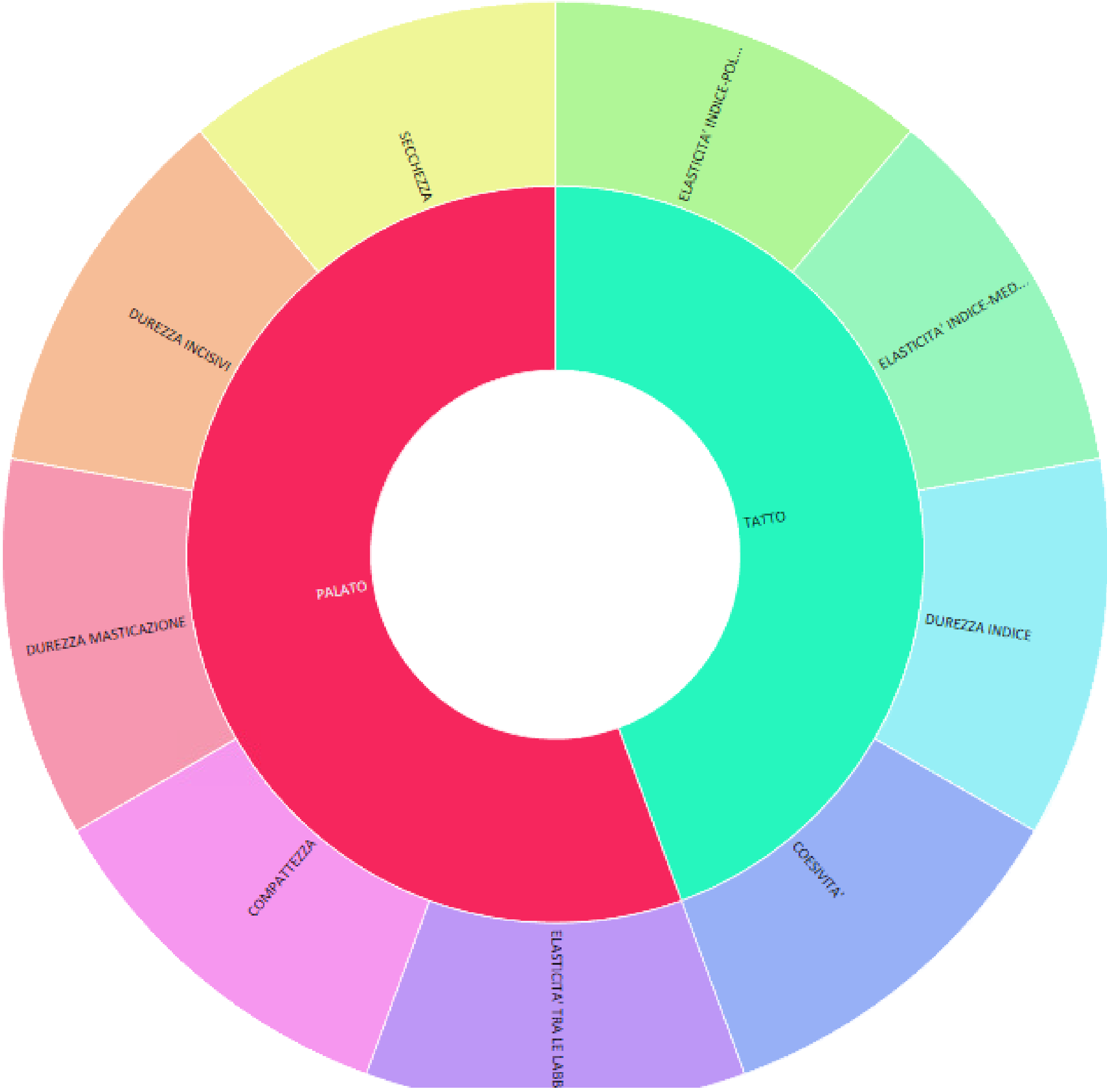
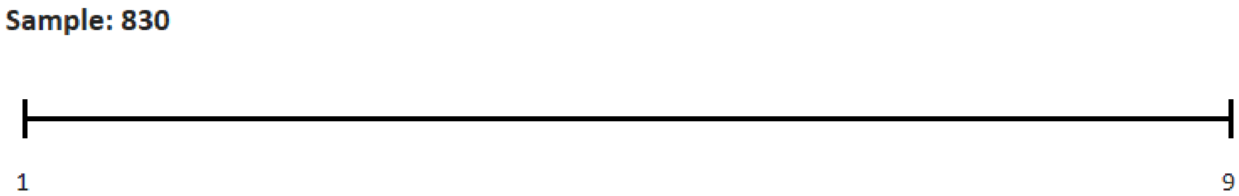


Figure 9 Sensory wheel

After clicking on an attribute, a line appears under the number of the sample.



The panelist can vote by placing the cursor on the line. During the first taste they can't see the number of their vote, they can only see, for each attribute, where they placed the vote for each sample bread; this help them to compare better the samples. Furthermore, the votes are real numbers with two significance figures after the decimal point.

Moreover, in order to explain the attribute with a more complex definition, a description is shown on the computer screen.

3. Bread preparation

In order to better understand the complexity of bread matter, below it is shown the difficulty of its preparation. In the Laboratory of Soremartec two people are needed to prepare 9 kilos of bread dough; the preparation consists of a series of different steps, which need to be followed properly, in order not to weaken attempts a day's work.

First powder mix, monoglyceride solution, saline solution, yeast solution and liquid semi-finished are prepared; it is important to weight with precision every single ingredient.

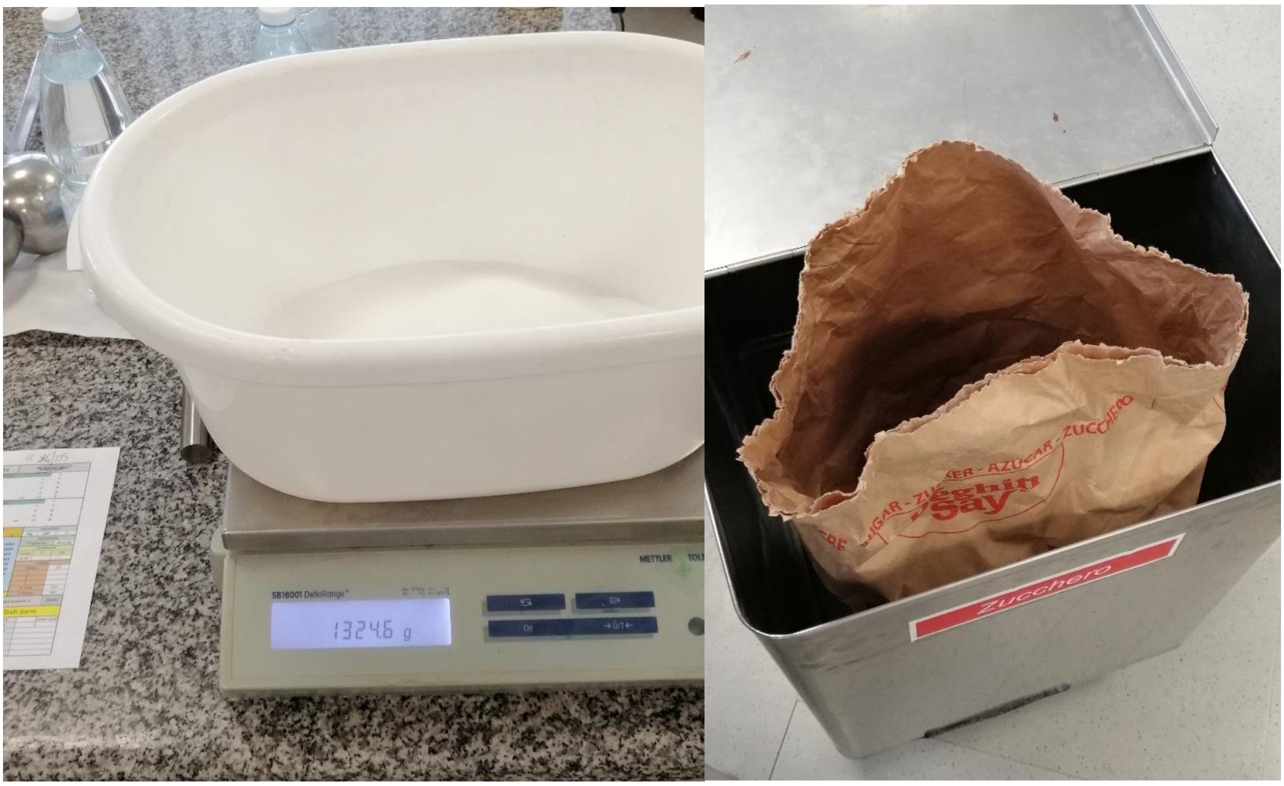


Figure 10 Weight of powder mix

Figure 11 Weight of liquid semi-finisched

Mix and solutions are used later during the cycle of dough. The cycle consists of five different steps in which different portions of the mix and solution are mixed together and put into a dough mixer for a fixed time interval. If these steps are not done properly, the bread might not rise.

Figure 12 Portion for the cycle of dough

Figure 13 Dough mixer

Subsequently, the dough is put in a blast chiller at a temperature of -2 degree Celsius in order to harden it. After that the dough need to be laminated to roll out it.

Figure 14 Lamination

After the lamination, the dough is wound around a tube and then it is spread on a baking pan.

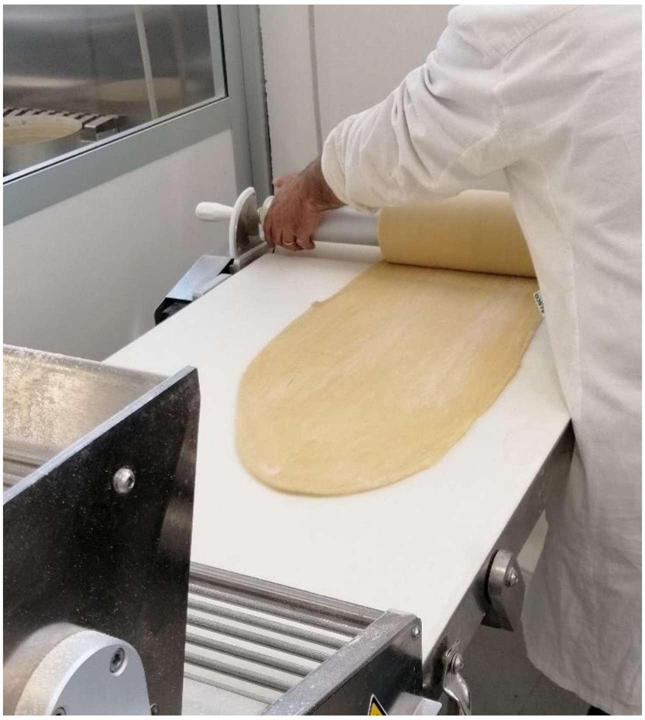


Figure 15 Lamination

Figure 16 Dough is spread on a baking pan

Next, the dough is ready to be put into a prover in order to rise. A tube with a small portion of dough is put in the prover to measure how fast the bread is rising.

Figure 17 Prover

Finally, the bread is baked.

Figure 18 Oven

Figure 19 Cooked bread

Figure 20 Cooked bread

Once the bread is cooked, it needs to cool in order to be cut into small pieces.

Figure 21 Cutting the bread

During the first taste the sample of bread are not shaven, whereas in the other two tastes the samples are shaved (height 25 mm), because of shaving samples gives aligned Texture profile analyzer curves.

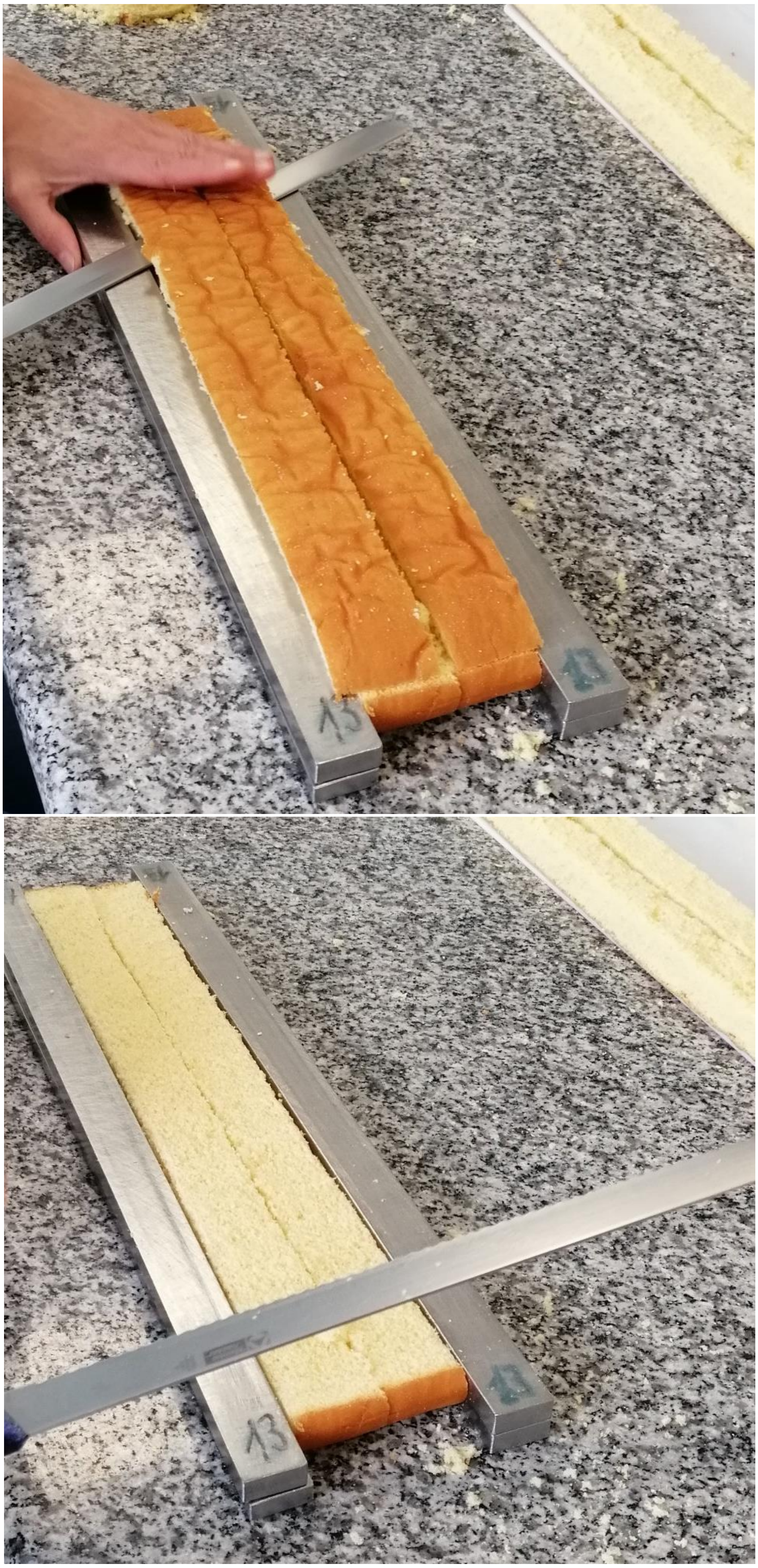


Figure 22 Shaving the bread

Figure 23 Bread samples

Figure 24 Packaging the samples

After each one of the bread samples has been wet with alcohol as food preservative, they need to be packed.

Figure 25 Packaging the samples

4. Tasters (panel)

As already explained, the panel is made up of two groups of people who come to taste the samples, one in the morning and the other in the afternoon; each group is made up of twenty people, females and males, and their age range is between eighteen and sixty years.

These people do not work in Soremartec, indeed the panel is called External Panel. They have been chosen after they presented a CV's to the Sensory Team of Soremartec. The fact that they do not work for Soremartec is very important, because it means that they do not prepare the product that they are going to taste and they are not influenced in their judgment by their work.

They have been trained for six months in order to became professional panel, but they work mainly with chocolate and hazelnut, so it is the first time they taste bread. Moreover, in this experiment, they have to taste rheological attributes, instead of the flavor.

Among the difficulties mentioned before, it must be added the fact that bread is not easy to evaluate: on the one hand, the preparation is very long and made up of a lot of steps and, for example, if a bread stays longer in an oven it becomes harder than the previous one, but they have the same recipe. On the other hand, if a sample bread is opened some minutes before the taste, it starts to lose moisture and becomes dry. As a consequence, it will seem harder than the other sample even if it is effectively not.

Unfortunately, bread is a sensitive matter; in addiction the votes that panelists give are based on subjectivity.

There are two processes during the generation of numeric votes: one is psychological, the person has to convert an energy into a sensation, the other consists in converting the sensation into a number. The second one could be influenced by the scaling.

Dealing with this experiment, scaling allows to collect both ranked data and absolute vote of each sample: this is the reason why in this case, scaling was preferred to rank.

Different types of scaling methods are available:

- category rating
- magnitude estimation
- mark on a line

The first method consists in letting the panelist choose among different, limited and previously agreed categories.

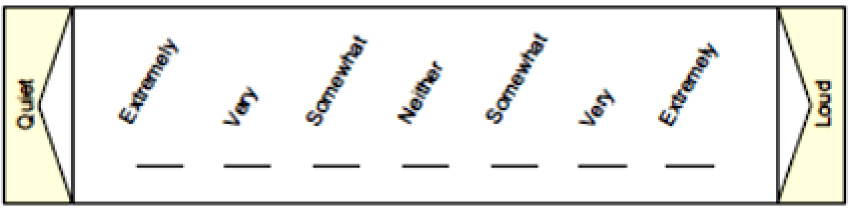


Figure 26 Category scaling

The second and third methods are similar: the panelist is asked to vote on a line using real number. In the case of a magnitude estimation, the line has no superior bound, while in the other case the line has both inferior and superior bound. Marking a line is referred as using a graphic-rating scale and the response is recorded as the distance of the mark from one end of the scale, where usually the end is considered lower. The end-anchor lines may be used to help the panelist avoid end effects that is the reluctance of subjects to use the ends of scale: psychological effects, as already told, should be considered in order to obtain the best sensory data as possible.

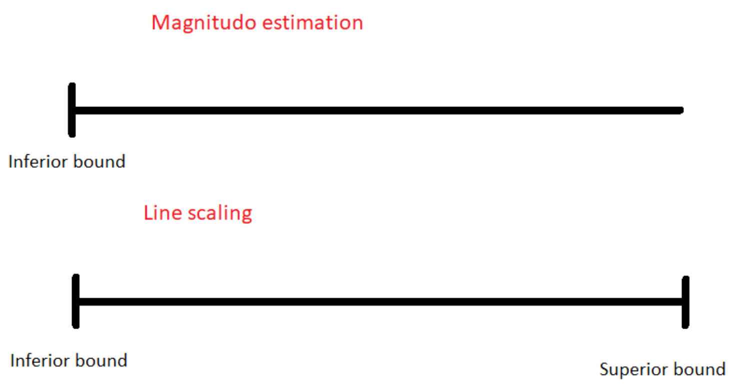


Figure 27 Scaling

Dealing with the experiment, line scaling is chosen to let the panelist have some freedom, but not too much. More precisely, the Nine Point Hedonic Scale is used during the sensory test because of its simplicity, accuracy and precision. The hedonic scale has been accepted by sensory professionals to test consumer preference and acceptability of foods.

In this case, anchors are not used to not influence the panel, but also because there are not samples of bread that can be used as absolutely sample, as already mentioned, bread has variable features that can not be replicated in a perfect way.

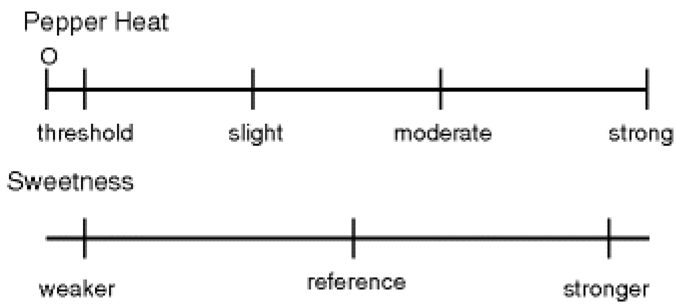


Figure 28 Anchors examples

After concerning the attributes, all the panelists share the same definition: this is very important because, when people deal with food, they use different words to express the same concept. This is the reason why it is important to have a shared sensory vocabulary among the panelists, in order to have a sensory panel who is evaluating the same features.

Before each test, tasters spend half an hour to review the concepts about the attributes and the way to test each sample. After, some of them go in the taste room, where they receive a plate where the sample are located.

Figure 29 Plate

The taste room contains five taste stations, each of them has a computer on which the software Compusense is installed; the panelist has to login using his personal ID and password and then he can start the taste.

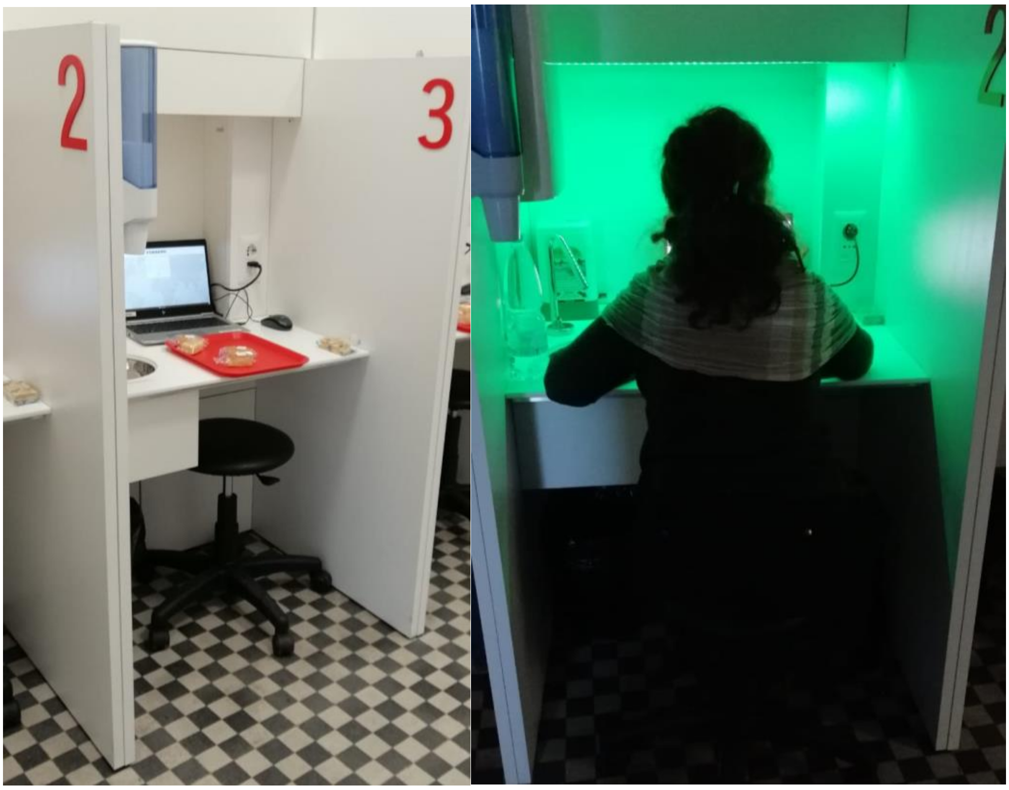


Figure 30 Sensory station

The light in the room is switched off to help the concentration of the panelist, above the computer there are colored lights (neon) that help the taster not to be confused by the color of the sample: for example, if a baking pan is taken out of the oven as the last one it will be more toasted than the first one, but this is inevitable, because it is impossible to take out of the oven all the breads at the same time.

For each group the taste lasts one hour and half, after the conclusion is possible to observe the complete data of the test.

5. Data format

There are two data sources of the experiment:

- TPA data and rheological data
- Sensory data

TPA measurements can be divided in two types: one is a table containing the information named in the first chapter, like hardness, elasticity and cohesiveness.

	A	B	C	D	E
1	**Test ID**	**Durezza**	**Forza 2° picco**	**Elasticità**	**Coesività**
2	Kinder Brioss STD 20	1015	885	0,843	0,513
3	Kinder Brioss STD 2	884	776	0,847	0,554
4	Kinder Brioss STD 3	932	825	0,856	0,531
5	Kinder Brioss STD 4	985	872	0,85	0,552
6	Kinder Brioss STD 5	1053	924	0,867	0,499
7	Kinder Brioss STD 6	999	869	0,845	0,512
8	Kinder Brioss STD 7	906	800	0,849	0,542
9	Kinder Brioss STD 8	1100	953	0,862	0,5
10	Kinder Brioss STD 9	872	766	0,841	0,528
11	Kinder Brioss STD 10	981	859	0,855	0,52
12	Kinder Brioss STD 11	955	841	0,864	0,502
13	Kinder Brioss STD 12	916	802	0,856	0,524
14	Kinder Brioss STD 13	1018	893	0,858	0,523
15	Kinder Brioss STD 14	909	804	0,867	0,579
16	Kinder Brioss STD 16	947	836	0,848	0,514
17	Kinder Brioss STD 17	1014	890	0,869	0,537
18	Kinder Brioss STD 18	1123	976	0,861	0,519
19	Kinder Brioss STD 19	996	873	0,85	0,52
20	Kinder Brioss STD 1	945	834	0,856	0,527
21	Kinder Brioss STD 21	907	797	0,874	0,573

Figure 31 TPA data

The other type deals with the curves of force measured by TPA probe.

