Microevolution Processes in Fieri Population as Estimated by Isonymy Techniques

Silvana Turku

Ph.D University of Elbasan, Albania,
e-mail: silvanaturku@yahoo.com

Hyrida Basha

Msc.University of Tirana, Albania
e-mail: hyridabasha@yahoo.com

Ilia Mikerezi

Prof. Dr. University of Tirana, Albania
e- mail: imikerezi@yahoo.com

Abstract *The genetic structure of Fieri population was analyzed through the distribution of surnames obtained by a sample of 3'471 marriages registered in the period 1946-2006. The data were recorded for every five years starting from 1946 (than 1951, 1956 and so on) to 2006. Estimations of Fisher's α, an indicator of richness of surnames, and v of Karlin-McGregor, indicator of migration rates, were obtained for both natives and immigrants. These indicators showed higher values in immigrants. The same was demonstrated even by log2k-log2S regression line. Two different estimations of relationships between all possible combinations of cohorts born in nine different decades were obtained through Euclidean distance and Lasker's coefficient of relationship, respectively. Comparisons of Ft, coefficient of total inbreeding and Fisher's α indicated that any decrease of Ft was followed by an increase of α as the result of recent migration.*

Keywords: *isonymy, surname distribution, marital structure, Euclidean distance and Lasker's coefficient.*

Introduction

In recent decades, surnames have been frequently used as genetic markers to evaluate different factors that influence the genetic structure of human populations. Considering that in most societies surnames are transmitted through the paternal line, they can be considered as neutral alleles of one locus on Y chromosome (Barrai et al, 1987; Barrai et al, 1989: Cavalli and Edwards, 1967; Zei et al, 1983). This property of surnames allows the investigation of the temporal trends in the variability of the population (Cavalli and Edwards, 1967; Rodriguez-Larralde, 1993).

The study of Fieri population is part of an investigation and it aims the estimation of the genetic relationships among all different local populations in Albania by using different methods. Therefore, it will be possible to cast more light on genetic history of Albanian population.

Fieri is located in South-Western part of Albania in a rather isolated geographical position. In this framework, there is an absence of genetic investigations in this population. Therefore, it seems interesting to study the frequency distribution of surnames in order to obtain population dynamics indicators especially related to migration. Moreover, this study aims to test whether the geographical position and migration patterns have affected the inbreeding level. This is the first the investigation of Fieri population, and, as in other researches, it will continue in the future by using other alternative genetic methods.(Gjika, 2005).

Materials and Methods

The records of 3471 marriages of individuals resident in Fieri were obtained from the registers of the Municipality. For each individual (subject) surname, gender, year and place of birth, and place of origin were recorded. A total of 6939 different surnames were counted. Among them, 4622 surnames were subjects born in Fieri, and 1317 surnames were subjects who had immigrated from elsewhere. All the records were grouped in 9 classes according to the date of birth of the bearers. In order to analyze chronological trends in marital structure the data were grouped in 7 classes according to the period of marriage.

The distribution of surnames was studied by sorting the main files by place of birth and period of birth. For each of both groups that were formed by the above sorting, several parameters were estimated. The unbiased isonymy (I) and Fisher's α were estimated according to Rodriguez-Larralde at al (1993):

$I = \Sigma_i q_i^2 - 1/n$ (q_i - relative frequency of the i-th surname in each cohort, n – total number of surnames in the same cohort) where the summation is over all surnames and $\alpha = 1/I$ (Zei G et al 1983)

Karlin-McGregor's v was calculated according to:

$v = (1-B)/ [B (n-1)]$, where *B* is the biased isonymy estimated as:

$B = \Sigma_i q_i^2$ (q_i - as defined above)

The relationship between cohorts born in different decades was studied according to the following formula:

$R = \Sigma_k (q_{ik}q_{jk})$, where q_{ik} and q_{jk} are the relative frequencies of surname *k* in cohorts *i* and *j* respectively. This parameter will be considered here as Lasker's coefficient.

The Euclidean distance between cohorts was also estimated according to Cavalli-Sforza et al (1967):

$D = \sqrt{(1 - \cos\theta)}$, where $\cos\theta = \Sigma_k \sqrt{(q_{ik}q_{jk})}$, with q_{ik} and q_{jk} as defined above.

Both groups of immigrants and natives to Fieri were compared by the slope of regression lines fitted to $\log_2 k$-$\log_2 S$ distributions according to where S is the number of surnames appeared k times (Barrai at al, 1987).

In order to analyze the endogamy, the marriages were subdivided into three groups according to the birthplace of the spouses: a) both partners born in Fieri, b) one partner born in Fieri and the other born elsewhere, c) both partners born elsewhere. Changes in premarital migration were studied by measuring the linear geographic distance (in kilometers) between birthplaces of mates. The total inbreeding (Ft) and its random component (Fr) were estimated from marital isonymy according to Crow at al (1965), and Crow at al (1980):

$Ft = Fr(1-Fn) + Fn$

Results and Discussion

1. Surname distribution by place birth

The distribution frequency of surnames was given as a function of the number of individuals carrying the same surname. (Caravelle at al, 1999)

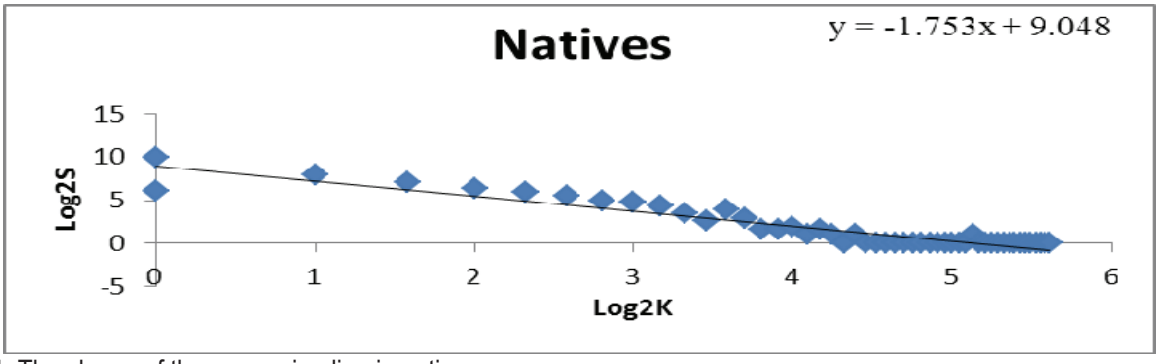

Fig 1-1: The slopes of the regression line in natives