	HU	HV	HW	HX	HY	HZ	IA	IB	IC	ID	IE	IF
1	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri	Kinder Bri
2	Force (g)	Distance (	Time (sec)	Force (g)	Distance (	Time (sec)	Force (g)	Distance (	Time (sec)	Force (g)	Distance (	Time (sec)
3	15,5	0	0	12,1	0	0	11,9	0	0	9,1	0	0
4	11	0,005	0,004	9,7	0,005	0,004	9,9	0,005	0,004	9,1	0,005	0,004
5	19,4	0,012	0,008	19,5	0,012	0,008	19,4	0,012	0,008	19,1	0,012	0,008
6	19	0,019	0,012	19,4	0,019	0,012	20,9	0,019	0,012	18,7	0,019	0,012
7	16,4	0,025	0,016	16,1	0,025	0,016	18	0,025	0,016	15,1	0,025	0,016
8	15,4	0,032	0,02	16	0,032	0,02	15,4	0,032	0,02	13,2	0,032	0,02
9	19,1	0,039	0,024	17,9	0,039	0,024	17,8	0,039	0,024	15,9	0,039	0,024
10	19,2	0,046	0,028	17,1	0,046	0,028	15,9	0,046	0,028	16,1	0,046	0,028
11	18,4	0,053	0,032	16,8	0,053	0,032	15,3	0,053	0,032	15,9	0,053	0,032
12	21,8	0,059	0,036	18,5	0,059	0,036	18,5	0,059	0,036	19,4	0,059	0,036
13	20,1	0,066	0,04	16,6	0,066	0,04	17,7	0,066	0,04	16,4	0,066	0,04
14	20,1	0,073	0,044	16,8	0,073	0,044	17,2	0,073	0,044	16	0,073	0,044
15	22	0,08	0,048	18,8	0,08	0,048	18,8	0,08	0,048	19,2	0,08	0,048
16	20,5	0,087	0,052	17,7	0,087	0,052	17,3	0,087	0,052	18,2	0,087	0,052
17	20,3	0,094	0,056	18,1	0,093	0,056	20	0,093	0,056	19,3	0,093	0,056
18	22,3	0,1	0,06	20,4	0,1	0,06	21,8	0,1	0,06	20,5	0,1	0,06
19	22,1	0,107	0,064	20,8	0,107	0,064	20,3	0,107	0,064	19,8	0,107	0,064
20	24,6	0,114	0,068	22,5	0,114	0,068	19,3	0,114	0,068	20,1	0,114	0,068
21	25,5	0,121	0,072	22,7	0,121	0,072	18,4	0,121	0,072	20,3	0,121	0,072

Figure 32 TPA data

Finally, there are the measurement related to the weight loss of the bread samples, after one night in a stove at 105 C ⁰ degrees. The weight loss of the bread measures indirectly the moisture related to the recipe.

Kinder Brioss STD 199									
Decodifica Piatto	Tara Piatto	Lordo Prima	Lordo Dopo	Delta prima	Delta dopo	**Perdita percentuale**	**Media**	Dev St	errore %
1	1,7755	5,681	4,727	3,906	2,952	24,427	24,279	0,236	0,972
2	1,7761	5,433	4,555	3,656	2,779	24,007			
3	1,7883	5,096	4,289	3,307	2,500	24,404			

Figure 33 weight loss data

Then, sensory data are collected by the software and exported in an Excel format: there is a line for each panelist and for each sample having the votes of each attribute.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Test_Nar	Section_N	Section_N	Sample_S	Session_	Session_	Sample_I	Sample_I	Sample_I	Design_P	Blinding_	Sample_T	Panelist_	Panelist_	Panelist_	Panelist_	Q1_
2	PANE BRI	First Sect	1	1	1	Session 1	1	1	STD	Position	701	Blind Cor	PE14	CIMMA.EL	Elisa		
3	PANE BRI	First Sect	1	1	1	Session 1	2	2	GLUTINE	Position	335	A	PE14	CIMMA.EL	Elisa		
4	PANE BRI	First Sect	1	1	1	Session 1	3	3	STD	Position	290	Blind Cor	PE14	CIMMA.EL	Elisa		
5	PANE BRI	First Sect	1	1	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE14	CIMMA.EL	Elisa		
6	PANE BRI	First Sect	1	2	1	Session 1	1	1	STD	Position	701	Blind Cor	PE07	BOSIO.PA	Paola		
7	PANE BRI	First Sect	1	2	1	Session 1	2	2	GLUTINE	Position	335	A	PE07	BOSIO.PA	Paola		
8	PANE BRI	First Sect	1	2	1	Session 1	4	3	STD	Position	290	Blind Cor	PE07	BOSIO.PA	Paola		
9	PANE BRI	First Sect	1	2	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE07	BOSIO.PA	Paola		
10	PANE BRI	First Sect	1	3	1	Session 1	1	1	STD	Position	701	Blind Cor	PE10	FERRINO.	Daniela		
11	PANE BRI	First Sect	1	3	1	Session 1	3	2	GLUTINE	Position	335	A	PE10	FERRINO.	Daniela		
12	PANE BRI	First Sect	1	3	1	Session 1	2	3	STD	Position	290	Blind Cor	PE10	FERRINO.	Daniela		
13	PANE BRI	First Sect	1	3	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE10	FERRINO.	Daniela		
14	PANE BRI	First Sect	1	4	1	Session 1	1	1	STD	Position	701	Blind Cor	PE21	TRAVAGLI	Marco		
15	PANE BRI	First Sect	1	4	1	Session 1	4	2	GLUTINE	Position	335	A	PE21	TRAVAGLI	Marco		
16	PANE BRI	First Sect	1	4	1	Session 1	2	3	STD	Position	290	Blind Cor	PE21	TRAVAGLI	Marco		
17	PANE BRI	First Sect	1	4	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE21	TRAVAGLI	Marco		
18	PANE BRI	First Sect	1	5	1	Session 1	1	1	STD	Position	701	Blind Cor	PE05	BALOCCO.	Pietro		
19	PANE BRI	First Sect	1	5	1	Session 1	3	2	GLUTINE	Position	335	A	PE05	BALOCCO.	Pietro		
20	PANE BRI	First Sect	1	5	1	Session 1	4	3	STD	Position	290	Blind Cor	PE05	BALOCCO.	Pietro		
21	PANE BRI	First Sect	1	5	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE05	BALOCCO.	Pietro		
22	PANE BRI	First Sect	1	6	1	Session 1	1	1	STD	Position	701	Blind Cor	PE17	RINALDI.(	Gabriele		
23	PANE BRI	First Sect	1	6	1	Session 1	4	2	GLUTINE	Position	335	A	PE17	RINALDI.(	Gabriele		
24	PANE BRI	First Sect	1	6	1	Session 1	3	3	STD	Position	290	Blind Cor	PE17	RINALDI.(	Gabriele		
25	PANE BRI	First Sect	1	6	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE17	RINALDI.(	Gabriele		
26	PANE BRI	First Sect	1	7	1	Session 1	2	1	STD	Position	701	Blind Cor	PE20	TIBALDI.P	Paolo		
27	PANE BRI	First Sect	1	7	1	Session 1	1	2	GLUTINE	Position	335	A	PE20	TIBALDI.P	Paolo		
28	PANE BRI	First Sect	1	7	1	Session 1	4	3	STD	Position	290	Blind Cor	PE20	TIBALDI.P	Paolo		
29	PANE BRI	First Sect	1	7	1	Session 1	3	4	PURAFARI	Position	846	Sample	PE20	TIBALDI.P	Paolo		
30	PANE BRI	First Sect	1	8	1	Session 1	2	1	STD	Position	701	Blind Cor	PE01	ANFOSSI.I	Marzia		
31	PANE BRI	First Sect	1	8	1	Session 1	1	2	GLUTINE	Position	335	A	PE01	ANFOSSI.I	Marzia		
32	PANE BRI	First Sect	1	8	1	Session 1	3	3	STD	Position	290	Blind Cor	PE01	ANFOSSI.I	Marzia		
33	PANE BRI	First Sect	1	8	1	Session 1	4	4	PURAFARI	Position	846	Sample	PE01	ANFOSSI.I	Marzia		
34	PANE BRI	First Sect	1	9	1	Session 1	3	1	STD	Position	701	Blind Cor	PE12	GALLIZIO.	Lorena		
35	PANE BRI	First Sect	1	9	1	Session 1	1	2	GLUTINE	Position	335	A	PE12	GALLIZIO.	Lorena		
36	PANE BRI	First Sect	1	9	1	Session 1	4	3	STD	Position	290	Blind Cor	PE12	GALLIZIO.	Lorena		
37	PANE BRI	First Sect	1	9	1	Session 1	2	4	PURAFARI	Position	846	Sample	PE12	GALLIZIO.	Lorena		
38	PANE BRI	First Sect	1	10	1	Session 1	4	1	STD	Position	701	Blind Cor	PE11	GALLINA.F	Fabio		
39	PANE BRI	First Sect	1	10	1	Session 1	1	2	GLUTINE	Position	335	A	PE11	GALLINA.F	Fabio		
40	PANE BRI	First Sect	1	10	1	Session 1	3	3	STD	Position	290	Blind Cor	PE11	GALLINA.F	Fabio		

PANE BRIOSS TERZO ASSAGGIO_1_ F

Figure 34 Sensory data

More specifically, sensory data can not have outlier data, or better said, it is impossible to classify a vote as outlier. Conversely, it is possible to classify a panelist as not precise, but during the experiment all the panelists are presumed to be valid because they were trained.

Otherwise, TPA instrument registered some curves, and so it is possible to collect data that are not reasonable; no preprocessing of data was needed because, in this case, the curves were obviously wrong, so they were deleted during the measurement and a new one sample was taken, in order to replace it.

6. Theory

Kolmogorov–Smirnov test

Kolmogorov–Smirnov test is a nonparametric test of the equality of continuous (or discontinuous) one-dimensional probability distributions; it can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test).
The Kolmogorov–Smirnov statistic measures the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution: the empirical distribution function F_n for n iid ordered observations X_i is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,x]}(X_i)$$

where $I_{[-\infty,x]}(X_i)$ is the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise.

For a given cumulative distribution function $F(x)$, the Kolmogorov–Smirnov statistic is:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $\sup_x$ is the supremum function. By the Glivenko–Cantelli theorem, D_n converges to 0 almost surely in the limit when n goes to infinity if the sample comes from distribution $F(x)$.

Glivenko-Cantelli Theorem

Assume that $X_1, X_2, \ldots$ are independent and identically-distributed random variables, $X_i \in R$, with cumulative distribution function $F(x)$ in common. The empirical distribution function for is defined by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,x]}(X_i)$$

where $I_{[-\infty,x]}(X_i)$ is the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise.

$F_n(x)$ is a sequence of random variables which converge to $F(x)$ almost surely by the strong law of large numbers for every (fixed) x: so F_n converges to F pointwise.

Glivenko and Cantelli proved uniform convergence of F_n to F :

$$||F_n - F||_\infty = \sup_{x \in R} |F_n(x) - F(x)| \rightarrow 0 \text{ almost surely.}$$

Two sample Kolmogorov–Smirnov test

The two-sample K–S test is a general nonparametric method for comparing two samples of underlying one-dimensional probability distributions; it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.
In this case, the Kolmogorov–Smirnov statistic is:

$$D_{n,m} = \sup_x | F_{1,n}(x) - F_{2,m}(x)|$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the first and the second sample respectively, and $\sup_x$ is the supremum function: The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution functions of two samples; the null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution.

The two-sample test checks whether the two data samples come from the same distribution, but this does not specify what that common distribution is. The Kolmogorov–Smirnov test it is devised to be sensitive against all possible types of differences between two distribution functions, but it needs that the number of samples must be large.

In the thesis only the two sample K-S test is used to test if sensorial grades and TPA measurements of different recipes are statistical different or not.

The MATLAB function for two sample K-S test is:

h = kstest2(x1,x2)

It returns a test decision for the null hypothesis that the data in vectors x1 and x2 are from the same continuous distribution. The alternative hypothesis is that x1 and x2 are from different continuous distributions. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise.

Spearman’s rank correlation coefficient

Rank correlation is a statistical dependence between the rankings of two variables.

Spearman's rank correlation coefficient or Spearman's rho, often denoted ρ, is a nonparametric measure of rank correlation and it is appropriate for both continuous and discrete ordinal variables. It measures how well the relationship between two variables can be described using a monotonic function.

The Spearman’s correlation coefficient is described as nonparametric for two reasons: first, a perfect Spearman correlation results when X and Y are related by any monotonic function, not only linear function (contrary to Pearson correlation). Secondly its exact sampling distribution can be obtained without requiring knowledge of the joint probability distribution of X and Y.

The sign of the Spearman correlation indicates the direction of association between the variables X and Y: if Y tends to decrease when X increases, the Spearman’s correlation coefficient is negative, while if Y tends to increase when X increases, the Spearman’s correlation coefficient is positive. A Spearman correlation of zero indicates that there is no connection between X and Y. Spearman correlation increases in magnitude as X and Y become closer to being perfect monotone functions of each other and, when X and Y are perfectly monotonically related, the Spearman correlation coefficient becomes 1.

A perfect monotone increasing relationship implies that, for any two pairs of data values X_i , X_j and Y_i, Y_j, $X_i - X_j$ and $Y_i - Y_j$ always have the same sign, while a perfect monotone decreasing relationship implies that these differences always have opposite signs.

Pearson correlation coefficient, given a pair of random variables X, Y is:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_X \sigma_y}$$

where cov is the covariance, σ_X is the standard deviation of variables X and σ_y is the standard deviation of Y.

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the rank variables. For a sample of size n , the n raw scores X_i, Y_i are converted to ranks rgX_i, rgY_i, and ρ is computed from:

$$\rho_{rg_x, rg_y} = \frac{cov(rg_x, rg_y)}{\sigma_{rg_X} \sigma_{rg_y}}$$

Kendall’s rank correlation coefficient

Kendall’s rank correlation coefficient, often denoted by τ, is a statistic used to measure the ordinal association between two measured quantities, like Spearman’s rank correlation coefficient does.

Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be a be a set of observations of the joint random variables X and Y respectively, such that all the values of (x_i, y_i) and (x_j, y_j), where $i < j$ are said to be concordant if the ranks for both elements agree (if both $x_i > x_j$ and $y_i > y_j$,or if both $x_i < x_j$ and $y_i < y_j$), otherwise they are said discordant.

The Kendall's correlation coefficient is defined as follow:

$$\tau = \frac{2[(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)]}{n(n-1)}$$

The coefficient must be in the range $-1 \le \tau \le 1$ because of the denominator is the total number of pair combinations.

The Kendall correlation between two variables will be high when observations have a similar or identical rank between the two variables, and low vice versa. Like Spearman coefficient, if the disagreement between the two rankings is perfect the coefficient has value −1, while if the agreement between the two rankings is perfect the coefficient has value 1. If X and Y are independent, then we would expect the coefficient to be approximately zero.

Spearman's and Kendall's rank correlation test

A tau test is a non-parametric hypothesis test for statistical dependence based on the tau coefficient, while a rho test is a non-parametric hypothesis test for statistical dependence based on the rho coefficient.

The R function that test association between paired samples, using one of Pearson's product moment correlation coefficient, Kendall's tau or Spearman's rho is:

cor.test(x, y, method, exact)

where **x** and **y** are the numeric vectors of data values (they must have the same length), **method** is a character string indicating which correlation coefficient is to be used for the test among Pearson, Kendall or Spearman and **exact** is a logical indicating whether an exact p-value should be computed for Kendall and Spearman coefficient (due to the small number of samples collected during the thesis, the continuity correction is not necessary).

Functional data analysis

Functional data commonly are curves of the form:

$$x_n(t_{j,n}) \in R \,, t_{j,n} \; \in [T_1, T_2] \,, n = 1,2,3 \ldots N \,, j = 1,2,3 \ldots J_n \,.$$

So there are N curves are observed on a common interval of time $[T_1, T_2]$; the values of the curves are never known at all points $t \in [T_1, T_2]$, but they are available only at some specific point $t_{j,n}$, which can be different for different curves x_n. In our case the instrument Texture Analyzer records the force measured by the probe every 0,04 seconds, so $t_{j,n}$ are equals for every curves.

We focus on situations where the number of points per curve is large. In this case, a fundamental idea of functional data analysis is that the objects we wish to study are smooth curves

$$x_n(t_{j,n}) \in R \,, t_{j,n} \; \in [T_1, T_2] \,, n = 1,2,3 \ldots N \,, j = 1,2,3 \ldots J_n \,.$$

for which the values x_n exist at any point t, but are observed only at selected points $t_{j,n}$.

First of all, we express the curves $x_n(t_{j,n})$ by means of basis expansion:

$$X_N(t) = \sum_{m=1}^{M} c_{m,n} \; B_m(t) \,, 1 \leq n \; \leq N$$

The B_m are some standard collection of basis functions, like splines wavelets, or sine and cosine functions.

This representation reflects the idea that the data are observations from smooth functions that share some shape properties, and so can be approximated as linear combinations of some M basic shapes B_m with M typically smaller than the number of observed time points J_n. If the number of points $t_{j,n}$ is very large, as in the curves of the Texture Analyzer force measurement, expansion succeeds in replacing the original scalar data $x_n(t_{j,n})$ by a smaller collection of the coefficients $c_{m,n}$: in fact, for each n, the curve $x_n$is represented by the column vector $c_n = [\, c_{n1}, c_{n2}, \ldots c_{nM}]$ of dimension M.

R package fda

R implements a package fda that allows to work with functional object in a easier way. Function like:

create.bspline.basis(rangeval, nbasis)

let us to convert our vector $[x_n(t_{j,n}), t_{j,n}]$, that represents a curve, in a functional object using different types of basis B_m, in this example we choose BSpline. **Rangeval** is a vector of length 2 containing the lower and upper boundaries of the range over which the basis is defined, while **nbasis** is the number of basis that we want to use.

A spline of order n is a piecewise polynomial function of degree $n - 1$ in a variable x. In order to build a Bspline we need the knots $[x_i, t_i]$ sorted in non decreasing order, they are the values of x

where the pieces of polynomial meet. For a given sequence of know the in a unique spline $B_{i,n}(x)$, up to scaling factor, satisfying:

$$B_{i,n}(x) = \begin{cases} 0 & if\ x \le t_i \quad or \quad x \ge t_{i+n} \\ \neq 0 & otherwise \end{cases}$$

For each finite knot interval where $B_{i,n}(x)$ is non-zero, the B-spline is a polynomial of degree $n - 1$, so it is a continuous function at the knots and when all knots are distinct, as they are for the Texture Analyzer curves, its derivatives are also continuous up to the derivative of degree $n - 1$.

There are other possible types of basis in fda package:

- Constant basis
- Exponential basis
- Fourier basis
- Monomial basis
- Polygonal basis
- Power basis

Bspline are used in the thesis for many reasons: Texture Analyzer curves are neither signal, so Fourier basis it cannot be the best choice, nor function having exponential behavior, so we choose Bspline because of their polynomial behavior (continuity and differentiability).

Before converting curves to R functional object, we need to choose ne right number of basis:

function.fd=create.bspline.basis(rangeval, nbasis)

function.fd$gcv

The second line of code shows the value of the generalized cross-validation or GCV criterion. If there are multiple curves, this is a vector of values, one per curve.

$$gcv = \frac{n * SSE}{(n - df)^2}$$

where n is the dimension of the input curve, SSE is the error sums of squares and it is a vector of the same size as gcv and df a degrees of freedom measure of the smooth. So GCV help to choose the right number of bases.

After choosing the bases and their number, we can convert the curve in a functional object using the following code:

smooth.basis(y, fdParobj)

where **y** is the vector of measurements and **fdParaobj** is the object made up of the bases.

After the raw data have been converted to functional objects, the simplest summary statistics are the pointwise mean and the pointwise standard deviation:

$$\bar{X}_N(t) = \frac{1}{N}\sum_{n=1}^{N} X_n(t)$$

$$SD_x(t)^2 = \frac{1}{N-1}\sum_{n=1}^{N} (\, X_n(t) - \bar{X}_N(t)\,)^2$$

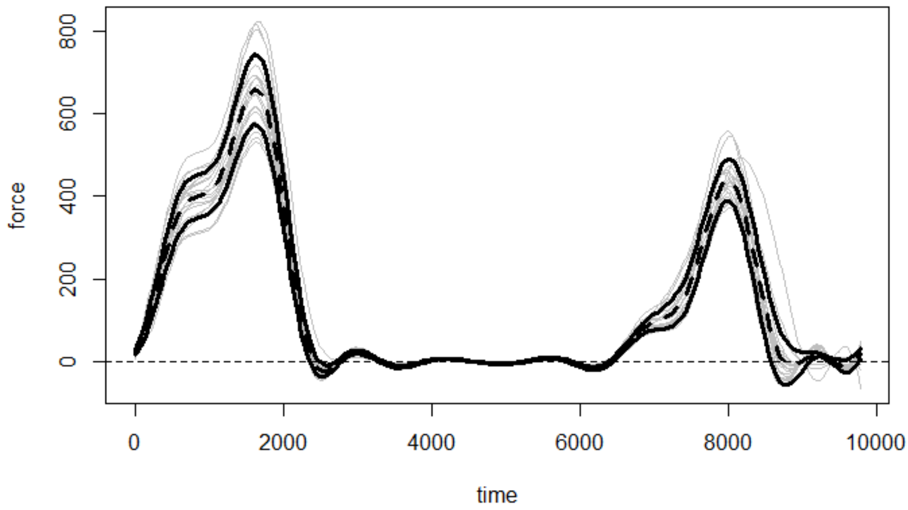


Figure 35 - An example of pointwise sample mean and standard deviation of 20 different curves

Two sample testing problem with permutation test

We observe two sample curves, for example two curves related to two different bread recipes, defined on the same interval $X_1, X_2, \ldots X_N$ and $X_{N+1}, X_{N+2}, \ldots X_{N+M}$. The first N curves are iid draws from a population with mean function μ_1 and the last M curves are iid draws from a population with mean function μ_2.

The function **t.perm(f1,f2,nperm)** test:

$$H_0 : \mu_1(t) = \mu_2(t), \text{t} \in [0, T]$$

Where **f1** and **f2** are functional object and **nperm** is the number of permutations used to build the test statics.

The test uses the statistics

$$T_{sup} = \sup_{t\in[0,T]} T(t)$$

$$T(t) = \frac{\bar{X}_N(t) - \bar{X}_M(t)}{(\, N^{-1}V_N(t) + M^{-1}V_M(t)\,)^{1/2}}$$

$$V_N(t)^2 = \frac{1}{N-1}\sum_{n=1}^{N} (\, X_n(t) - \bar{X}_N(t)\,)^2$$

$$V_M(t)^2 = \frac{1}{M-1} \sum_{n=N+1}^{M+N} (X_n(t) - \bar{X}_N(t))^2$$

The null hypothesis is rejected if T_{sup}(the normalized difference between the sample mean functions) is large for some t. The distribution under H_0 is approximated as follows: denote by π one of the $(N+M)!$ permutations of the index set $1,2, \dots N, N+1, \dots, N+M$. If H_0 holds the two samples $X_{\pi(1)}, \dots, X_{\pi(N)}$ and $X_{\pi(N+1)}, \dots, X_{\pi(N+M)}$ have the same mean; the empirical distribution of the $(N+M)!$ values of T_{sup}^{π} is used as an approximation to the null distribution of the test statistics. **nperm** gives the dimension of the subset of all the possible $(N+M)!$ permutations that the test is going to use.

The R function gives as output:

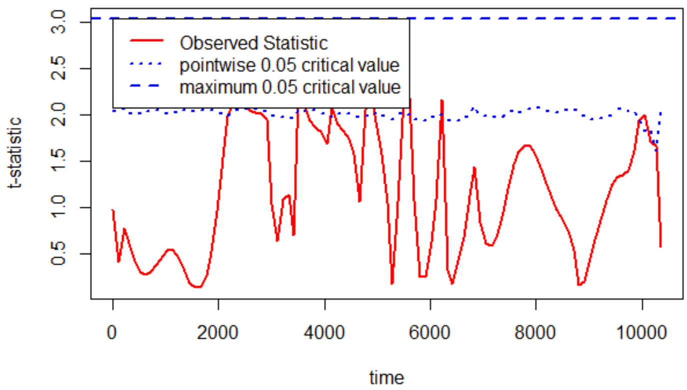


Figure 36 - Example 1

In the figure of the first example we can observe that the statistics never overcomes the maximum critical values, so the curves don't are not significantly different.

In the figure of the second example we observe that the statistics overcomes more than one time the maximum critical values, so the curves are significantly different.

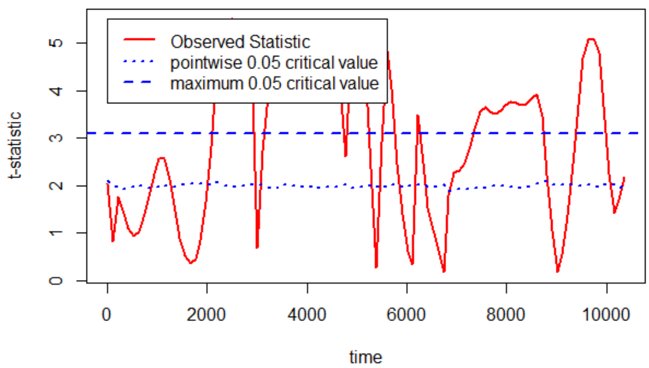


Figure 37- Example 2

Pearson's chi-squared test

Suppose to have n observations of random sample from a population that are classified into k mutually exclusive classes with observed frequencies x_i for $i = 1, \dots, k$, moreover a null hypothesis gives the probability p_i that an observation belongs to the ith class.

So, we have the expected numbers $m_i = np_i$ for all i, where

$$\sum_{i=1}^{k} p_i = 1$$

$$\sum_{i=1}^{k} m_i = n \sum_{i=1}^{k} p_i = \sum_{i=1}^{k} x_i$$

Pearson says that if the null hypothesis is correct, the limiting distribution of the statistics

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i} = \sum_{i=1}^{k} \frac{x_i^2}{m_i} - n$$

is the χ2 distribution as $n \to \infty$.

For a fixed sample size, greater differences $(x_i - m_i)$ produce larger X^2 values and stronger evidence against the null hypothesis. The p-value is the probability that, under H_0, X^2 is at least large as the observed value. The chi-squared approximation improves as m_i increase, usually $m_i \geq 5$ is sufficient for a decent approximation. The chi-squared distribution is concentrated over nonnegative values, with mean equal to its degree of freedom df and standard deviation $\sqrt{2df}$; when df increases, the distribution concentrates around larger values an is more spread out.

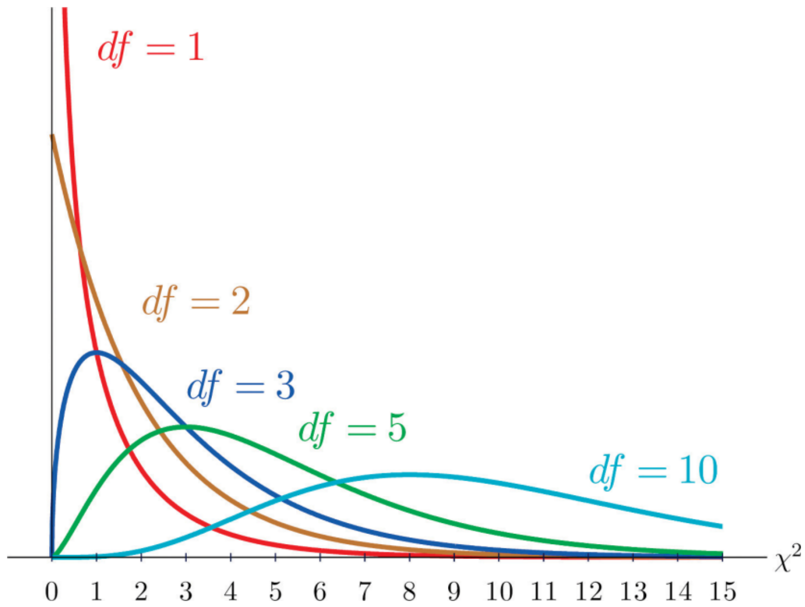


Figure 38 - Chi-squared distributions

Contingency tables

A contingency table is a special type of frequency distribution table, where two variables, row and column, are shown simultaneously; they are used in statistics to summarize the relationship between several categorical variables

	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

Figure 39 - Example of contingency table

Testing the frequencies of vectors of dimension (n,1)

We are looking for significant differences between the frequencies in a vector of observation.

Chi-squared test tell us if the frequencies are random or not, where we are interested in non random ones.

The R function

chisq.test(data)

performs chi-squared contingency table tests and goodness-of-fit tests., where **data** is the vector that we want to analyze whose entries must be non-negative integers. Because of **data** is a vector, then a goodness-of-fit test is performed (x is treated as a one-dimensional contingency table): in this case, the hypothesis tested is whether the population probabilities equal those in p, or are all equal if p is not given, like in our case.

Testing the frequencies of matrix of dimension (n,n)

We are looking for significant differences between the frequencies in a matrix of observation.

Chi-squared test tell us if the frequencies are random or not, where we are interested in non random ones.

The R function

chisq.test(data)

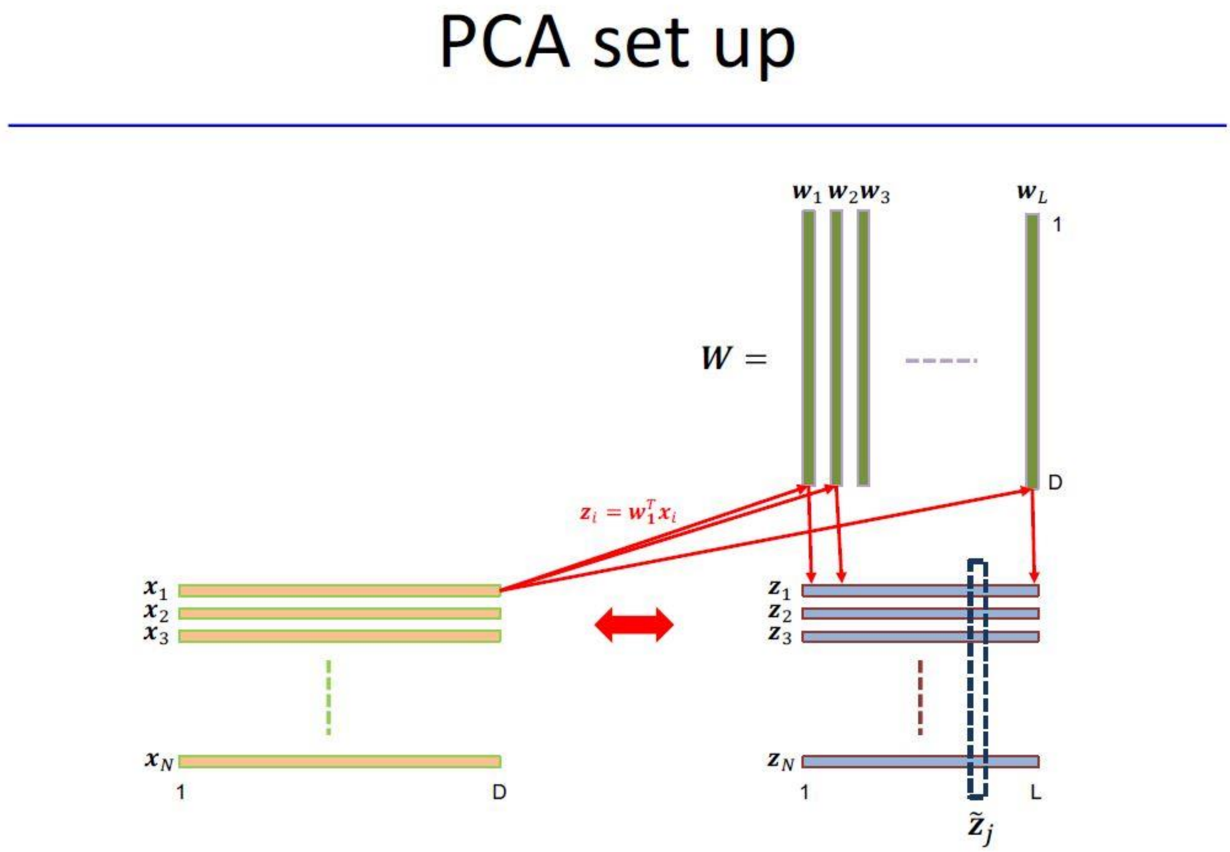
performs chi-squared contingency table tests and goodness-of-fit tests., where **data** is a matrix with at least two rows and columns. **Data** is taken as a two-dimensional contingency table: the entries of x must be non-negative integers and cases with missing values are removed. Then Pearson's chi-squared test is performed of the null hypothesis that the joint distribution of the cell counts in a 2-dimensional contingency table is the product of the row and column marginals.

Principal component analysis

Principal component analysis is an approach to derive a low dimensional set of features from a large set of variables.

PCA is an orthogonal linear transformation that takes the data to anew coordinate system of the in a space which dimension is smaller than the initial dimension.

If a data matrix X (NxD dimension) is considered, which sample mean of each column has been shifted to zero, where each rows is a different repetition of an experiment and each column is a feature of the experiment, the PCA transformation is defined by a set of D-dimensional vectors (W matrix) of coefficients that map each row of X to a new vector of principal component scores (Z matrix).



In the thesis PCA is used not to reduce the number of features but in order to have a one dimensional representation of the frequency table, 4x4 table that will be discussed later, to obtain a rank by the score vector, even when the rank is not so evident.

So only the first principal component direction is used, that is the one along which the observations vary the most.

7. First taste results

Once the sensory taste data and measurements of the Texture Profile Analyzer tool are collected, the recipes are compared in pairs, using sensory data at first and then the instrument data. The non-parameter test of Kolmogorov Smirnov is used to compare the samples (null hypothesis: the samples come from the same population) because, analyzing their histograms, the data are not normal distributed. Indeed, it is not considered plausible in this case to use a parametric test that has normal distribution.

Kolmogorov's test, during the first taste, does not show significant differences between the numerical ratings given by the tasters to the different recipes for each attribute evaluated. The test is also performed among the different attributes of the same field, such as the three different hardness types that are analyzed by the tasters: *Durezza Masticazione*, *Durezza Incisivi* and *Durezza Indice.*

HARDNESS	**A**	**B**	**D**	**C**
D. Masticazione vs D. Incisivi	Statistically Similar	S. Similar	Statistically Different	S. Similar
D. Masticazione vs D. Indice	S. Similar	S. Similar	S. Similar	S. Similar
D. Indice vs D. incisivi	S. Similar	S. Similar	S. Similar	S. Similar

ELASTICITY	**A**	**B**	**D**	**C**
E. Indice e Medio vs E. Labbra	S. Similar	S. Similar	S. Similar	S. Similar
E. Indice e Pollice vs E. Indice e Medio	S. Similar	S. Similar	S. Similar	S. Similar
E. Indice e Medio vs E. Labbra	S. Similar	S. Similar	S. Similar	S. Similar

	A	**B**	**D**	**C**
Coesività vs Compattezza	S. Similar	S. Similar	S. Similar	S. Similar

Recipes	**TPA Hardness**	**Sensory hardness**	**TPA elasticity**	**Sensory elasticity**	**TPA cohesiveness**	**Sensory cohesiveness**
D-B	S. Different	S. Similar	S. Similar	S. Similar	S. Similar	S. Similar
D-A	S. Different	S. Similar	S. Different	S. Similar	S. Different	S. Similar
D-C	S. Different	S. Similar	S. Similar	S. Similar	S. Different	S. Similar
A-B	S. Similar	S. Similar	S. Different	S. Similar	S. Different	S. Similar
A-C	S. Similar	S. Similar	S. Similar	S. Similar	S. Similar	S. Similar
B-C	S. Similar	S. Similar	S. Different	S. Similar	S. Different	S. Similar

Durezza Indice is taken into account as sensory hardness because, since the votes of the sensory hardness are not statistically different among the recipes with significant frequency (see the table of comparisons above), it is clear that logically looked more like the Texture Analyzer's measurement method.

The Kolmogorv test is conducted only on the Hardness and not on the force measured during the second crush (second peak) because they are linearly dependent as you can see below:

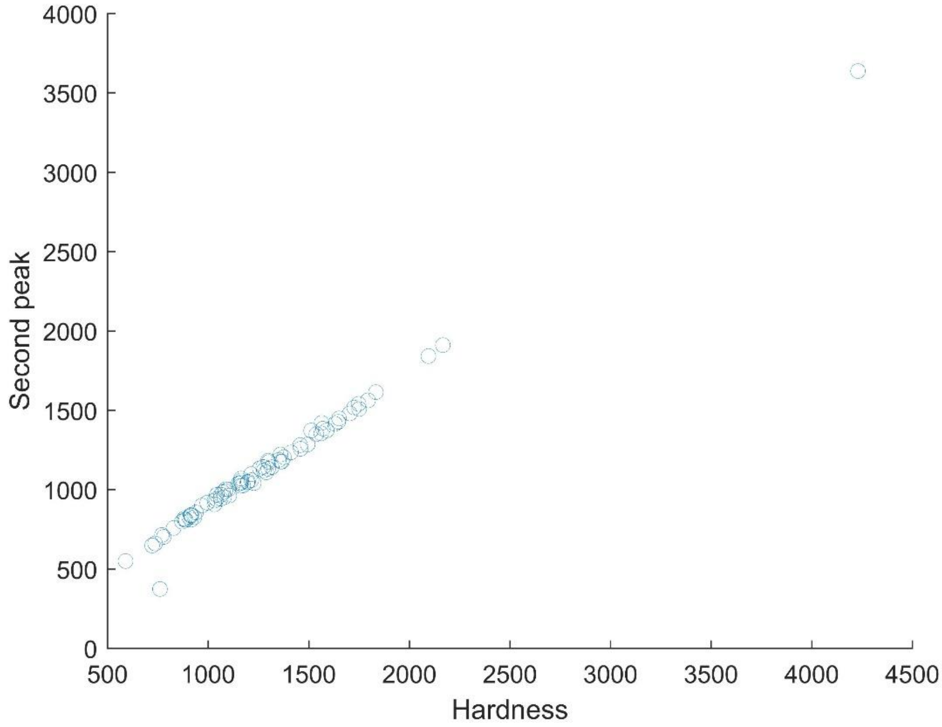


Figura 1 Linear dependence

Since the sensory observations are no significantly different, the rank given by each taster, thanks to the votes expressed during the tasting, is analyzed: during the experiment, in fact, it is taken into account that the subjectivity of the use of a numerical scale. In addition, the tasters are not used to evaluate bread, so often they rate on the average of the scale (it is a common behavior when the panelists are not confident)

The tables below show the frequencies at which the various recipes are ranked as first, second, third and fourth respectively, where the first place indicates greater intensity of the attribute and the fourth place indicates less intensity. The rows of the tables sum up to 33, which is the number of the panel tasters who participated at the tasting.

Elasticità Indice e Pollice	**A**	**C**	**B**	**D**
1^	10	8	10	5
2^	8	8	9	8
3^	7	9	7	10
4^	8	8	7	10

Elasticità Indice e Medio	A	C	B	D
1^	7	10	11	5
2^	8	6	11	8
3^	10	9	5	9
4^	8	8	6	11

Elasticità Labbra	A	C	B	D
1^	9	6	8	10
2^	7	8	8	10
3^	12	7	9	5
4^	5	12	8	8

Coesività	A	C	B	D
1^	5	13	7	8
2^	14	8	5	6
3^	9	5	3	16
4^	5	8	17	3

Compattezza	A	C	B	D
1^	5	11	8	9
2^	8	11	6	8
3^	10	4	7	12
4^	10	7	12	4

Secchezza	A	C	B	D
1^	10	7	7	9
2^	10	12	1	10
3^	7	9	10	7
4^	6	6	15	6

Durezza Indice	A	C	B	D
1^	6	7	8	12
2^	4	13	7	9
3^	15	6	5	7
4^	8	7	13	5

Durezza Masticazione	A	C	B	D
1^	6	7	6	14
2^	13	10	2	8
3^	11	10	5	7
4^	4	6	20	3

Durezza Incisivi	A	C	B	D
1^	10	7	3	13
2^	8	11	2	12
3^	13	10	6	4
4^	3	5	21	4

In some cases the tables allow to see more clearly which recipe is put at the first, second, third and fourth place for each attribute, hence it is possible to start the search for a correlation between sensory voting ranks and the average measurements of the Texture Profile Analyzer.

For each attribute, such as "Elasticity", it is looked for a correlation among the values of the only "Elasticity" attribute measured by the instrument. The search for a correlation between ranks is carried out using the Kendall's correlation test and Spearman's correlation test: they both prove the null hypothesis that there is no correlation between the samples, estimating both Kendall's tau and Spearman's rho.

The ranks are obtained using PCA reduction on the frequency's tables

ATTRIBUTO	PANI	MEDIA TPA	RANK TPA	RANK CLASSIFICA
ELASTICITA_INDICE_POLLICE	A	0,8485	1	3
ELASTICITA_INDICE_POLLICE	C	0,8477	2	2
ELASTICITA_INDICE_POLLICE	B	0,814	4	1
ELASTICITA_INDICE_POLLICE	D	0,8374	3	4
ELASTICITA_INDICE_MEDIO	A	0,8485	1	2
ELASTICITA_INDICE_MEDIO	C	0,8477	2	3
ELASTICITA_INDICE_MEDIO	B	0,814	4	1
ELASTICITA_INDICE_MEDIO	D	0,8374	3	4
ELASTICITA_TRA_LABBRA	A	0,8485	1	1
ELASTICITA_TRA_LABBRA	C	0,8477	2	4
ELASTICITA_TRA_LABBRA	B	0,814	4	2
ELASTICITA_TRA_LABBRA	D	0,8374	3	3
COESIVITA	A	0,63	2	3
COESIVITA	C	0,6462	1	2
COESIVITA	B	0,5473	3	1
COESIVITA	D	0,5422	4	4
COMPATTEZZA	A	0,63	2	1
COMPATTEZZA	C	0,6462	1	2

COMPATTEZZA	B	0,5473	3	4
COMPATTEZZA	D	0,5422	4	3
DUREZZA_INDICE	A	1276,3	2	3
DUREZZA_INDICE	C	1048,9	4	4
DUREZZA_INDICE	B	1141,6	3	1
DUREZZA_INDICE	D	1620,8	1	2
DUREZZA_MASTICAZIONE	A	1276,3	2	2
DUREZZA_MASTICAZIONE	C	1048,9	4	4
DUREZZA_MASTICAZIONE	B	1141,6	3	3
DUREZZA_MASTICAZIONE	D	1620,8	1	1
DUREZZA_INCISIVI	A	1276,3	2	3
DUREZZA_INCISIVI	C	1048,9	4	4
DUREZZA_INCISIVI	B	1141,6	3	2
DUREZZA_INCISIVI	D	1620,8	1	1
SECCHEZZA	A	20,738	3	3
SECCHEZZA	C	21,7	2	4
SECCHEZZA	B	20,325	4	2
SECCHEZZA	D	23,204	1	1

The use of the two methods did not yield significant results on any attribute;

ATTRIBUTE	RHO	p-value	TAU	p-value
Elasticità Indice e Pollice	-0,4	0,6	-0,33	0,4969
Elasticità Indice e Medio	-0,2	0,8	0	1
Elasticità Labbra	0,2	0,8	0	1
Coesività	0,4	0,6	0,33	0,4969
Compattezza	0,6	0,4	0,33	0,4969
Durezza indice	0,4	0,6	0,33	0,4969
Durezza Masticazione	1	$2{,}2 * 10^{-16}$	1	0,04154
Durezza Incisivi	0,8	0,2	0,66	0,1742
Secchezza	-0,4	0,6	-0,33	0,4969

By looking at the rho and tau estimates, correlations are not statistically significant, but in the case of hardness attribute they are all positive: so tasters and tools are evaluating the samples with a concurring growth in sign, although they are not linked by a purely linear relationship. However, in order to obtain a rank on the values of the Texture Profile Analyzer, only the average measurements are considered without taking into account the variance between the samples of a single recipe.

The fact that the rho and tau estimators for *Secchezza* are negative is related to the way they are estimated by weight loss: weight loss, indeed, is the measure that indicates how much water is lost from bread after spending a night in the stove at 105 degrees, that is the humidity of the bread. So, the measure of the water lost during this process is inversely proportional to the feeling of dryness perceived on the palate, which measures how much saliva the bread absorbs.

Dealing with elasticity, since sensory data are not significant, this possible negative correlation is not considered.

Trying to extract information by reducing the frequencies of each attribute to a single unique rank is probably too specific, thus it is taken into account all the information that the observed frequencies tables returned. Then a Chi Quadro test of Pearson on the frequencies is used (null hypothesis: the samples come from the same population of average and variance estimated by the table) which shows that:

ATTRIBUTE	TEST RESULT
Durezza Masticazione	Rejection H0 - Very significant (p-value < 0,02)
Durezza Incisivi	Rejection H0 – Significant (0,02 < p-value < 0,05)
Durezza Indice	Rejection H0 - Very significant
Elasticità Indice e Pollice	H0 is not rejected
Elasticità Indice e Medio	H0 is not rejected
Elasticità Labbra	H0 is not rejected
Secchezza	Rejection H0 - Significant
Coesività	Rejection H0 - Very significant
Compattezza	H0 is not rejected

The test results confirm what can be noticed immediately: the attributes concerning hardness the D recipe has always received more votes at first place, while the B recipe has always received more votes at last place with significant differences in grades (after fixing the attribute and position, the more significant frequencies are highlighted in yellow in the tables and the significance is measured again with a Chi Quadro test). The most significant differences both belong to the attributes of *Durezza al palato,* which is probably the most common one that people face every day compared to *Durezza al tatto*.

Concerning the *Coesività* and *Secchezza* attributes, the B recipe is significantly judged to be the less cohesive and the less dry.

From the frequency tables, therefore, it is noted that the attributes related to the elasticity of the samples are not well understood or are too difficult for tasters to evaluate.

Subsequently, the entire curves (force,time), measured by the Texture profile analyzer, are analyzed instead of considering only the two peaks of maximum hardness and elasticity. The instrument measures curves of (hardness, distance, time), but the distance depends only on the height of the loaves that is not considered important.

The curve is given by a vector of force measurements (grams) that are evaluated every 0.004 seconds by the tool, and, thanks to the **fda** package of the R software, the curve can be approximated by the use of B-Spline. The number of bases to use for each set of curves in each recipe is chosen considering the value of the generalized cross-validation index (GCV).

The curves below show the GCV index for all the possible numbers of basis that can be used: as can be seen from the 20 bases onwards, the index stabilizes its decrease, so a threshold of maximum variability explained by the addition of Bases is reached.

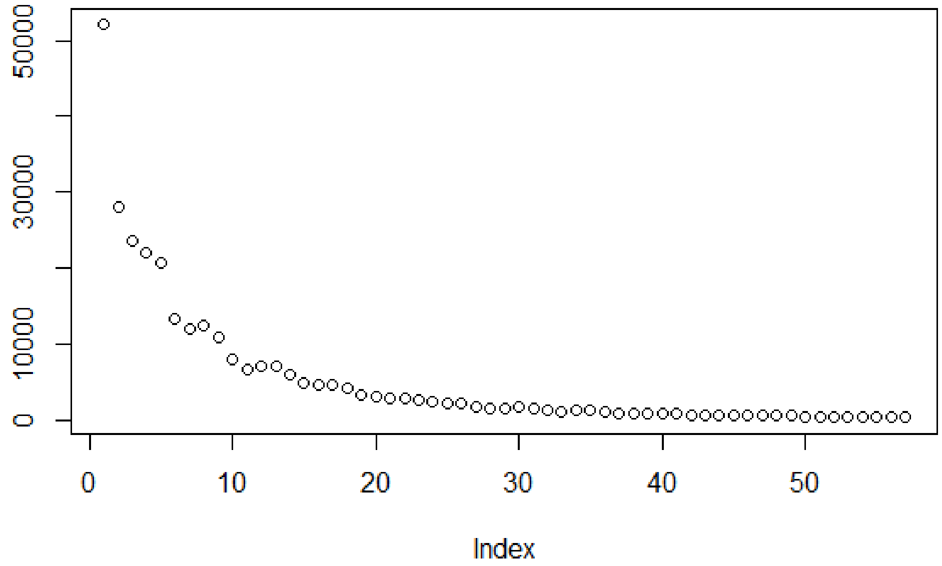


Figure 2 C recipe, index on y label and number of basis on x label

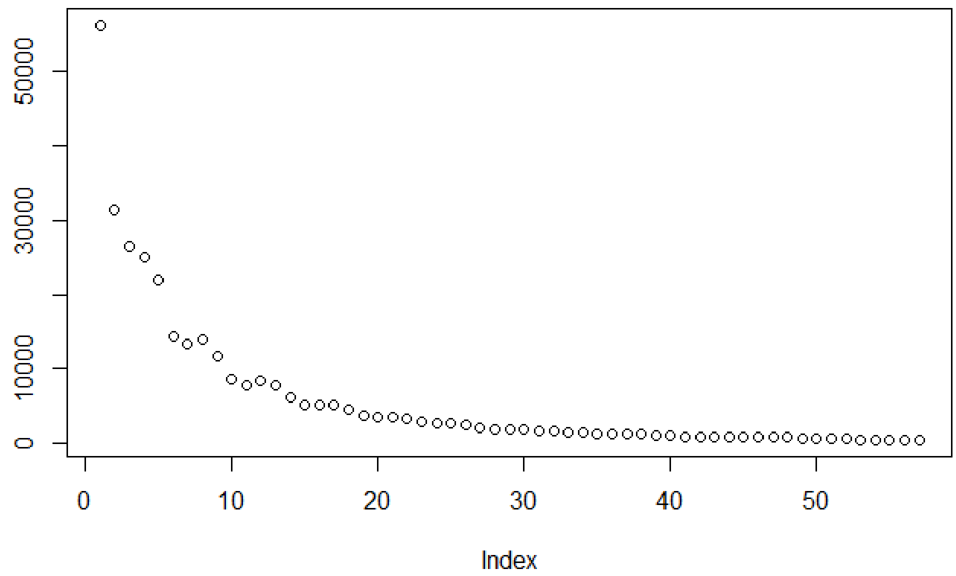


Figure 3 A recipe, index on y label and number of basis on x label

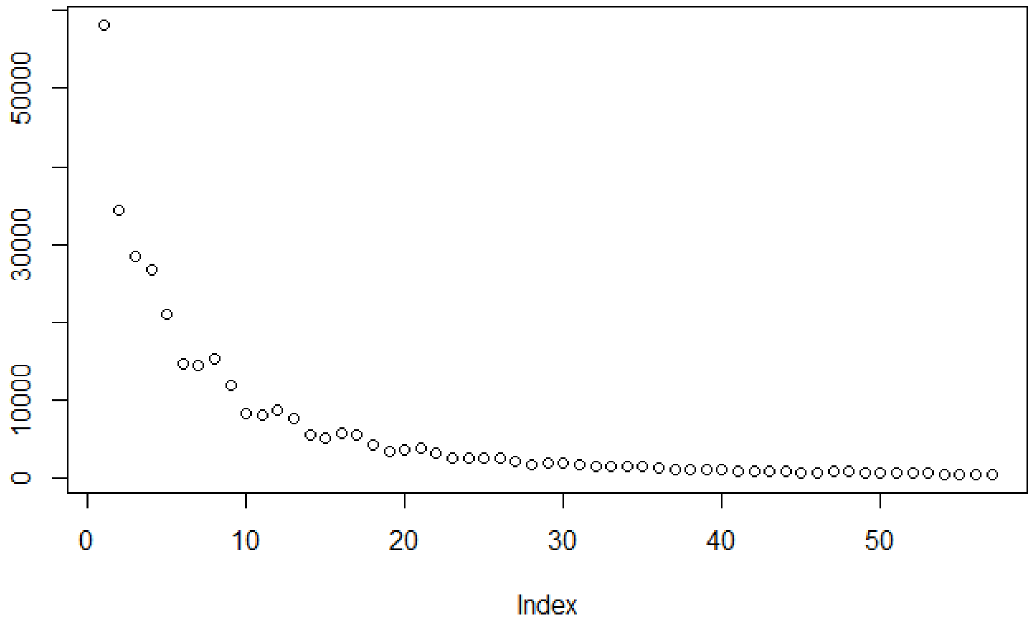


Figure 4 B recipe, index on y label and number of basis on x label

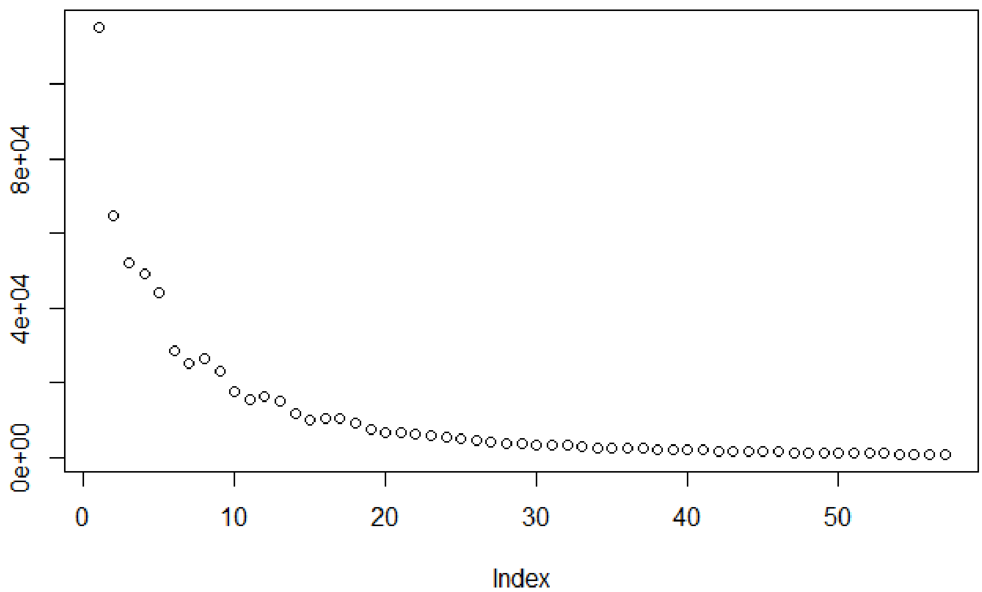


Figure 5 D recipe, index on y label and number of basis on x label

Once the number of bases is chosen, a functional object is created in order to approximate the curves. The figures below show the approximate curves of the measurements made for all the recipes; from the functional object the average point function and the standard time deviation are obtained for each recipe.

In the figure below are represented the average point curve (black dashed line) and the average point curve $\pm$ standard point deviation (thick black lines).

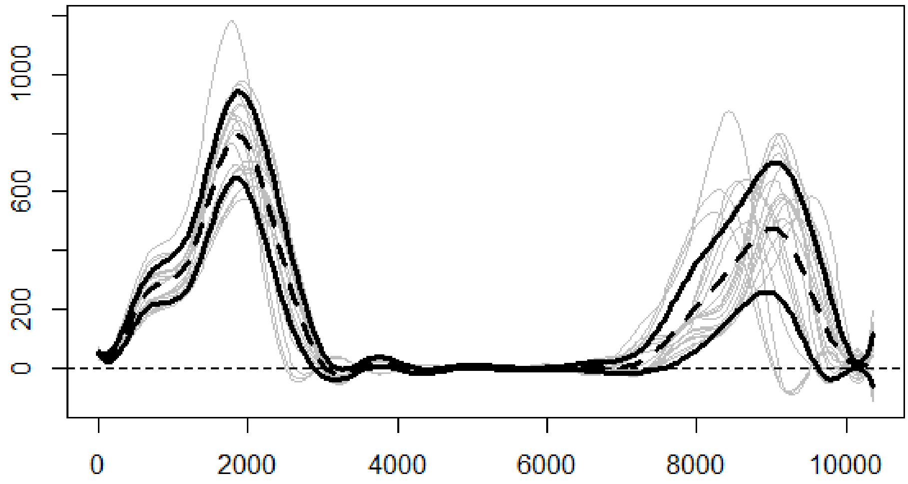


Figure 6 C recipe, force on y label and time on x label

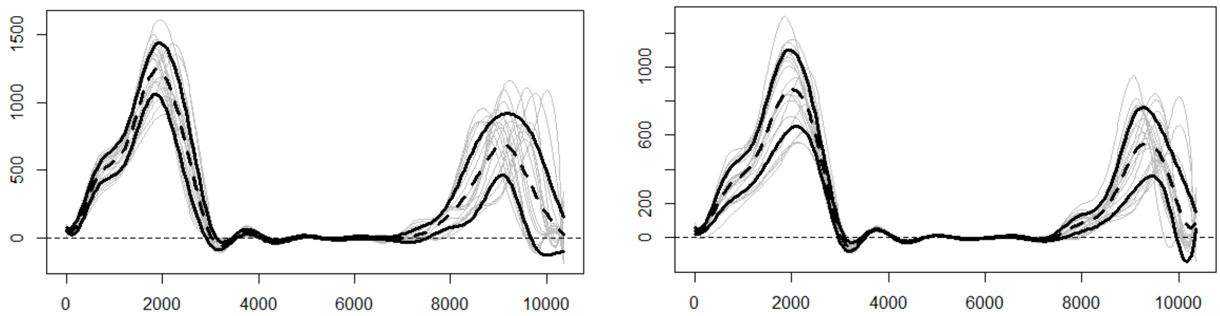


Figure 7 D and B recipes, force on y label and time on x label

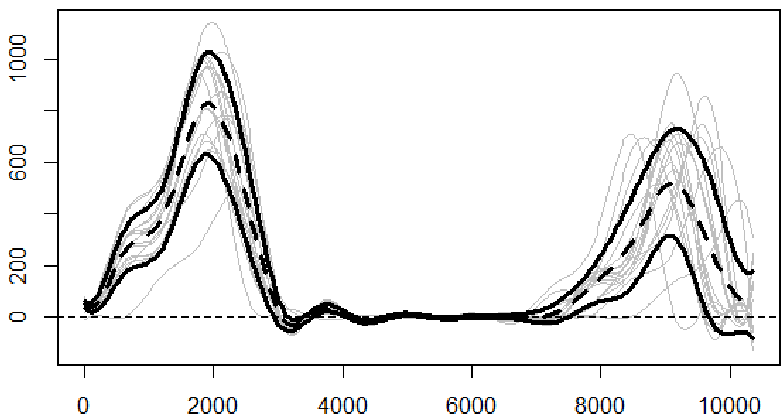


Figure 8 A recipe, force on y label and time on x label

Below the average punctual functions for each recipe are shown:

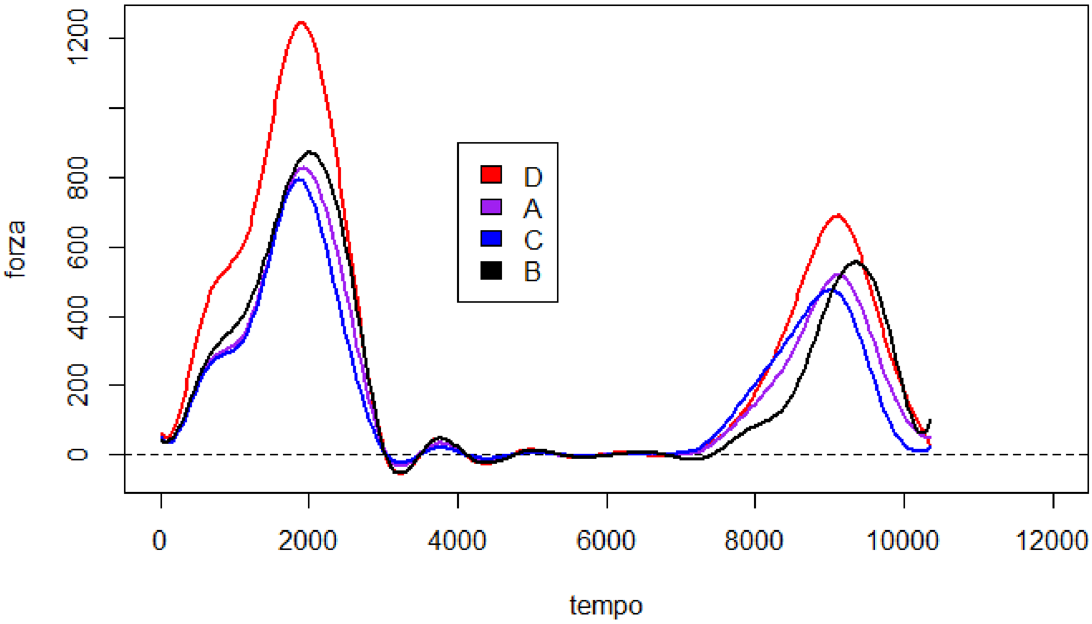
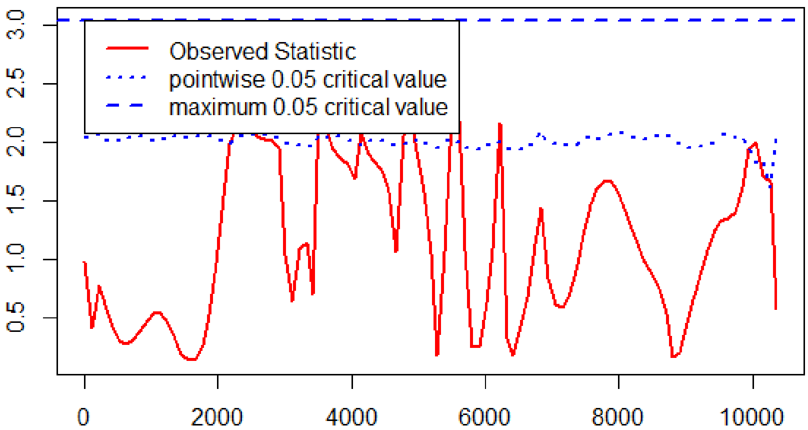


Figure 9 Force on y label and time on x label

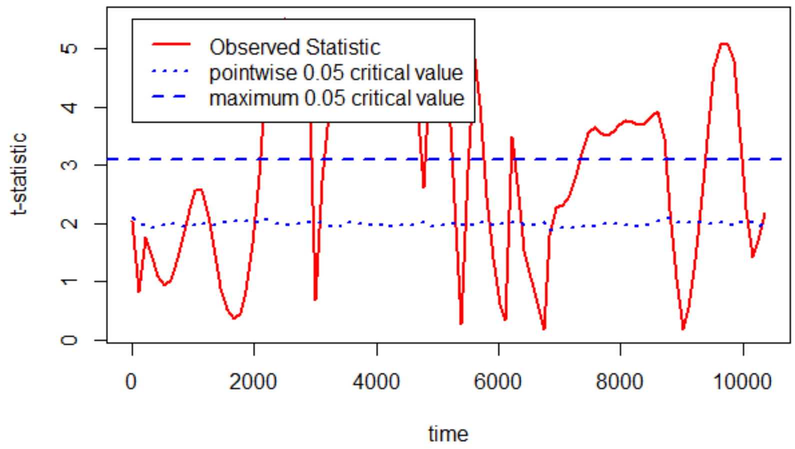
The figure above shows that the D recipe is perceived as the hardest by the instrument, which is in agreement with sensory assessments. The image shows that some curves are truncated at the end: this is due to the fact that, in order to use the test that will be introduced later, the curves must have the same length. As the curve of the C recipe is the one of the smallest lengths, the curves of the other recipes have been truncated on purpose. Indeed, the Texture Profile Analyzer tool stops recording the force a few seconds after returning to the position corresponding to the height of the bread sample. In addition, the curves are not aligned with each other within each recipe: this is due to the fact that the loaves do not have the same height.

The **t.perm** test of the R software's fda package is used to compare the curves of different recipes, (it tests as null hypothesis the fact that the sample averages of normalized recipes are not statistically different). As output the test returns the figures below:

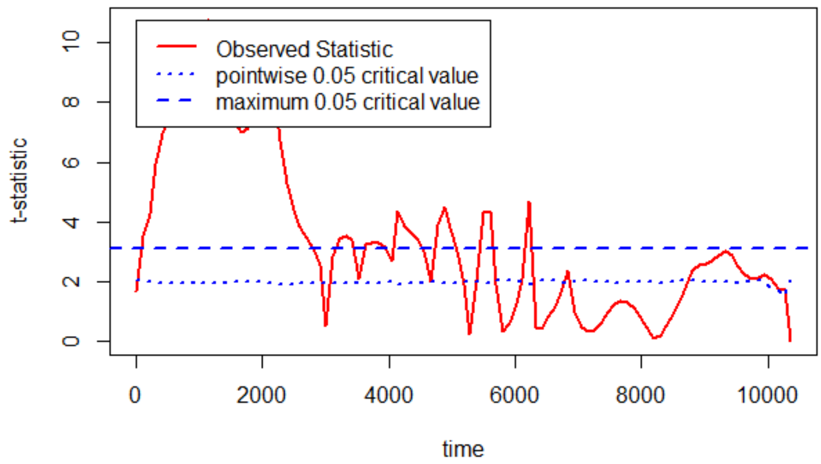
C - A



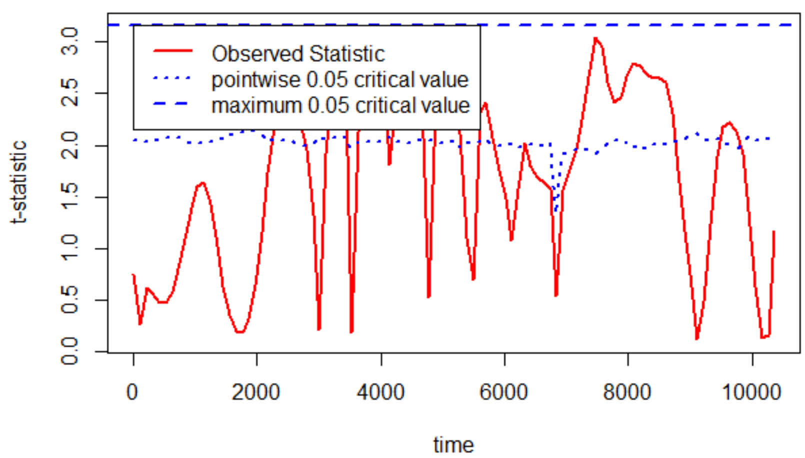
C - B


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0
1
2
3
4
5
0
2000
4000
6000
8000
10000

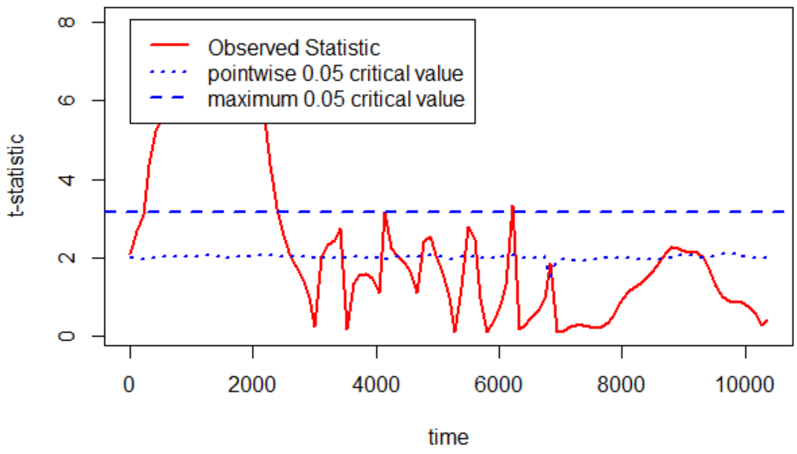
C - D


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0
2
4
6
8
10
0
2000
4000
6000
8000
10000

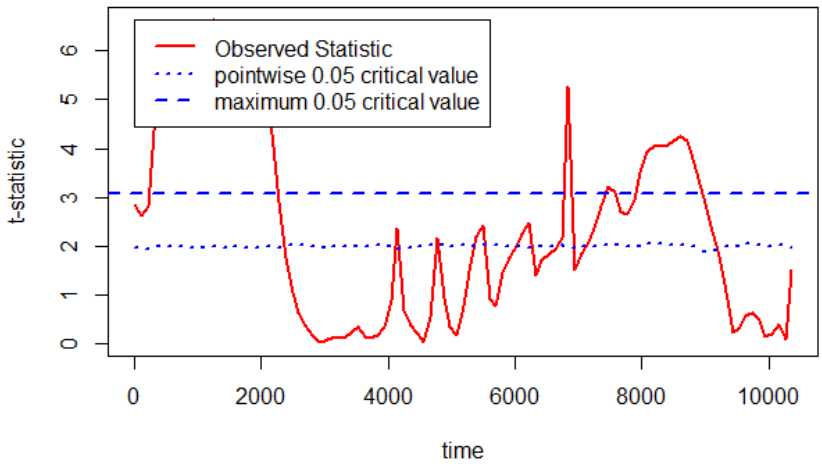
A - B


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
2000
4000
6000
8000
10000

A - D



B - D



The images show that A - B and C - A recipe pairs are not statistically different. Indeed, the observed statistic never exceeds the maximum critical value; this is in line with the Kolmogorov Smirnov's test on hardness measured by the Texture Profile Analyzer.

The instrument does not see statistical differences between the recipes B - C, while the t.perm test sees the two recipes as diverse. This can be explained by looking at the graph: up to the component 2000 the difference between the curves is less than the maximum critical value, so they are not statistically different. On the contrary, around the value 2000, the curves are considered different: this fact is coherent with Kolmogorov Smirnov's test, because around this value it is placed the maximum peak of force measured by the instrument.

Moreover, the recipes pairs C- D, A - D and B - D are judged statistically different, in accordance with the Kolmogorov Smirnov's test on the hardness measured by the Texture Profile Analyzer. The curves differ mainly in the first part, that is during the first crush of the bread: this can be explained by the overdosage of the gluten component, so the glutinic net is more resistant to the first crush but then, once compressed, the bread returns to behave similarly to recipes with standard gluten dose.

Therefore, the D recipe is the hardest of the four for both sensory tasting and laboratory measurements. Curve analysis reveals more information than the individual hardness analyzed earlier, but, unfortunately, this reasoning can not be done with elasticity due to the absence of an elasticity curve.

Changing the curves to have the force spikes aligned would lead to improvements in the analysis? The **register.fd** function of the **fda** package of the R software aligns on the average punctual function of the curves by transforming or distorting the domain arguments of each curve with a non-linear transformation that tightly preserves the order.

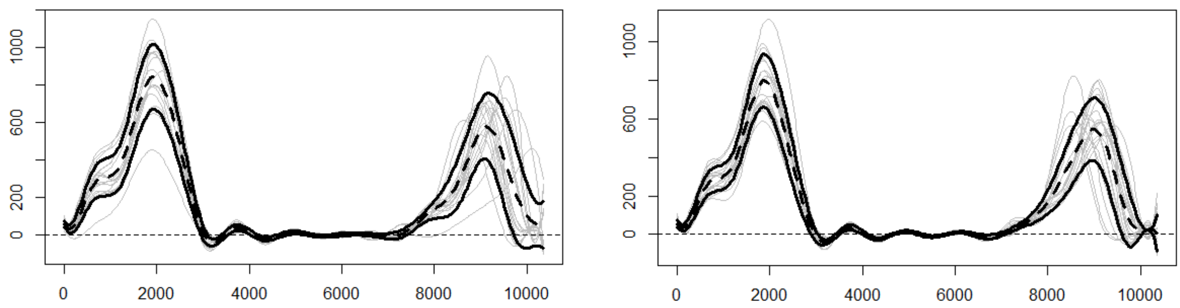


Figure 10 A and C recipes aligned

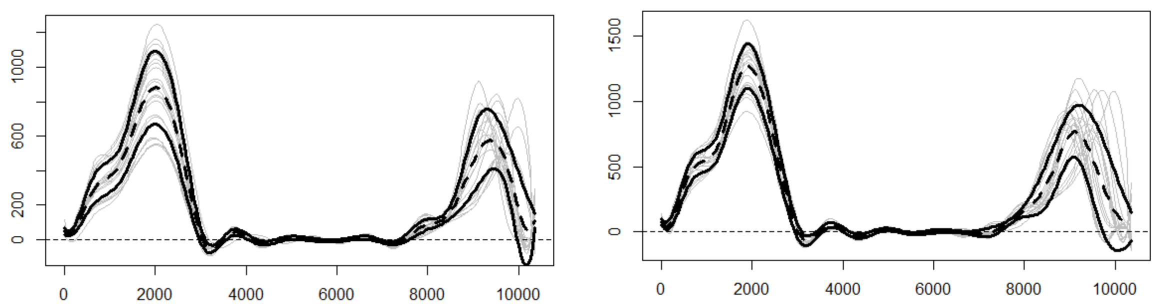
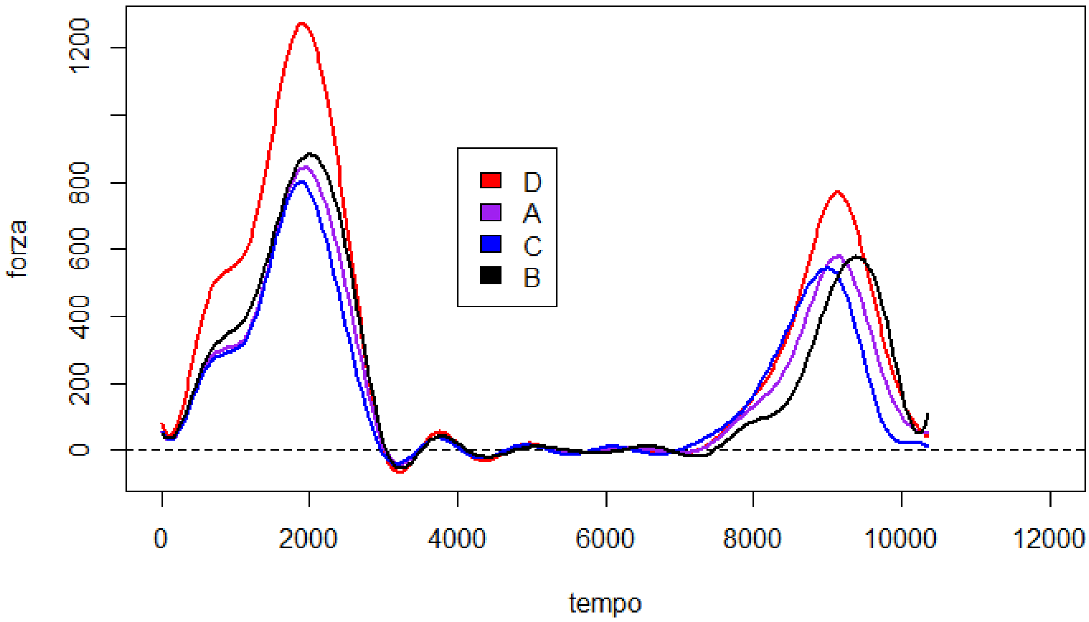
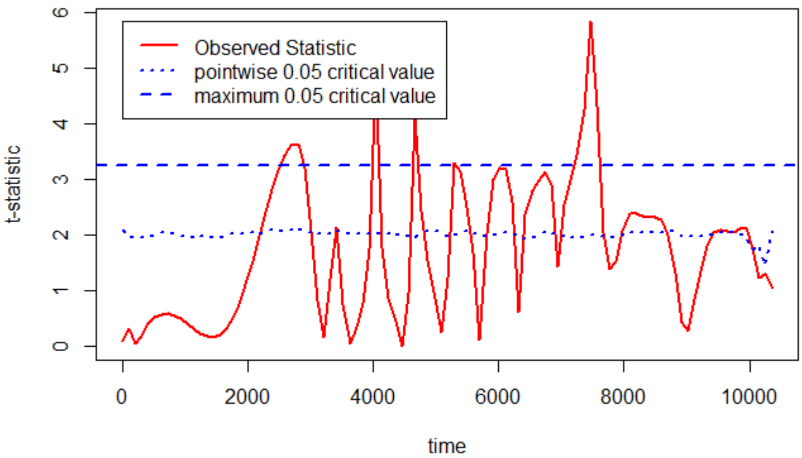


Figura 11 B and D recipes aligned

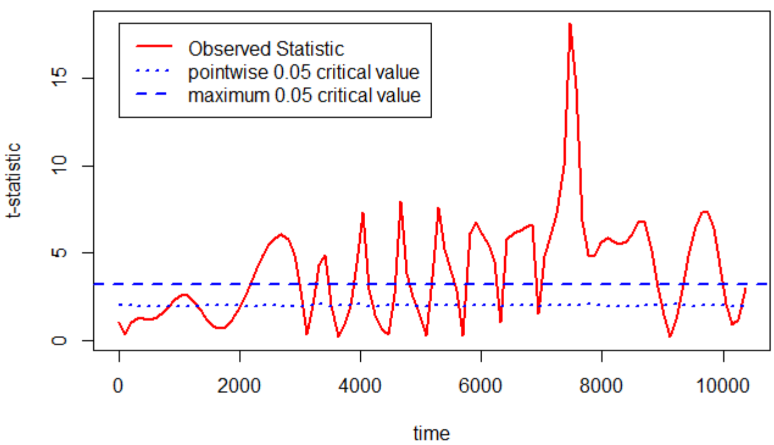


The t.perm test is repeated on the aligned curves:

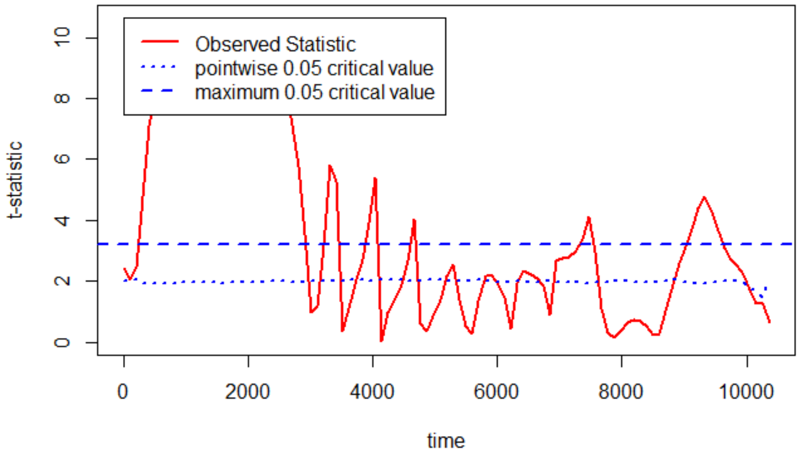
C - A



C - B

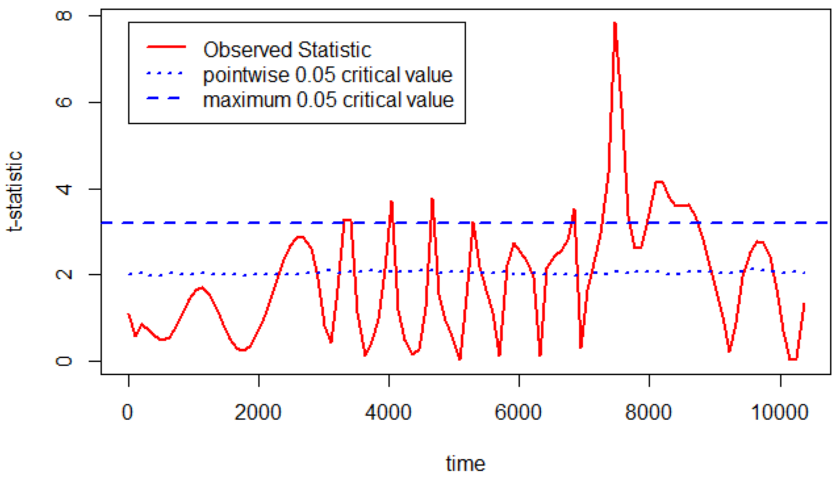


C - D



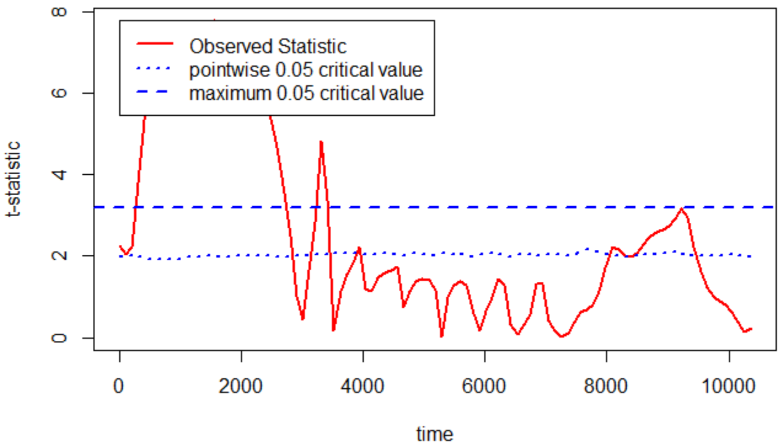
Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0
2
4
6
8
10
0
2000
4000
6000
8000
10000

A - B



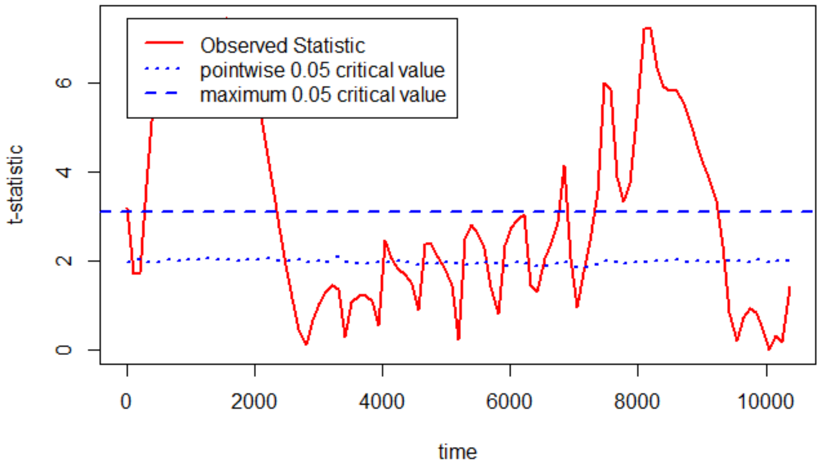
Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0
2
4
6
8
0
2000
4000
6000
8000
10000

A - D



Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
time
0
2
4
6
8
0
2000
4000
6000
8000
10000

B - D



At this point all the pairs of recipes analyzed are different; probably aligning them amplify the noise and make them paradoxically less comparable from a logical point of view.

Then the sensory grades for each attribute and pairs of recipes are compared in order to obtain the frequencies at which one bread is perceived as harder, more elastic, etc. The results for each attribute are reported:

DUREZZA INDICE

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	14	16	11	21	13	14
Greater the second one	19	17	22	12	20	19

DUREZZA MASTICAZIONE

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	16	23	13	25	9	9
Greater the second one	17	10	20	8	24	24

DUREZZA INCISIVI

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	17	26	15	27	10	7
Greater the second one	16	7	18	6	23	26

ELASTICITÀ INDICE E POLLICE

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	18	17	18	14	20	20
Greater the second one	15	16	15	19	13	13

ELASTICITÀ TRA LE LABBRA

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	19	18	16	15	12	16
Greater the second one	14	15	17	18	21	17

ELASTICITÀ INDICE E MEDIO

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	16	12	19	14	20	20
Greater the second one	17	21	14	19	13	13

COESIVITÀ

	A - C	A - B	A - D	C- B	C - D	B - D
Greater the first one	13	22	17	21	18	12
Greater the second one	20	11	16	12	15	21

COMPATTEZZA

	A - C	A - B	A - D	C- B	C - D	B - D
Maggiore il primo pane	12	16	13	22	16	15
Maggiore il secondo pane	21	17	20	11	17	18

SECCHEZZA

	A - C	A - B	A - D	C- B	C - D	B - D
Maggiore il primo pane	21	21	15	24	15	12
Maggiore il secondo pane	12	12	18	9	18	21

By analyzing the frequencies of the columns of the tables using a Chi Quadro test, significant frequencies are highlighted in yellow.

By analyzing them, it can be noted that:

- The hardness attribute is perceived better on the palate than the touch, probably due to the fact that people are daily trained to judge the hardness of a food using bite and chewing compared to touch.
  The hardness measured by chewing coincides with the statistically significant differences observed with the Texture Profile Analyzer.
  Considering the *Durezza Indice*, only the difference between A - D recipes is significant, where D is perceived as harder. Looking at the frequencies between the D - B and D - C pairs,

although differences are not statistically significant, they show that D is perceived harder than the other two.

Finally, the significant frequencies involve all comparisons related to the B recipe, which is perceived softer, in accordance with the rankings obtained from the votes of sensory tastes. So, it seems that the *Durezza Indice* evaluates more clearly the comparison with the recipe perceived as softer, while the *Durezza Masticazione* evaluates more clearly the comparison with the recipe perceived as harder.

- Regarding elasticity, unfortunately no significant frequencies are observed due to the difficulty of evaluating the samples.

- Dealing with dryness, the only statistically significant difference is between the C - B pair, however, this is not consistent with the observed weight drops where C is measured as less dry than B.

- *Coesività* attribute is statistically significant for the C - B recipes pair, which is consistent with the test made on the measurements of the Texture Profile Analyzer, as can be seen from the figure below:

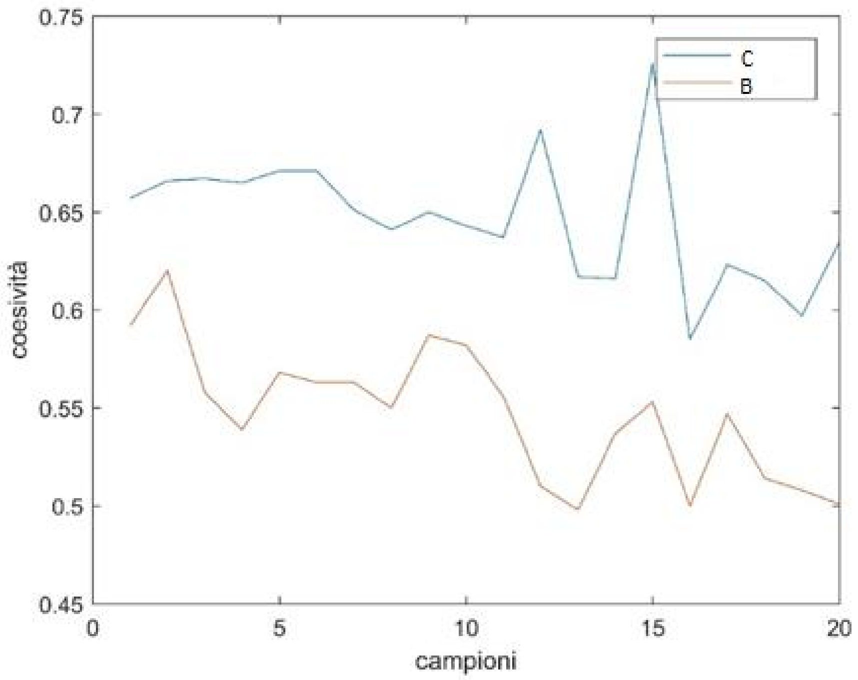


Figura 12 Coesività between C and B

8. Planning the second taste

The first taste results show that the bread samples seem to be very similar. So, for the second taste, two recipes change to make the bread more different. The panel taste again four different recipes.

8.1 Choice of the bread recipes

Recipes A and D don't change, while B and C do:

- **E**

  This recipe is obtained by increasing the rate to 40% of grease (sunflower oil), which consists of an increase to 3,2% on absolute term on the dough; by adding grease, the bread is expected to be softer than B.
- **F**

  This recipe is obtained by adding a rate to 0,3% of a specific enzyme in the mixture of powders (flour, gluten, premix, ...) of the standard recipe; by adding the enzyme the bread is expected to be chewier than recipe C.

As already said, bread samples are shaven (height 25 mm) to give aligned Texture profile analyzer curves.

Figure 40 Bread sample

8.2 Choice of the software layout for the taste

During the second taste, in order to try to make more significance the votes of the panel, they are integer numbers; indeed, two significance figures after the decimal point may be too much and can mark distinction between bread that panel do not recognize. Moreover, panelists do not see the number that they are marking on the line. As a consequence, tasters have only nine possible values to rate the samples, so many bread samples can have the same vote; maybe, in this taste rank could be less easy to find.

After the first taste, panelists told that they find difficult to evaluate the bread samples because, according to them, there were too many attributes. However, it was decided not to change them because only three tastes will be carried out, so it is more accurate to have the same attributes to analyze each of them.

Moreover, the wheel layout is no more used; instead of the panelist sees for each attribute the line scaling of all the bread samples, with the code of the sample above. So, the panelist is asked to focus more on the rank for each attribute with respect to the single vote.

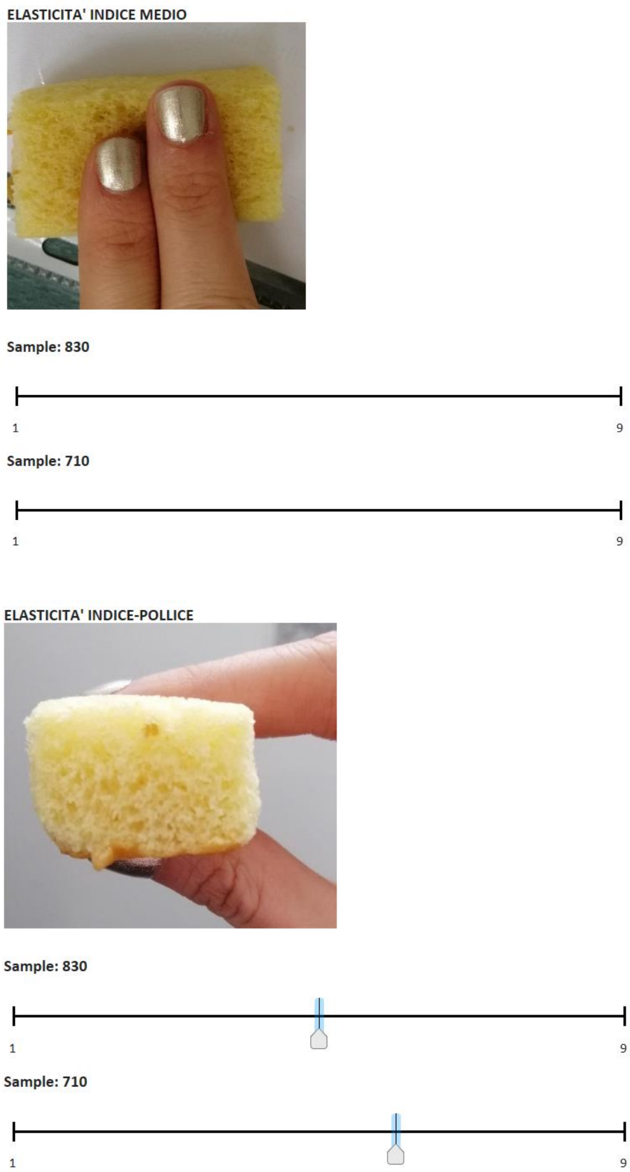


Figure 41 Compusense layout for the second tast

In order to help the panelist to remember the right attribute definition, images are added below the attribute name to help its definition.

9. Second taste results

Once the sensory data and measurements of the Texture Profile Analyzer tool are collected, each pairs of recipes are compared using the Kolmogorv test, first using sensory data and then the tool data.

Second taste data does not show significant differences between the numerical ratings given by the tasters to the different recipes for each attribute evaluated. The test is again used to compare the different attributes of the same field, for example the three different hardness types that were analyzed by the tasters: *Durezza medio e Indice, Durezza incisivi* and *Durezza masticazione.*

HARDNESS	**A**	**E**	**D**	**F**
D. Masticazione vs D. Incisivi	Statistically Similar	S. Similar	S. Similar	S. Similar
D. Masticazione vs D. Indice	S. Similar	S. Similar	S. Similar	S. Similar
D. Indice vs D. incisivi	S. Similar	S. Different	S. Similar	S. Similar

ELASTICITY	**A**	**E**	**D**	**F**
E. Indice e Medio vs E. Labbra	S. Similar	S. Similar	S. Similar	S. Similar
E. Indice e Pollice vs E. Indice e Medio	S. Similar	S. Similar	S. Similar	S. Similar
E. Indice e Medio vs E. Labbra	S. Similar	S. Similar	S. Similar	S. Similar

	A	**E**	**D**	**F**
Coesività vs Compattezza	S. Different	S. Similar	S. Similar	S. Similar

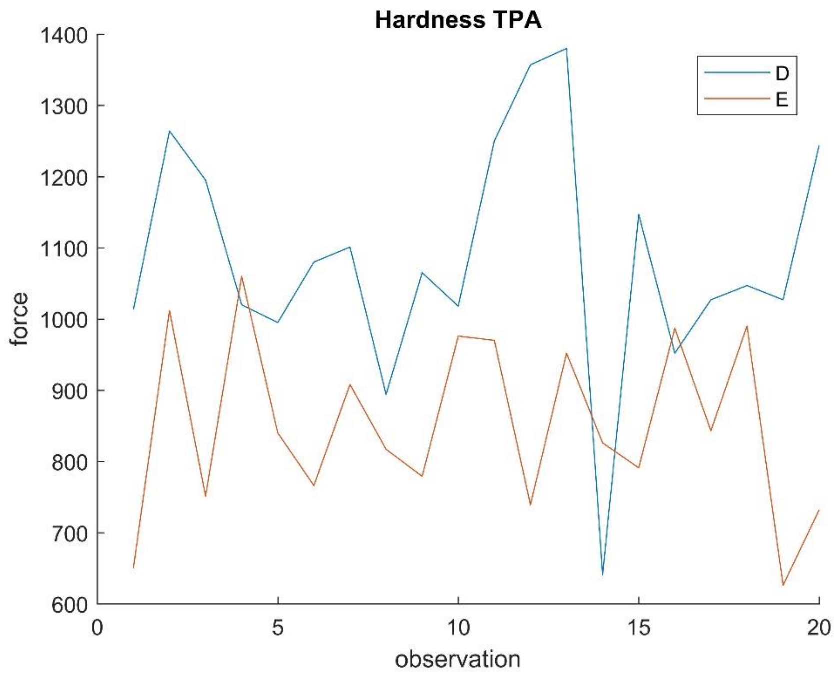
Recipes	**TPA Hardness**	**Sensory hardness**	**TPA elasticity**	**Sensory elasticity**	**TPA cohesiveness**	**Sensory cohesiveness**
D - E	S. Different	S. Similar	S. Similar	S. Similar	S. Different	S. Similar
D - A	S. Different	S. Different	S. Different	S. Similar	S. Similar	S. Similar
D - F	S. Different	S. Different	S. Different	S. Similar	S. Different	S. Similar
A - B	S. Similar	S. Similar	S. Different	S. Similar	S. Different	S. Similar
A - F	S. Similar	S. Similar	S. Different	S. Similar	S. Different	S. Similar
E - F	S. Similar	S. Similar	S. Different	S. Similar	S. Different	S. Similar

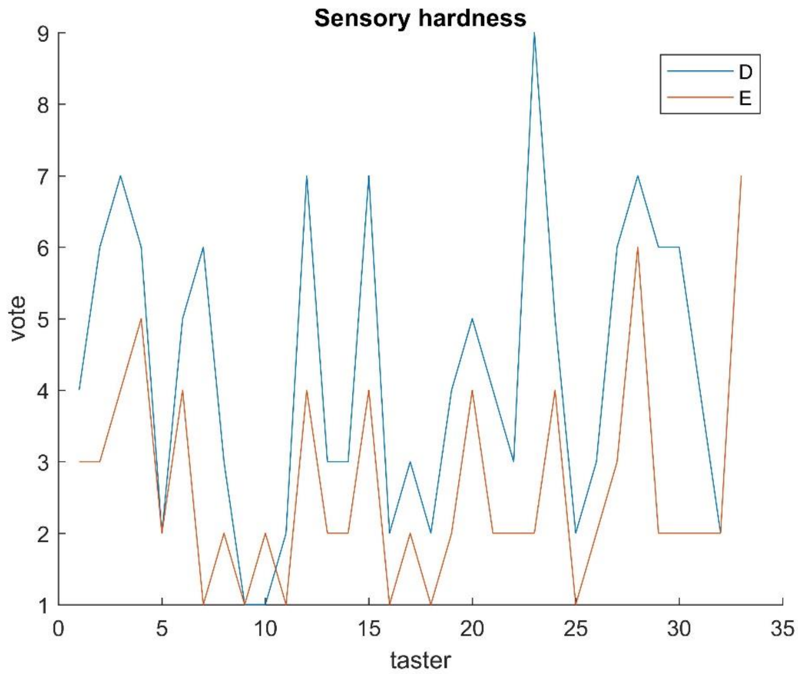
The *Durezza Indice* is taken into account as sensory hardness.

As in the previous taste, there are no significant sensory differences for the attributes of elasticity and cohesiveness. However, dealing with hardness attribute, it occurs that the sensory results and the results of the Texture Analyzer coincide and, more interestingly, it coincides the perception of diversity between the pairs D - E, D - A and D - F. As it will be shown later, indeed, the D recipe is perceived as the hardest even in this taste and probably the introduction of votes by integer numbers accentuates the difference.

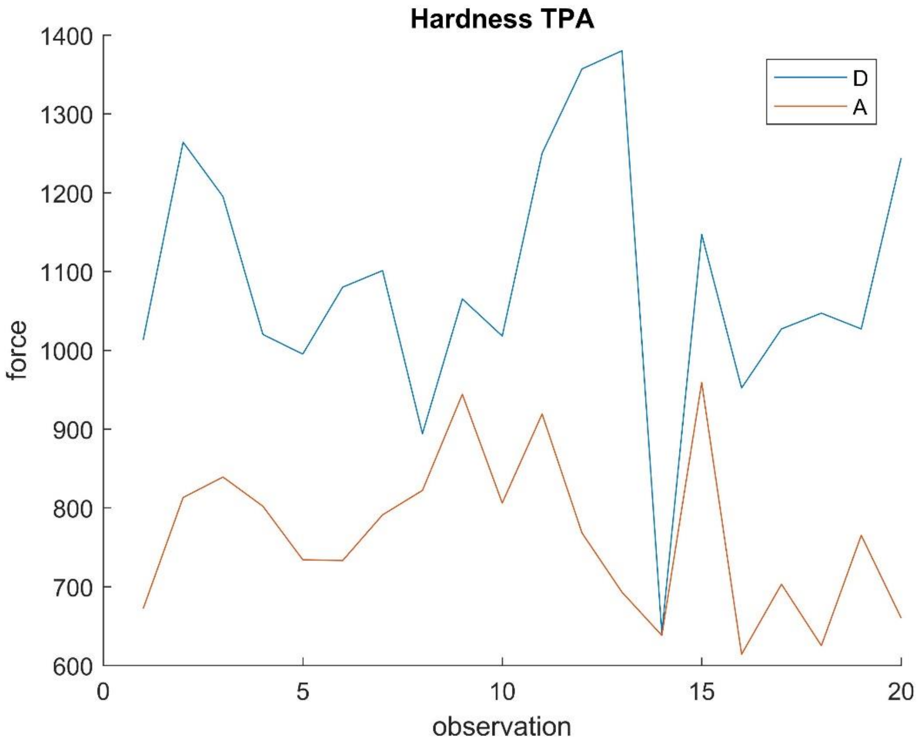
Let's look at the voting and measurement vectors of the hardness tool of each recipe pair.

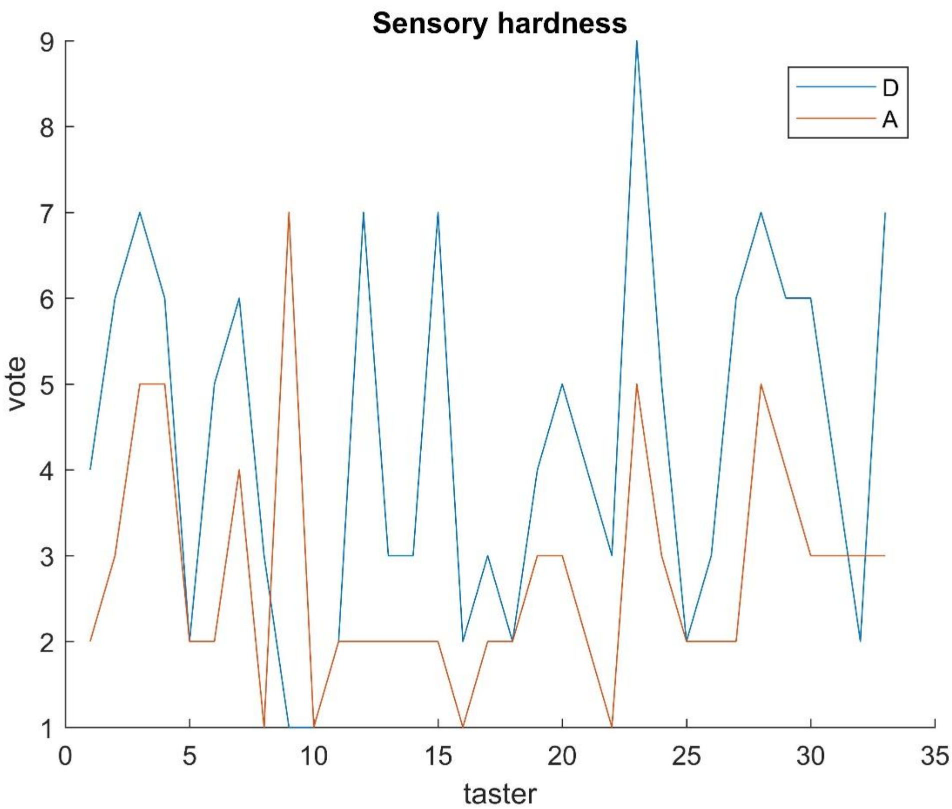
D - E



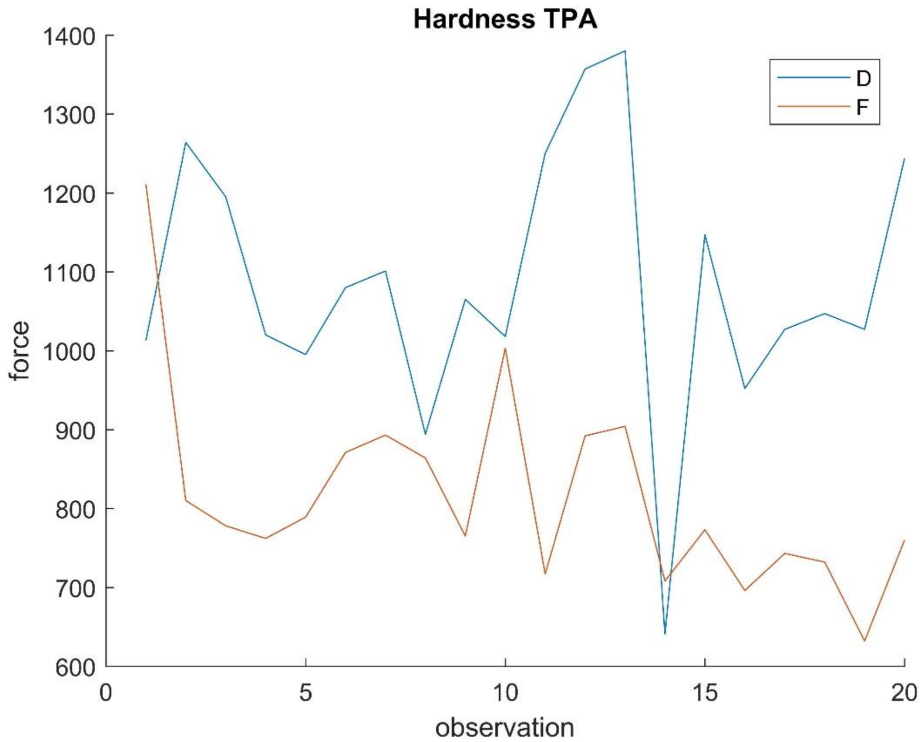


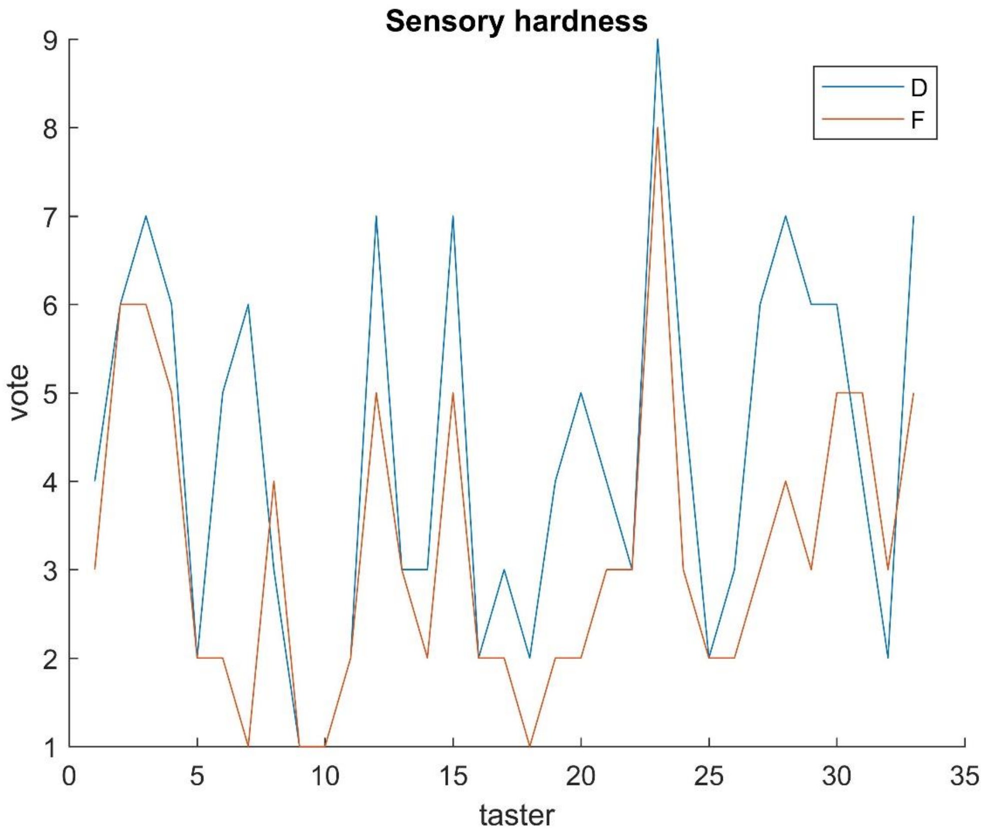
D – A


Hardness TPA
D
A
force
observation
1400
1300
1200
1100
1000
900
800
700
600
0
5
10
15
20

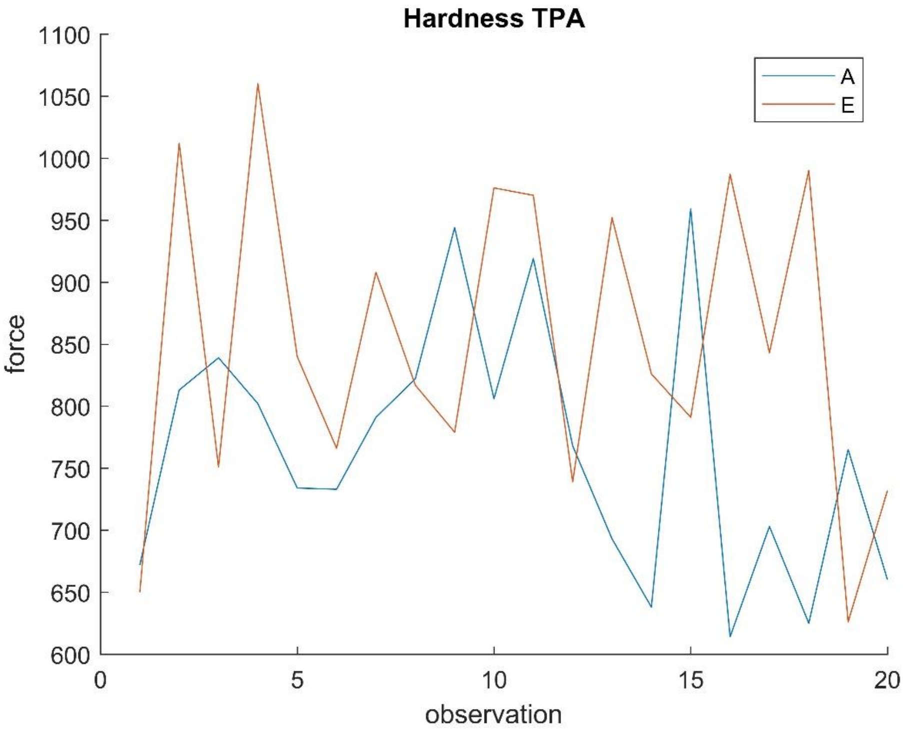

Sensory hardness
D
A
vote
taster
9
8
7
6
5
4
3
2
1
0
5
10
15
20
25
30
35

D – F

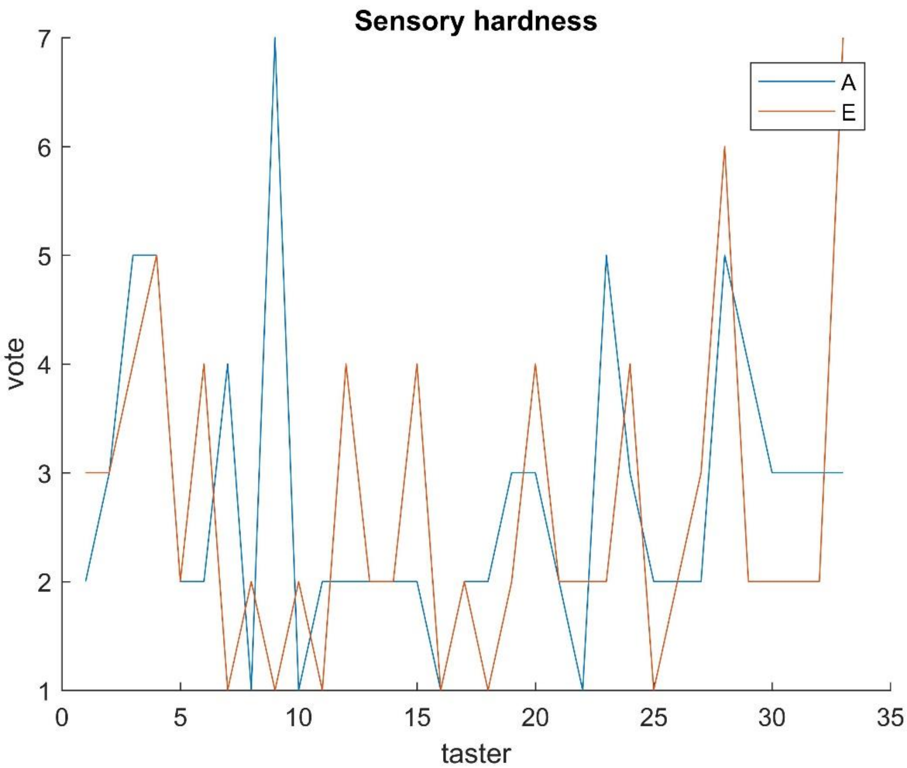

Hardness TPA
force
observation
D
F
1400
1300
1200
1100
1000
900
800
700
600
0
5
10
15
20


Sensory hardness
vote
taster
D
F
9
8
7
6
5
4
3
2
1
0
5
10
15
20
25
30
35

A – E

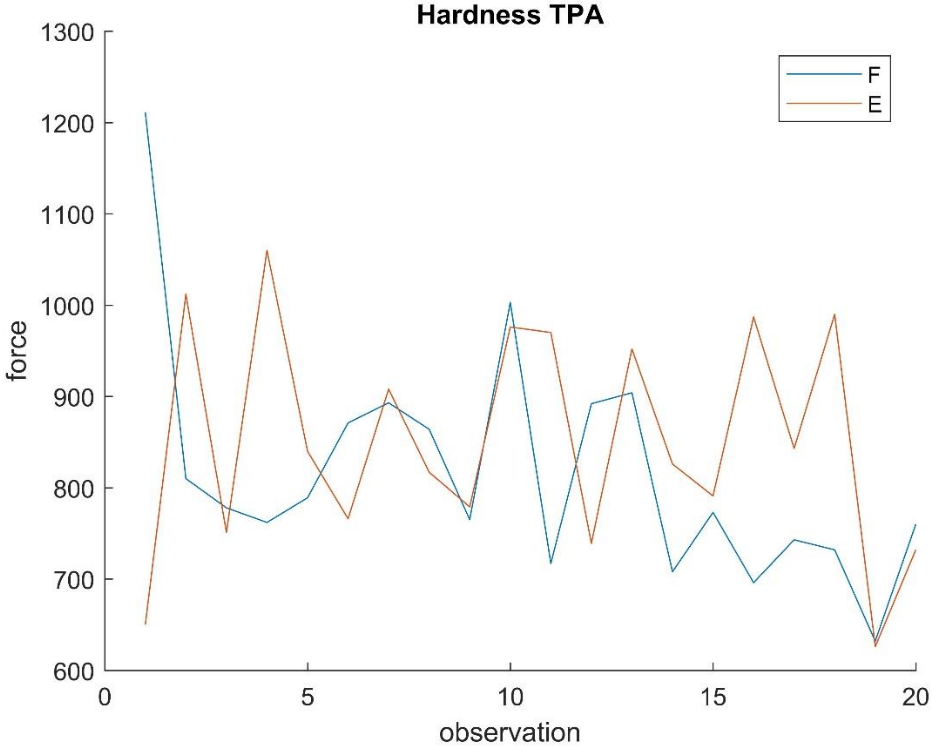


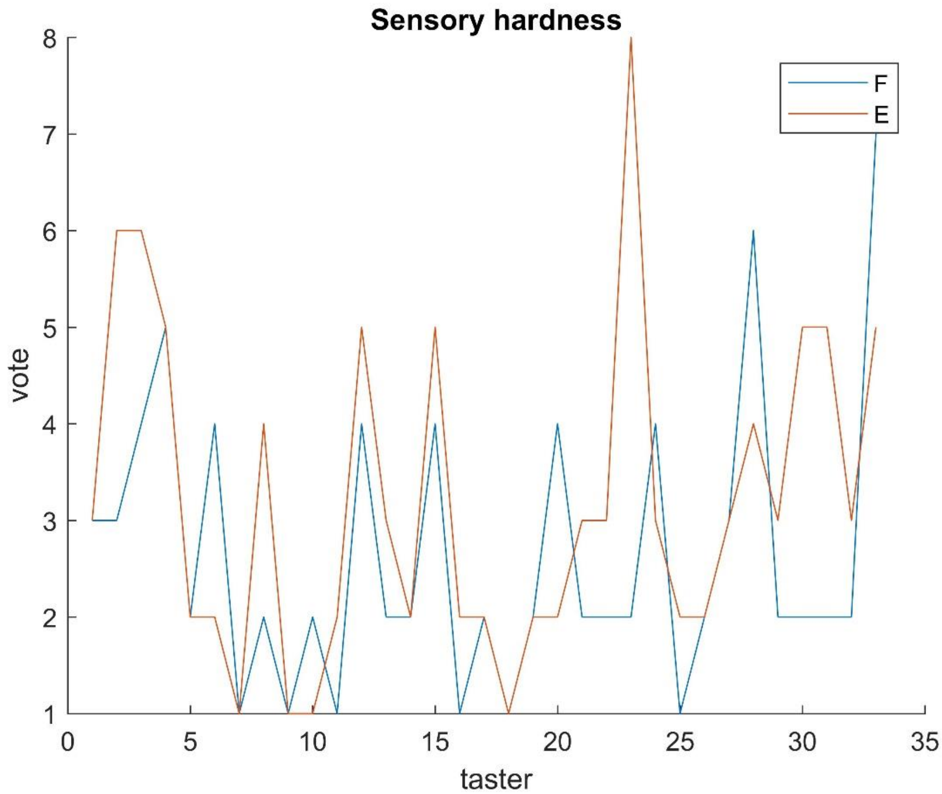
Hardness TPA
force
observation
A
E
1100
1050
1000
950
900
850
800
750
700
650
600
0
5
10
15
20



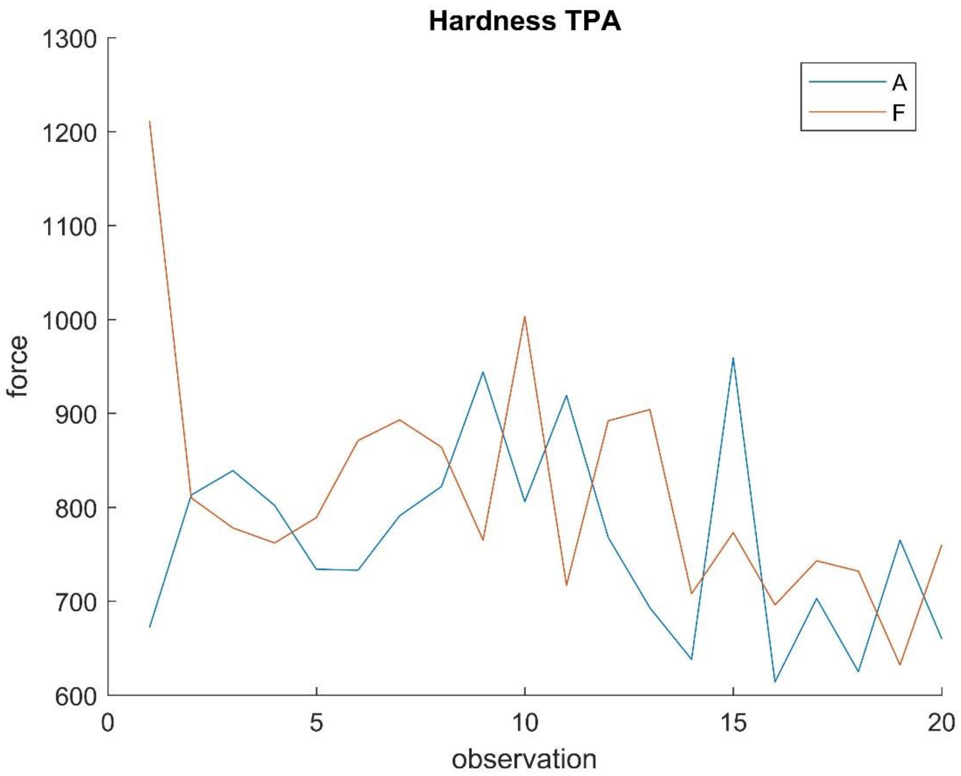
Sensory hardness
vote
taster
A
E
7
6
5
4
3
2
1
0
5
10
15
20
25
30
35

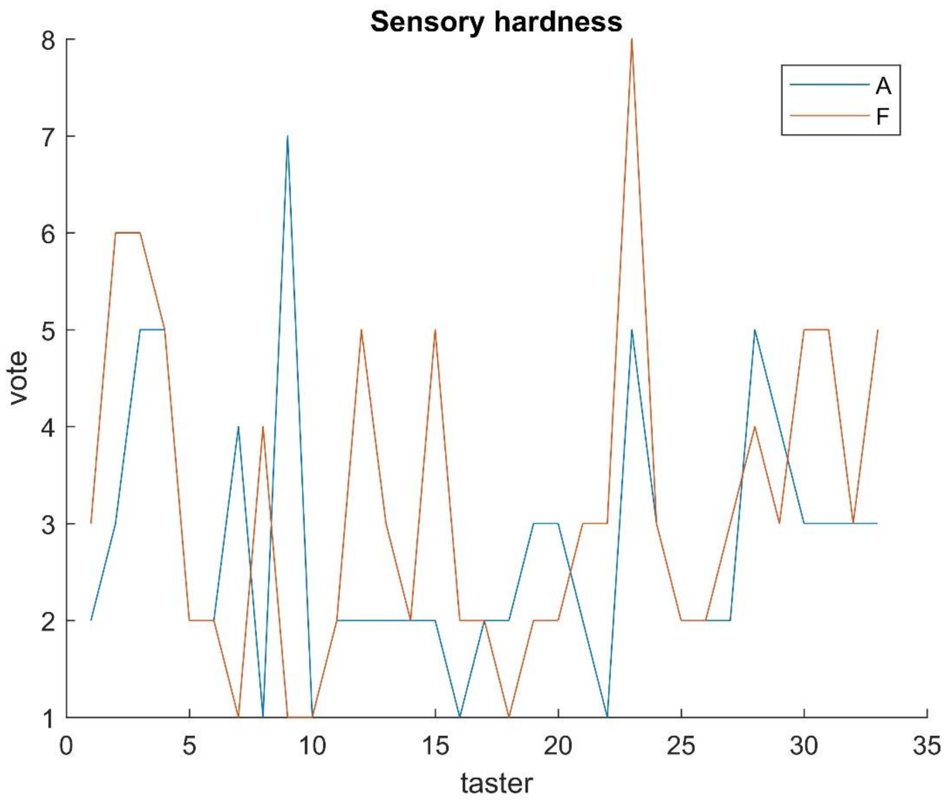
A - F


Hardness TPA
F
E
force
observation
1300
1200
1100
1000
900
800
700
600
0
5
10
15
20


Sensory hardness
F
E
vote
taster
8
7
6
5
4
3
2
1
0
5
10
15
20
25
30
35

F – E


Hardness TPA
force
observation
A
F
1300
1200
1100
1000
900
800
700
600
0
5
10
15
20


Sensory hardness
vote
taster
A
F
8
7
6
5
4
3
2
1
0
5
10
15
20
25
30
35

The average of the hardness measured by the Texture Analyzer, out of the 20 samples analyzed for each recipe, are as follows:

Mean D: 1085,8 g

Mean E: 850,75 g

Mean A: 765 g

Mean F: 815,15 g

Δ (D - E) = 235,05 g

Δ (D - A) = 320,80 g

Δ (D – F) = 270,65 g

Δ (A - E) = 85,75 g

Δ (A – F) = 50,15

Δ (E - F) = 35.60 g

A difference between the averages greater than 235.05 g makes the tasters to assign votes that are statistically different in terms of hardness, that agree with those ones of the Texture Analyzer. However, differences minor than 85.75 make not the tasters and the Texture Analyzer able to perceive differences.

Next, it is possible to analyze the rank given by each taster thanks to the votes expressed in the tasting. The tables below show the frequencies at which the various recipes are ranked at first, second, third and fourth place respectively, where the first place indicates greater intensity of the attribute and the fourth place indicates less intensity. Since only nine votes are possible, there are recipe receiving the same vote from the taster, so the rows of the tables no longer add up to 33.

Elasticità Indice e Pollice	**A**	**D**	**E**	**F**
1^	10	13	8	12
2^	8	5	13	14
3^	7	7	5	5
4^	8	8	7	2

Elasticità Indice e Medio	**A**	**D**	**E**	**F**
1^	12	9	7	11
2^	8	4	13	11
3^	7	10	7	9
4^	6	10	6	2

Elasticità Labbra	**A**	**D**	**E**	**F**
1^	16	11	12	13
2^	1	6	13	9
3^	8	6	6	8
4^	8	10	2	3

Coesività	**A**	**D**	**E**	**F**
1^	16	15	10	16
2^	8	9	9	6
3^	3	1	9	8
4^	6	8	5	3

Compattezza	**A**	**D**	**E**	**F**
1^	9	22	11	17
2^	9	5	8	8
3^	5	3	9	6
4^	10	3	5	2

Secchezza	**A**	**D**	**E**	**F**
1^	7	26	4	11
2^	6	4	11	13
3^	8	3	12	8
4^	12	0	6	1

Durezza Indice	**A**	**D**	**E**	**F**
1^	2	27	4	8
2^	6	2	18	14
3^	11	2	8	11
4^	14	2	3	0

Durezza Masticazione	**A**	**D**	**E**	**F**
1^	6	28	3	10
2^	8	4	11	14
3^	12	1	12	7
4^	7	0	7	2

Durezza Incisivi	A	D	E	F
1^	4	26	2	13
2^	8	6	9	15
3^	16	1	17	3
4^	5	0	5	2

From the previous tables it is clearer in some cases than others, which recipe is put at the first, second, third and fourth place for each attribute. Hence, again, the search for a correlation is conducted between the ranks of sensory votes and the ranks given by the Texture Profile Analyzer's average measurements, which is carried out by Kendall's correlation test and Spearman's correlation test. The ranks are obtained again using PCA reduction on the table frequencies.

ATTRIBUTO	**PANI**	**MEDIA TPA**	**RANK TPA**	**RANK CLASSIFICA**
ELASTICITA_INDICE_POLLICE	A	0,8241	4	3
ELASTICITA_INDICE_POLLICE	F	0,8853	1	1
ELASTICITA_INDICE_POLLICE	E	0,8529	2	2
ELASTICITA_INDICE_POLLICE	D	0,8473	3	4
ELASTICITA_INDICE_MEDIO	A	0,8241	4	3
ELASTICITA_INDICE_MEDIO	F	0,8853	1	2
ELASTICITA_INDICE_MEDIO	E	0,8529	2	1
ELASTICITA_INDICE_MEDIO	D	0,8473	3	4
ELASTICITA_TRA_LABBRA	A	0,8241	4	4
ELASTICITA_TRA_LABBRA	F	0,8853	1	2
ELASTICITA_TRA_LABBRA	E	0,8529	2	1
ELASTICITA_TRA_LABBRA	D	0,8473	3	3
COESIVITA	A	0,4914	4	3
COESIVITA	F	0,5668	1	2
COESIVITA	E	0,5151	2	1
COESIVITA	D	0,4981	3	4
COMPATTEZZA	A	0,4914	4	4
COMPATTEZZA	F	0,5668	1	2
COMPATTEZZA	E	0,5151	2	3
COMPATTEZZA	D	0,4981	3	1
DUREZZA_INDICE	A	765	4	3
DUREZZA_INDICE	F	815,15	3	2
DUREZZA_INDICE	E	850,75	2	4
DUREZZA_INDICE	D	1085,8	1	1
DUREZZA_MASTICAZIONE	A	765	4	3
DUREZZA_MASTICAZIONE	F	815,15	3	2
DUREZZA_MASTICAZIONE	E	850,75	2	4
DUREZZA_MASTICAZIONE	D	1085,8	1	1
DUREZZA_INCISIVI	A	765	4	3
DUREZZA_INCISIVI	F	815,15	3	2
DUREZZA_INCISIVI	E	850,75	2	4
DUREZZA_INCISIVI	D	1085,8	1	1
SECCHEZZA	A	24,976	1	4

SECCHEZZA	F	23,409	4	2
SECCHEZZA	E	24,409	2	3
SECCHEZZA	D	23,992	3	1

The results are summarized in the following table:

ATTRIBUTE	RHO	p-value	TAU	p-value
Elasticità Indice e Pollice	0,8	0,2	0,66	0,1742
Elasticità Indice e Medio	0,6	0,4	0,33	0,4969
Elasticità Labbra	0,8	0,2	0,66	0,1742
Coesività	0,6	0,4	0,33	0,4969
Compattezza	0,4	0,6	0,33	0,4969
Durezza indice	0,4	0,6	0,33	0,4969
Durezza Masticazione	0,4	0,6	0,33	0,4969
Durezza Incisivi	0,4	0,6	0,33	0,4969
Secchezza	-0,8	0,2	-0,66	0,1742

Looking at the rho and tau estimates, it can be observed that the perception of tasters agrees with the measurements of the instrument, although no correlation index is significant, except for the attribute *Secchezza* for the same reason already explained: the fact that the rho and tau estimators for dryness are negative is related to the way they are estimated by weight loss. Indeed, weight loss is the measure that indicates how much water is lost from bread after spending a night in the stove at 105 degrees, that is the humidity of the bread. So, the measure of the water lost during this process is inversely proportional to the feeling of dryness perceived on the palate, which measures how much saliva the bread absorbs.

However, it must be taken into account that to obtain a rank on the values of the Texture Profile Analyzer only the average measurements are used, without considering the variance between the samples of a single recipe.

Subsequently, a Chi Quadro test of Pearson is carried out on the frequencies observed by the tables (null hypothesis: the samples come from the same population of average and variance estimated by the table), which shows that:

ATTRIBUTE	TEST RESULT
Durezza Masticazione	Rejection H0 - Very significant (p-value < 0,02)
Durezza Incisivi	Rejection H0 - Very significant
Durezza Indice	Rejection H0 - Significant (0,02 < p-value < 0,05)
Elasticità Indice e Pollice	H0 is not rejected
Elasticità Indice e Medio	H0 is not rejected
Elasticità Labbra	H0 is not rejected
Secchezza	Rejection H0 - Significant

Coesività	Rejection H0 - Very significant
Compattezza	Rejection H0 - Significant

The test results confirm what can be noticed observing the tables: in the attributes concerning hardness the D recipe receives more votes as first place, with differences in significant votes. After fixing the attribute and position, the frequencies more significant are highlighted in yellow in the tables and the significance is measured again with a Chi Quadro test). E recipe in this tasting does not always turn out to be the softest.

The most significant differences belong to the attributes of hardness: in fact, as noted above, the D recipe always excels, it is also found at the first place of the *Secchezza* attribute. This is interesting: being perceived as the driest, that means presenting less moisture and therefore water, affects the perception of the hardness of bread. Instead, A recipe is perceived as the less dry, so the wettest, in accordance with the measurements of weight drops.

Even during the second taste, the frequency tables show elasticity attributes of the samples are not well understood: indeed, a single significance occurs in a tables at the third position and not at the first or last position.

Next, the entire curves (force,time) measured by the Texture Profile Analyzer, are analyzed. Curves for A, D, E and F recipes are reported. In the figure below the average point curve (black dashed line) and the average point curves ± standard point deviation (thick black lines) are shown.

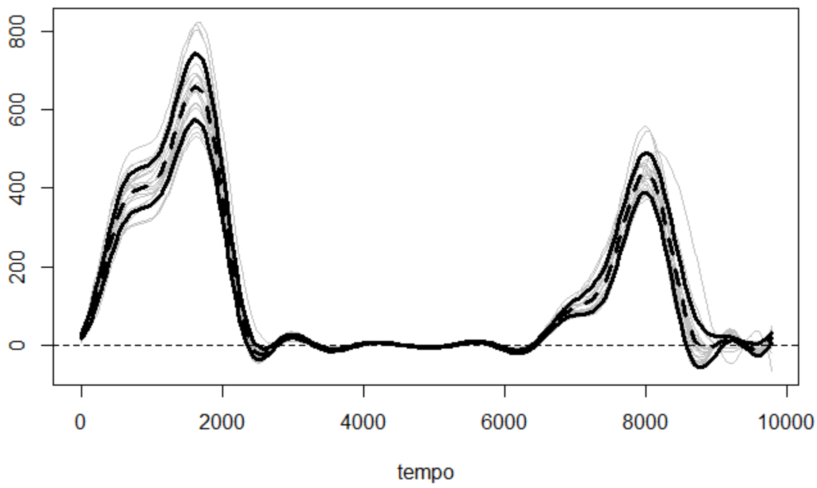


Figure 13 A recipe force curve

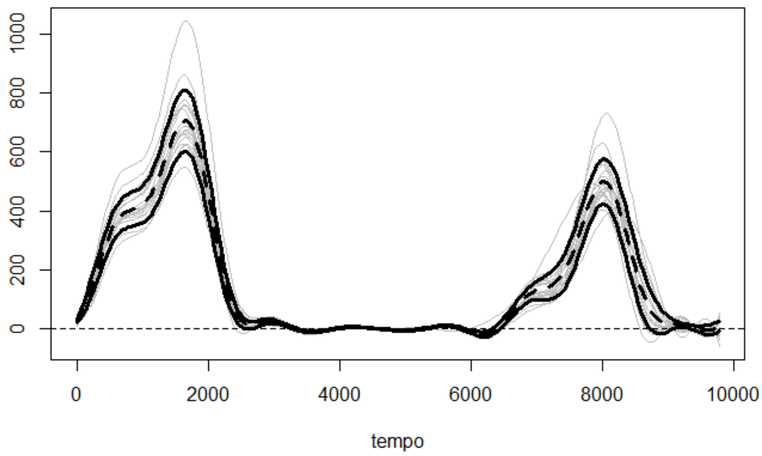


Figure 14 F recipe force curve

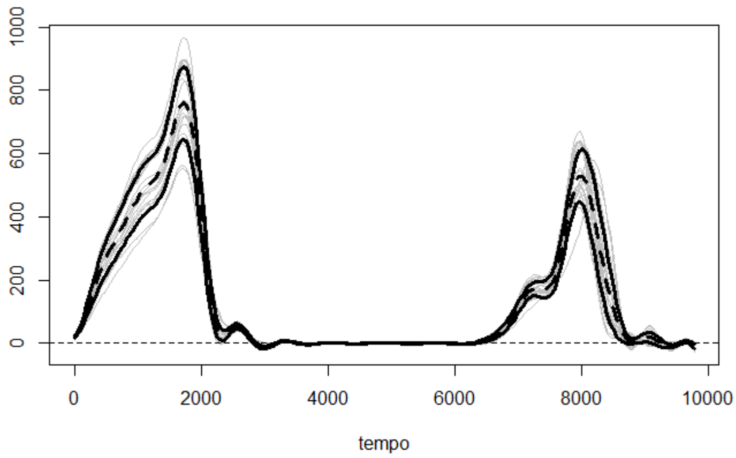


Figura 15 E recipe force curve

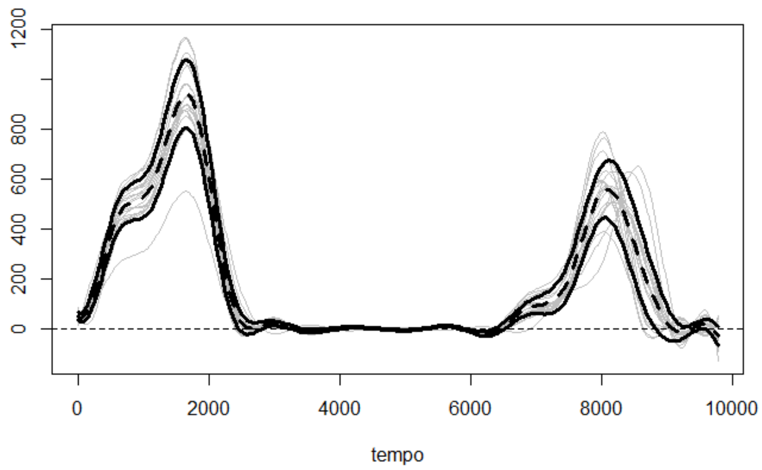
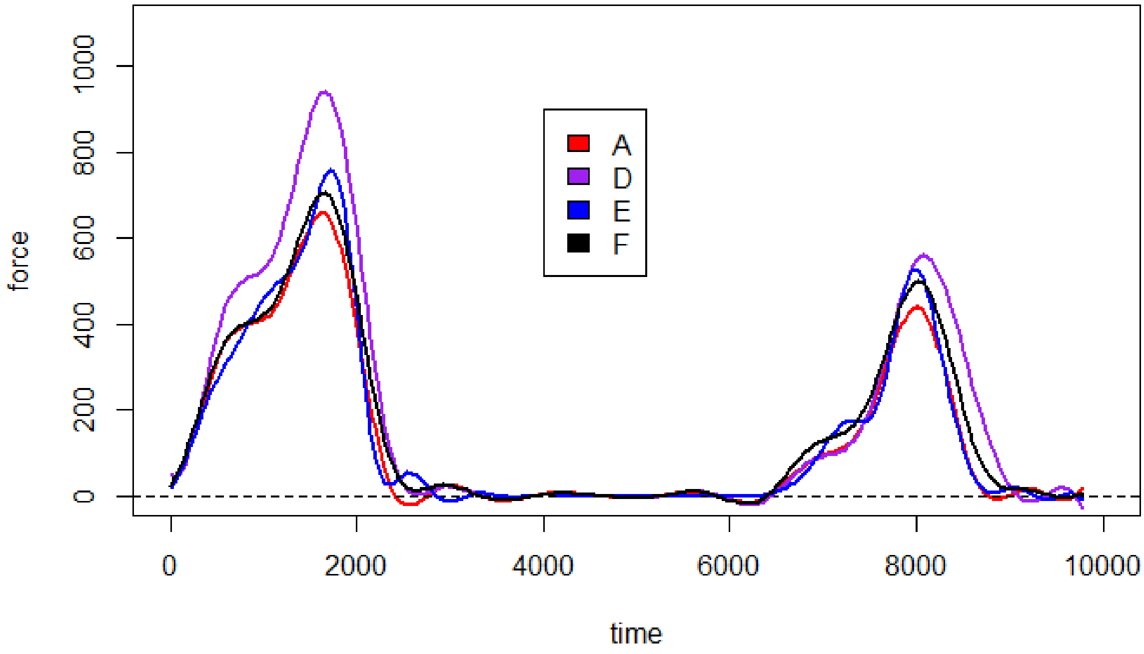


Figura 16 D recipe force curve

Below the average punctual functions for each recipe are represented:

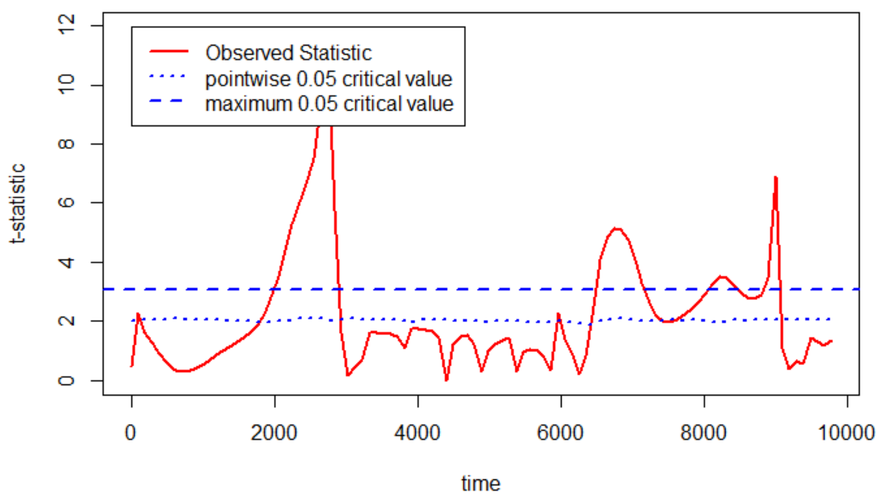


The image shows that the D recipe is perceived as the hardest by the instrument, which is in agreement with sensory assessments.

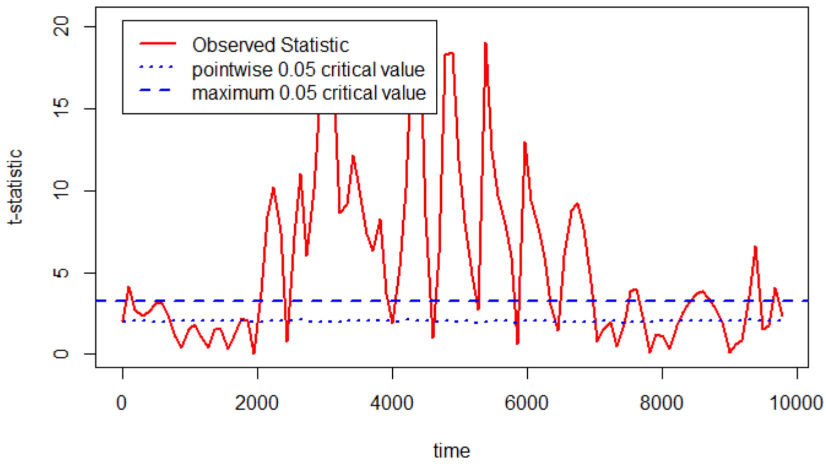
During the second taste the sampes, as mentioned above, are shaved, which is the reason why the curves are more aligned than the first taste: so, in this case, **register.fd** function is not used to align curves.

The **t.perm** test is repeated and the output of the test returns the graphs below:

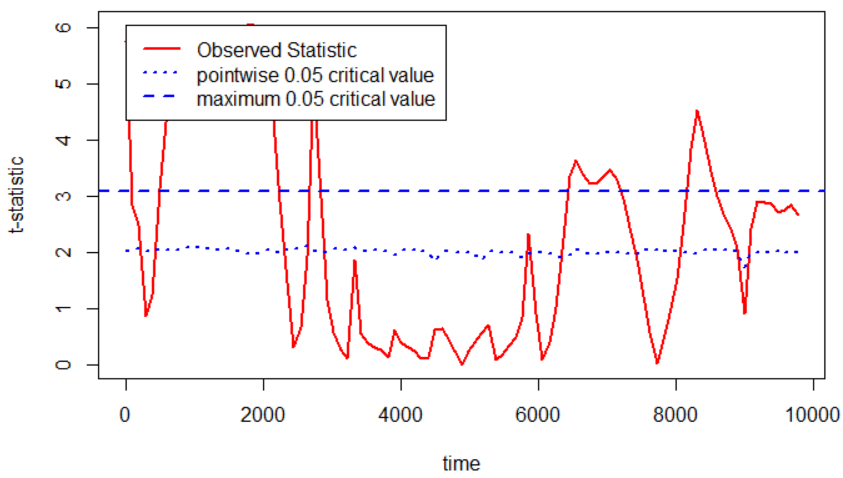
F – A



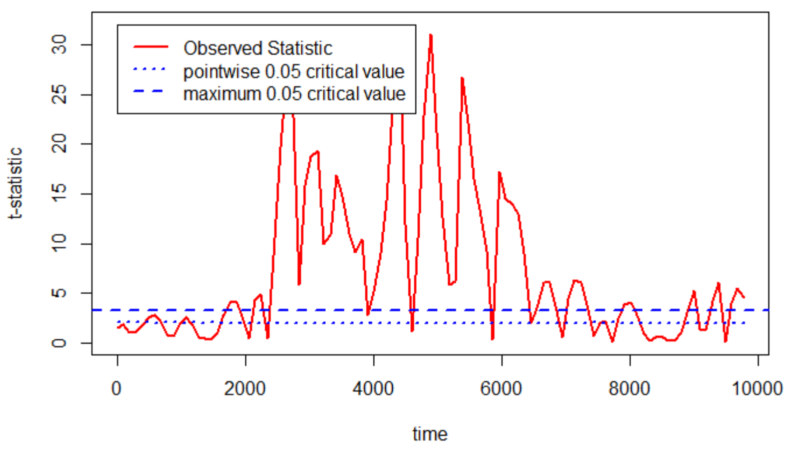
F – E


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
0
5
10
15
20
0
2000
4000
6000
8000
10000
time

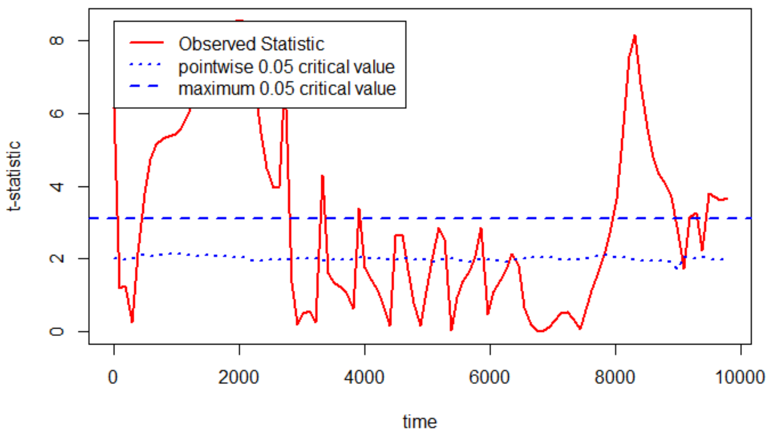
F - D


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
0
1
2
3
4
5
6
0
2000
4000
6000
8000
10000
time

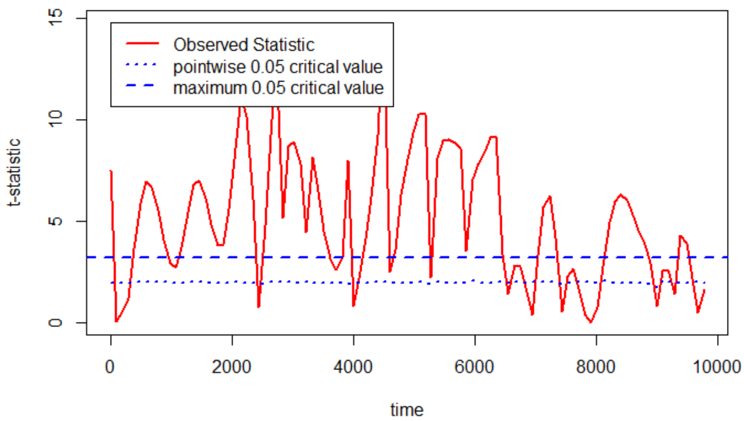
A - E


Observed Statistic
pointwise 0.05 critical value
maximum 0.05 critical value
t-statistic
0
5
10
15
20
25
30
0
2000
4000
6000
8000
10000
time

A - D



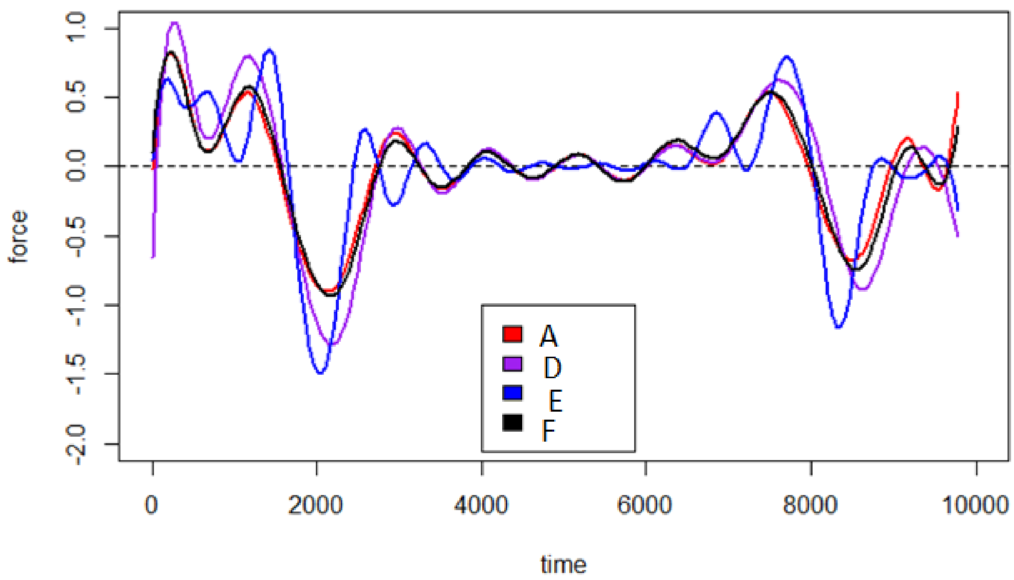
E - D



The graphs confirm that D recipe is again the one that differs the most from the others on the first and the second peaks measured by the Texture Analyzer. Moreover F recipe differs from the A in insignificant places of the curves: they differ in the places where the Texture Analyzer probe rises after compressing the bread, so this can be interpreted as F recipe returns to the probe a force that decreases more slowly than the A one. Also E recipe presents a particular behavior. This differs from the A and F curves in the points of the curve where the force results null for each recipe: in those sections, in fact, the probe is returning to its initial position and then descending, so it measures no force. In those traits (after component 2000 and before component 8000) the D recipe does not differ statistically significantly from the recipes A and F, as it is logical to expect. This behavior is common only in the analyses in which the E recipe is involved.

In fact, if you look closely at the average punctual curves you notice that the recipe presents a faster decrease after the first peak, compared to the other recipes, and a slower growth during the second peak: this could explain the differences shown in the test.

Below the graph of the first derivatives of the curves of each recipe is shown. The first derivative of the E recipe is very different from the others; therefore, the difference is not related to the peaks of force reached by the probe, but to the different growth and decrease of the curve.



Then, the sensory grades of the recipe pairs are compared for each attribute by using the frequencies at which one bread is perceived as harder, more elastic, etc.

The results for each attribute are reported:

DUREZZA INDICE

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	3	2	6	23	26	11
Greater the second	30	31	27	10	7	22

DUREZZA MASTICAZIONE

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	2	7	12	21	28	16
Greater the second	31	26	21	12	5	17

DUREZZA INCISIVI

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	2	3	8	18	27	23
Greater the second	31	30	25	15	6	1

ELASTICITÀ INDICE E POLLICE

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	15	11	16	11	14	14
Greater the second	18	22	17	22	19	19

ELASTICITÀ TRA LE LABBRA

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	17	11	13	12	10	8
Greater the second	16	22	20	21	23	25

ELASTICITÀ INDICE E MEDIO

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	17	13	15	12	13	17
Greater the second	16	20	18	21	20	16

COESIVITÀ

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	13	10	16	21	11	15
Greater the second	20	23	17	12	22	18

COMPATTEZZA

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	6	8	8	13	18	15
Greater the second	27	25	25	20	15	18

SECCHEZZA

	A - F	A - E	A - D	F - B	F - D	B - D
Greater the first	3	7	11	21	27	15
Greater the second	30	26	22	12	6	18

By analyzing the frequencies of the columns of the tables using a Chi Quadro test, significant frequencies are highlighted in yellow.

By analyzing them, it can be noted that:

- The hardness attribute is perceived better on the palate than the touch, this is due to the fact that people are daily trained to judge the hardness of a food using bite and chewing compared to touch.
  Durezza Indice points out each pair of recipes as significantly different and identifies D as the hardest recipe and A as the softest.

- The *Elasticità labbra* has three comparisons with significant frequencies in which D and E turn out to be perceived as the most elastic, but unfortunately the attribute does show neither significance findings nor consistence with the results of Kolmogorov's test on texture Analyzer data.

- About *Secchezza*, the F recipe is perceived as drier. The result is not in accordance with the rank found by tasters' votes, but in accordance with the percentages of the weight drop.

- About *Compattezza*, the attribute shows that the A recipe is the easiest to swallow than other recipes.

Compared to the previous taste, it is possible to observe the increase of statistically significant frequencies: this can be due to both clearer evidence between the recipes of the loaves and an improvement in the awareness of tasters during tasting.

10.Planning the third taste

The third taste would investigate how difficult is to analyze bread samples using sensory evaluation like votes. As usually, during the taste four samples are given to the panelists, but this time two of them have the same recipe and the panelists do not know that.

9.1 Choice of the bread recipes

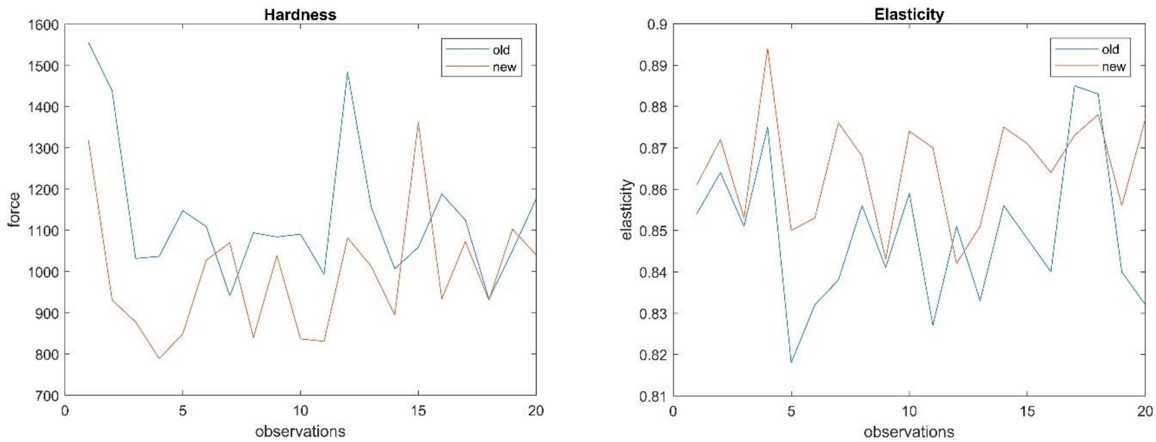
For the third taste the recipes used are A, the standard one, D, the one that during the previous taste is voted as the hardest, and a new one:

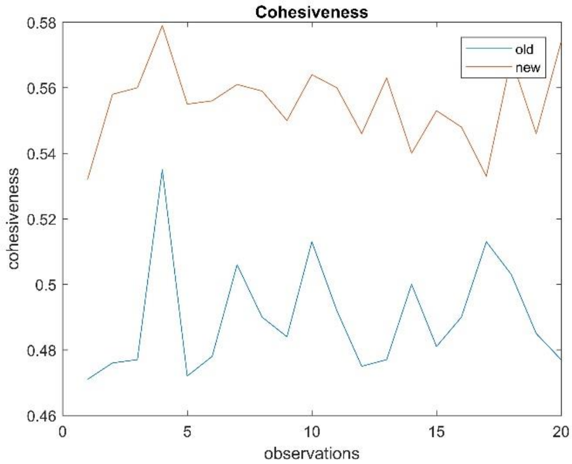
- **G**
  This recipe is obtained by adding a new enzyme to the standard recipe in order to obtain a very soft bread.

As the secondo test, bread samples are shaven (height 25 mm) to give aligned Texture profile analyzer curves.

In order to accentuate differences between breads, the D samples having a shelf life of one month are used: the reason is that, as expected, the hardest bread after one month packed and kept in stock, is harder than the beginning and Kolmogorov test on TPA data of the two breads confirms that.

- ➢ K. test on TPA data of the 2 breads confirms that the bread, which is packed after one month and kept in stock, is harder in comparison with the one at the beginning. Thus, the expectation is confirmed.





As expected, the old one is harder, less cohesive and less elastic.

9.2 Choice of the software layout for the taste

The layout does not change with compared to the second taste, but this time the panelist could see his vote while he is marking the line of the bread sample. Votes are again real numbers, but with only one significance figure after the decimal point.

11. Third taste results

First of all, hardness sensory data are analyzed because, in the previous tastes, it seems to be the best understood attribute.

RECIPE	Durezza masticazione	Durezza incisivi	Durezza indice
A1-A2	Not Statistical Different	Not S. Different	Not S. Different

Kolmogorov's test shows that panelists do not see the difference between between bread samples of the same recipe. *Durezza Masticazione* is chosen as the most informative attribute for the third taste because *Durezza Indice* and *Durezza Incisivi* do notvshow any statistical difference between the recipes. So, comparing the different votes of *Durezza Masticazione,* each pair of recipes are analyzed:

RECIPE	Sensory hardness
A1 - A2	Not S. Different
A1 - D	Not S. Different
A1 - G	Not S. Different
A2 - G	Not S. Different
A2 - D	S. Different
D - G	S. Different

Tasters recognize the D recipe as the different one, except in the case of A1; A2 is perceived softer than the D one.

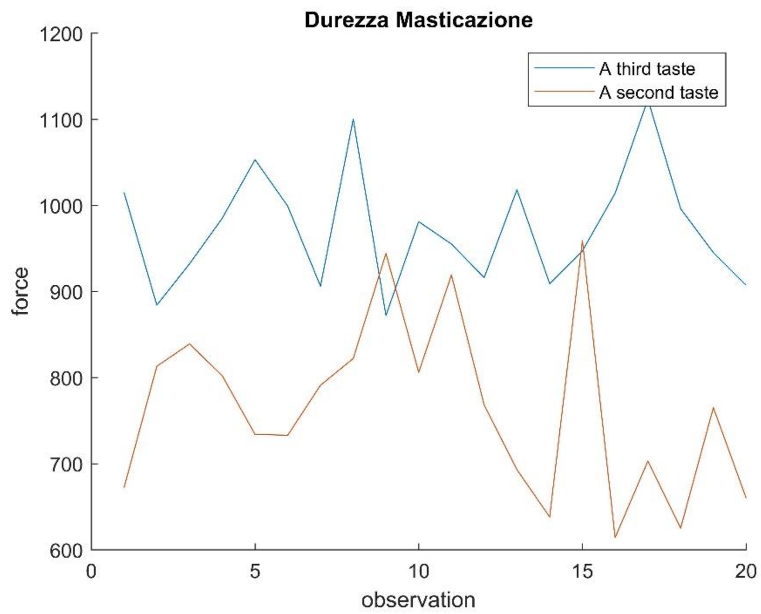


Figure 42 Comparison between the two standard recipes

The figure above shows that the standard recipe of the third taste is harder than the standard one of the second taste, even if the recipe is the same, and Kolmogorov's test confirms that.

Unfortunately, it is not possible to compare the TPA data of the third taste with the one of the first taste because the bread sample have not been shaved.

The result could seem illogical, but as it was stated before, bread matter is unpredictable. This is the reason why it is so difficult to evaluate that in a sensorial way: there are no anchors (bread sample taken as reference) because it is very hard to obtain two identical breads, with the same identical characteristics. So, the figure could explain why the tasters consider one time A and D as different and the second time they do not.

Moreover, the force curves of the two recipes are analyzed using **fda** R package:

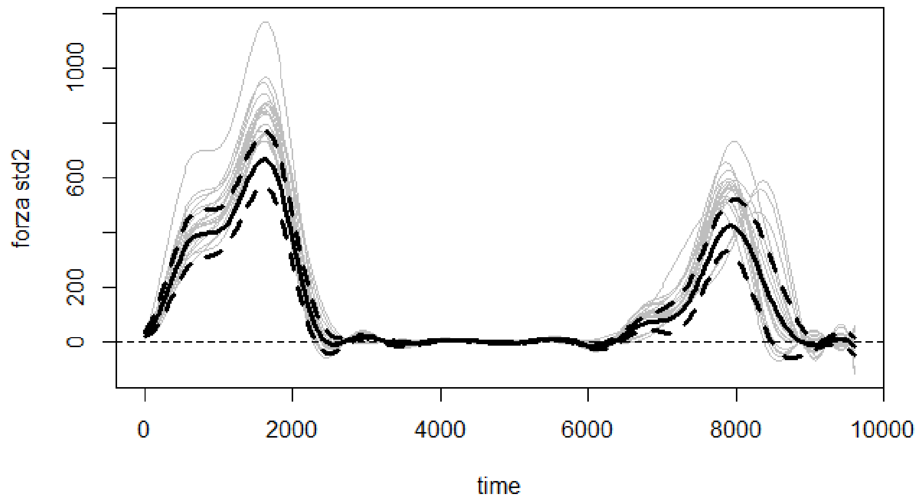


Figure 43 A recipe of second taste

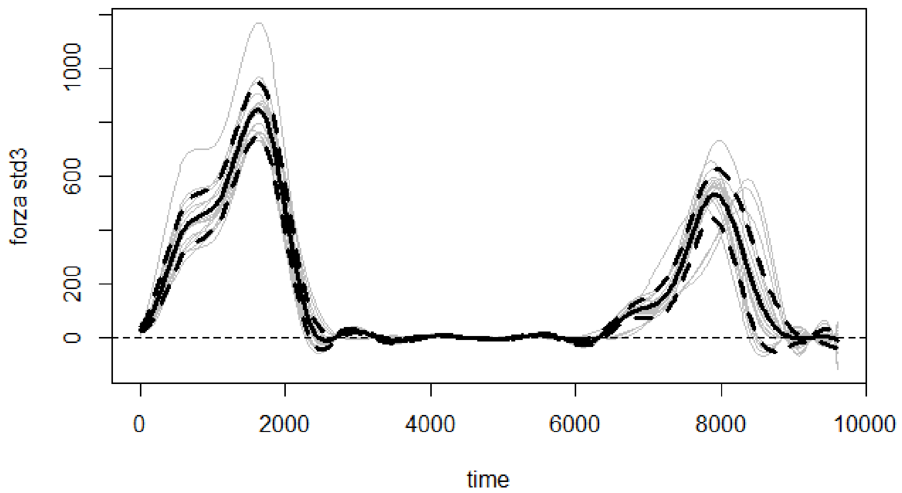


Figure 44 A recipe of third taste

Then **t.perm** test is applied to the curves:

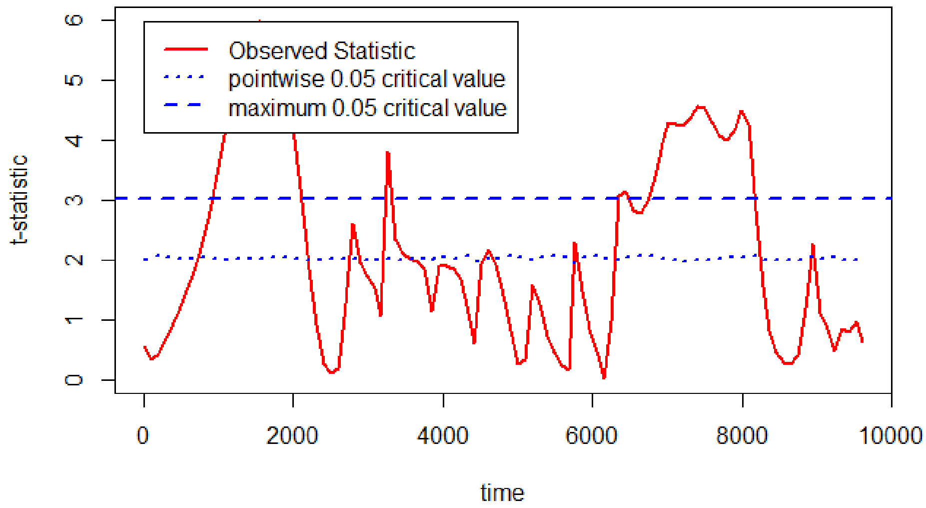


Figure 45 Output of t.perm test

Accordingly to the more specific analysis on the curves, they are different with regard to the two peaks of curves. Indeed, the curves differ on the components around the two peaks of force.

Moreover, the TPA data of the three recipes are analyzed by Kolmogorov's test and the instrument underlines that all the bread recipes different:

RECIPE	TPA hardness
A – D	Statistical Different
A – G	S. Different
D - G	S. Different

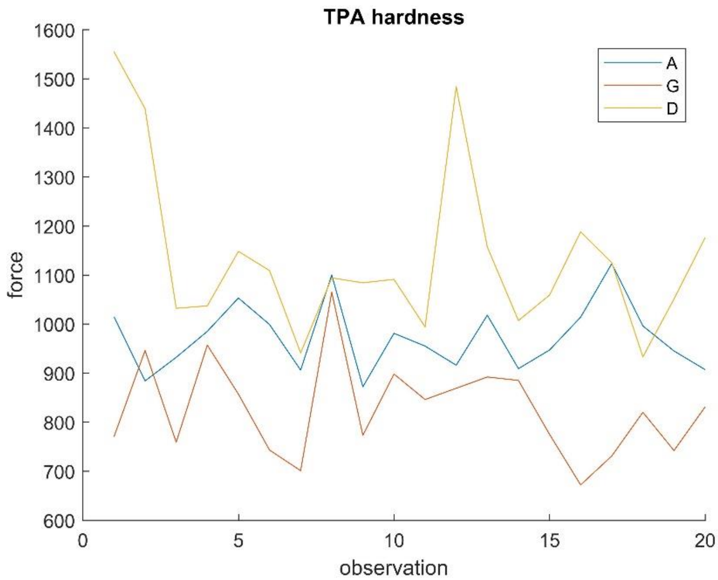


Figure 46 TPA bread measuments

During the second taste analysis, differences on average hardness greater than 235.05 g AND make the tasters able to perceive statistically different hardness and match with those of the Texture Analyzer. In this case the average of the hardness measured by Texture Profile are:

Mean D: 1135,20 g

Mean G: 826,60 g

Mean A: 972,85 g

Δ (D - G) = 308,25 g

Δ (D - A) = 152,35g

Δ (A – G) = 142,25g

By looking at the averages it is possible to observe that how the D recipe is harder than the G one: in fact, the difference is greater than 235,05g so, as expected, the tasters see the bread as different. The same analysis cannot be done for the other two comparisons because the differences are very similar, so the panelist are not so accurate.

The frequencies of the comparison of each pair of recipes are given below and they are analyzed by a Chi Quadro test:

DUREZZA INDICE

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	13	13	14	13	17	20
Greater the second	17	17	16	17	1	10

DUREZZA MASTICAZIONE

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	3	17	14	25	24	15
Greater the second	27	13	16	5	6	15

DUREZZA INCISIVI

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	10	17	18	22	20	16
Greater the second	20	13	12	8	10	14

The significance frequencies are highlighted in yellow. By observing the tables above, it is proved that the attribute *Durezza Masticazione* is the more significant and that the recipe D is seen as the hardest one. So, the pairs comparison is more precise than the comparison between the votes. It can be also seen that the other two attributes of hardness, as noticed before, do not present significant result.

Below the plots of the force measured by Texture Analyzer on the recipes D and G are shown:

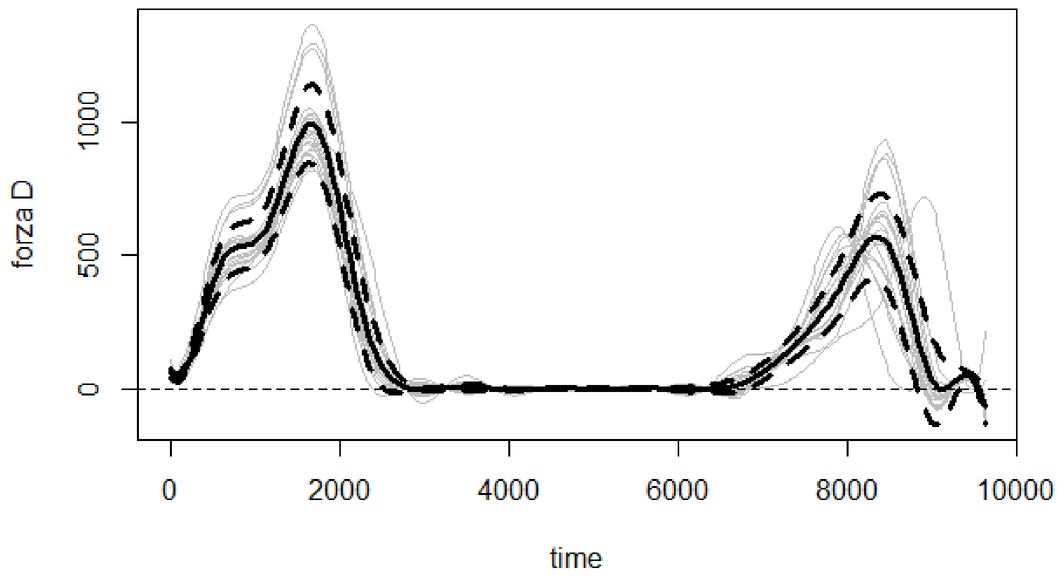


Figure 47 Force curve of D

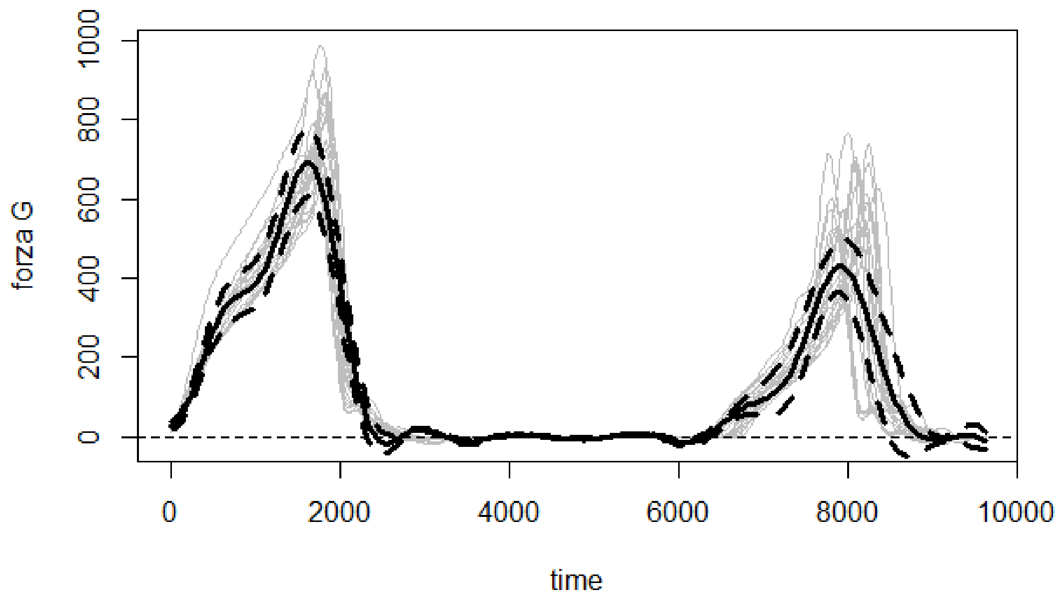
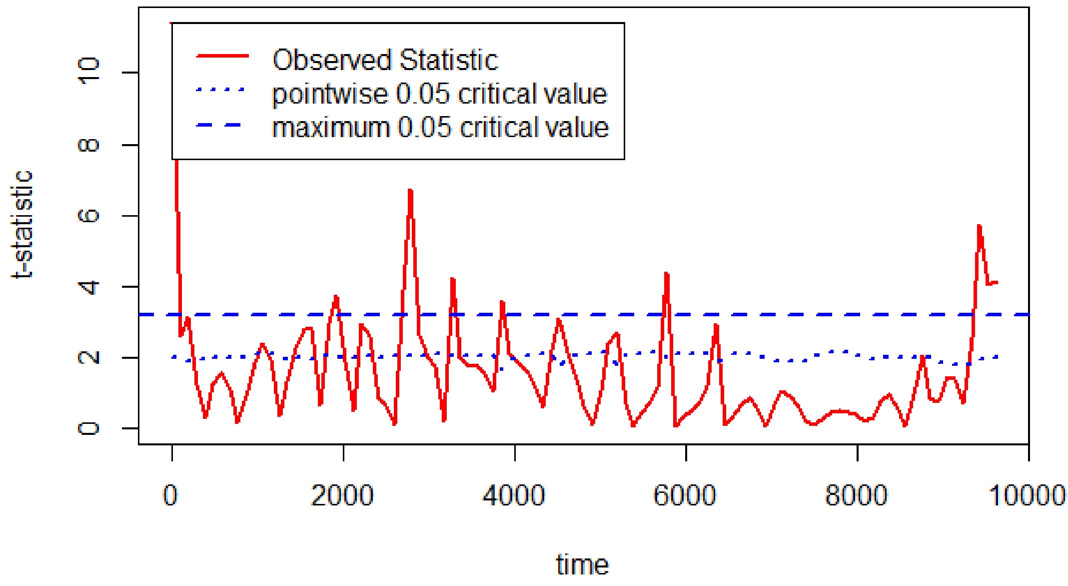


Figure 48 Force curve of G

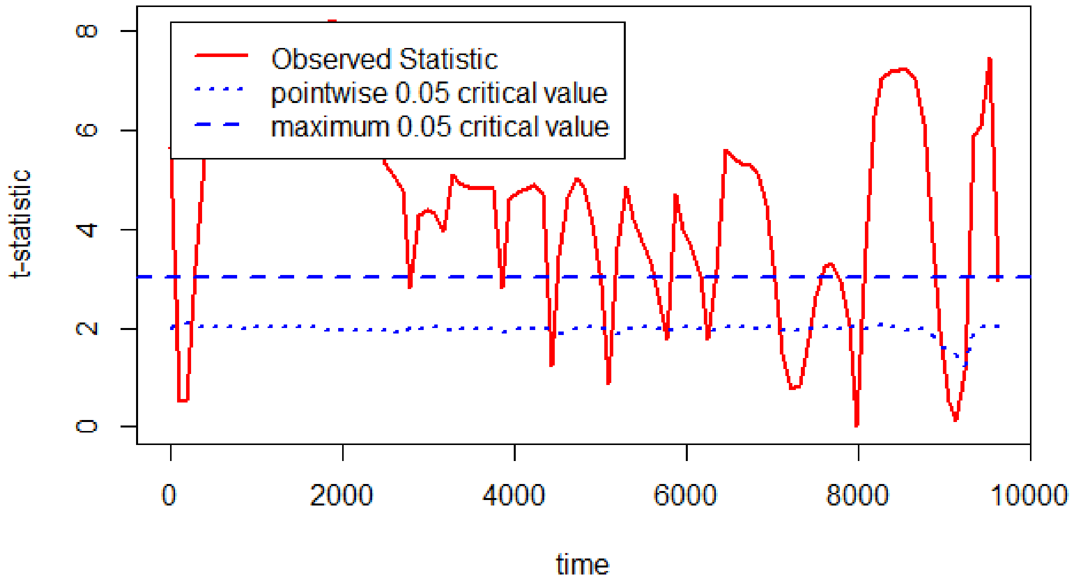
Then **t.perm** test is used to compare each pairs of recipe:

A – D



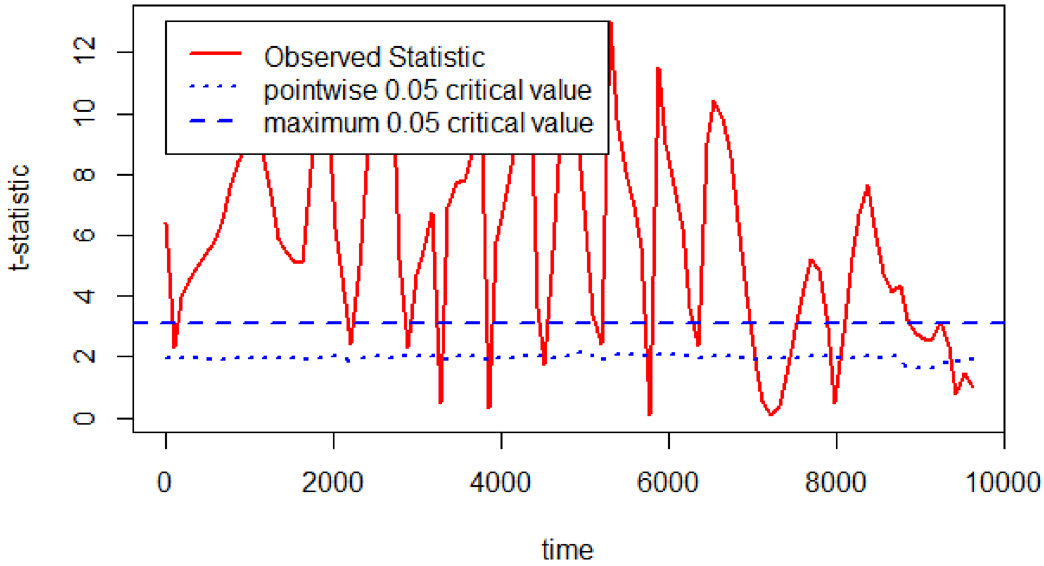
The plot shows that the test sees the recipes different, but not so much: indeed, the observed statistic is almost below the maximum critical value and it is consistent with the observations of the panelists.

D – G



The plot shows that the test sees the recipes very different and that is consistent with the fact that tasters, during the comparison of each pair of recipes, perceive the recipe D as the hardest.

A – G



The test sees the two recipes as very different, but the panelists do not perceive the same thing: that is consistent with the fact that TPA can measure difference between the recipe more precisely.

Finally, the frequencies tables of the comparison among each pair of recipes are given below:

ELASTICITÀ INDICE E MEDIO

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	17	15	11	8	11	12
Greater the second	13	15	19	22	19	18

ELASTICITÀ INDICE E POLLICE

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	13	12	15	14	15	13
Greater the second	17	18	15	16	15	17

ELASTICITÀ TRA LE LABBRA

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	13	9	16	11	12	16
Greater the second	17	21	14	19	18	14

COESIVITÀ

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	5	10	15	19	20	13
Greater the second	25	20	15	11	10	17

COMPATTEZZA

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	10	14	14	20	18	16
Greater the second	20	16	16	10	12	14

SECCHEZZA

	A1 - D	A1 – A2	A1 - G	D – A2	D -G	A2 - G
Greater the first	7	13	13	20	19	14
Greater the second	23	17	17	10	11	16

As can be noticed, the tables do not show many significant frequencies (the significant frequencies are measured by Chi Quadro test and highlighted in yellow). Tasters find it difficult to evaluate these attributes, or they are not well logically linked with the TPA features.

Even if there are not so many significance frequencies D recipe is generally evaluate as drier and less elastic and this is coherent with the fact that it has a shelf life of one month.

Instead A1 and A2 samples are not seen as different, coherently with the hardness attributes.

12. Comparison between the sensory taste

As explained before, the way of testing the bread samples differs among the three tastes: during the first taste, the attributes are placed on a wheel in order to emphasize the order of the attributes that the panelist has to observe, so the panelist votes all the attributes each time only for one sample. Then, after he votes, the grade remains on the line where the taster marks the votes, so he can focus on the order he is placing bread samples. Moreover, the votes are real numbers with two significance figures after the decimal point.

By analyzing the first taste results, it can be supposed that real numbers with two significance figures after the decimal point are too precise for panelists, because no significance differences are observed on the sensory data. So, during the second taste, integer number are used. These ones emphasize the difference between bread samples, but they leave less freedom to the tasters because, in this case, they only have nine possible values to vote.

Furthermore, the wheel is replaced by a test where the taster has to vote, for each attribute, all the samples together, in order to emphasize the importance of the comparison among them. Moreover, images are added to the descriptions in order to help the panelist remember the definition of each attribute.

Then, during the third taste, real numbers with one significance figure after the decimal point are used as votes. Finally, during the third taste, the real number appears above the line where they place the cursor, in order to help the panelists to better understand the way they are voting.

At the end of the last taste, panelists thanks to have seen the vote while they were voting: it help them to be aware of how they are evaluating samples. They also see, after they finished the second and third taste, a page where simple statistics, as average and standard deviation, show the trend of the votes of the panel for each sample and for each attribute. So, the more tasters feel involved in the sensory process, the more is explained to them, the better they understand the more they are interested in doing the best they can.

Moreover, the data of the third taste give significance evidence, so probably real numbers with one significance figure after the decimal point as votes are the more reasonable.

13. Conclusions

After having analyzed and compared sensory data and rheological data, at the end of the experiment it is possible to claim that hardness is the sensory attribute better understood: in fact, significant frequencies of the table, where the recipes are compared in pairs, are observed only in relation to this one. But it has to be considered that hardness has three different sensory types of attribute, *Durezza Masticazione*, *Durezza Indice* and *Durezza Incisivi*, and usually only *Durezza Masticazione* and *Durezza Indice* gives a larger number of significance frequencies. So, probably time when the panelist tastes *Durezza Incisivi* lasts too little, it is very hard to evaluate it precisely. Conversely, panelist tasting of the other two attributes lasts more because of the mastication or because the taster can press slowly the bread sample while he cannot bite slowly.

Moreover, after the three tastes, bounds are found by linking rheological measurements and sensory data: a difference between the averages of Texture analyzer greater than 235.05 g makes the tasters to perceive statistically, by Kolmogorov's test, different hardness that agree with those of the Texture Analyzer. However, differences less than 85.75 make the tasters and the Texture Analyzer not able to perceive statistically different hardships. The bounds are confirmed also by the third taste in which, for a difference greater than 235,05 g, tasters perceive difference between bread samples, while for difference of the average hardness around 100 g tasters do not see the bread sample as difference, while the instrument does.

Instead, searching for a rank correlation, Spearman's and Kendall's correlation, between the rank of the sample breads given by the votes of the panelist and the rank of the sample breads given by the averages of the texture analyzer measurements does not give significant results. But, at least, Spearman's and Kendall's test show that panelists agree with the hardness measurement of the texture analyzer because of the positive correlation, while for elasticity is not true.

In fact, probably elasticity is too difficult to evaluate by the tasters: indeed, the structure of the bread sample changes after two or more compression, so if the taster tries to evaluate more than two times the sample, he could be confused about the evaluation. Moreover, the texture analyzer does not measure directly elasticity, but it results from the ratio between the times of the two compressions, so probably the fingers and the lips of the panelist are not sensible enough to the time or the way the sample return to its primary position.

As proof of this, even the frequencies by analyzing the table where the recipes are compared in pairs do not give significance results and that is the simplest way to analyze data and, if the frequencies don't show significant difference, also Kolmogorov's test does not show significance results.

Then, dealing with the other sensory attributes *Secchezza*, *Compattezza* and *Coesività,* it can be observed that their definitions are linked: indeed, if a bread sample is drier, then it is also more compact and less cohesive according to the panelists. But probably they are well logically linked to the cohesiveness measured by texture analyzer. In fact, during the planning of the first taste, panelists were already not satisfied by the definition of *Coesività* because it is not a common word used in everyday life; furthermore, it is not so easy to replicate the measurement of texture analyzer by hands. Finally, texture analyzer does not measure directly cohesiveness, but it results from the ratio between two areas.

Furthermore *Secchezza* is not difficult to understand but it can not be measured by texture analyzer. Consequently, it was added for the sake of completeness to the attribute. However, as explained before, it is measured as weight loss, that indicates how much water was lost from bread after spending a night in the stove at 105 degrees, that is the humidity of the bread. So, the measure of the water lost during this process is inversely proportional to the feeling of dryness perceived on the palate, which measures how much saliva the bread absorbs. Therefore, they are inverse and well logically linked, but they are not properly the same thing.

Moreover, these three attributes are also linked to hardness. Indeed, a drier bread sample is perceived also as harder, but unfortunately, a lack of statistical significance results is observed neither related to the table of frequencies of comparison between the pairs of recipes, nor using test correlation on the rank or Komogorov's test on the votes.

Dealing with the third taste, the goal is to verify if tasters, who did two tastes and are trained, could be aware that two of the four samples came from the same recipe. The result is very interesting: panelists do not see the two samples of the same recipe as different, either by analyzing the votes using Kolmogorov's test and the table of frequencies of comparison between the pairs of recipes. They also recognize the hardest recipe chosen on purpose as the easier to find in order to help them because of the two equal samples.

They show one more time that they could recognize well hardness and the hardest recipe. However, according to the bound mentioned before, they cannot distinguish bread samples that are too similar in average measurements of texture analyzer.

Moreover, the third taste shows that A recipe of the second taste and A recipe of the last taste differ in hardness measured by texture analyzer. This is related to the sensitive matter of bread, as told before. So, the experiment shows that when people analyze bread it must be taken into account that is not possible to replicate exactly a recipe, so different analysis and results can be obtained by the same recipe done at different times.

About **fda** R package, it can be said that analyzing (time,force) curves of each recipe is more interesting than considering only the two peaks of force. First of all, the **t.perm** test of the package is as precise as Kolmogorov's test on the first peak , in fact if Kolmogorov's test sees the recipes as different also **t.perm** does. Secondly it allows to analyze more precisely the differences between the hardness of each recipe.

For example, during the second taste, F recipe differs from the A in insignificant places of the curve, namely they differ in the places where the Texture Analyzer probe rises after compressing the bread. Thus, it can be explained as F recipe returns a force to the probe that decreases more slowly than the A. Also E recipe differs from A and F recipes by this particular behavior: it differs from the A and F curves in points of the curve where the force results null for each recipe. In those sections, in fact, the probe is returning to its initial position and then descending, so it measures no force. It is important to consider that in those traits (after component 2000 and before component 8000) the D recipe does not differ statistically significantly from the recipes A and F, as is logical to expect. This behavior is present only in the analyses in which the E recipe is involved. **fda** package allows also to analyze the derivatives of the curves, so it can be noticed that the first derivatives of the curves of each recipe differ from E. Therefore, it is not a difference related to the peaks of force reached by the probe but to the different growth and decrease of the curve. So, more accurate analysis can be done. Unfortunately, they can not be easy related to sensory evaluation. This is the reason why nothing more that **t.perm** test is used for the analysis of the bread samples data.

Moreover, **fda** package allows to see easily the average curves of the curves registered by texture analyzer for all the twenty samples analyzed for each recipe. So, a first analysis can be done only by looking at them to see which recipe is the hardest or the softest.

Unfortunately, this reasoning cannot be done with elasticity o cohesiveness because they are not measured directly by the instrument but is obtained, so there are not (time, cohesiveness) or (time, elasticity) curves and this is limiting.

Bibliography

1. Piotr Kokoszka, Matthew Reimherr, Introduction to functional data analysis (2017)
2. Walter A. Shewhart, Samuel S. Wilks, An introduction to categorical data analysis (2019)
3. TPA document – Rheological Laboratory of Soremartec Spa
4. Harry T. Lawless, Hildegarde Heymann, Sensory evaluation of food (2010)
5. Morten Meilgaard, Gail Vance Civille, B. Thomas Carr, Sensory evalutation techniques (1999)

Sitography

1. https://en.wikipedia.org/wiki/Kolmogorov–Smirnov_test
2. https://www.mathworks.com/help/stats/kstest2.html
3. https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient
4. https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient
5. https://en.wikipedia.org/wiki/B-spline
6. https://en.wikipedia.org/wiki/Chi-squared_test#Pearson's_chi-squared_test
7. https://en.wikipedia.org/wiki/Contingency_table
8. https://www.compusense.com/en/sensory-services/
9. https://en.wikipedia.org/wiki/Retrogradation_(starch)
10. https://en.wikipedia.org/wiki/Starch
11. https://en.wikipedia.org/wiki/Principal_component_analysis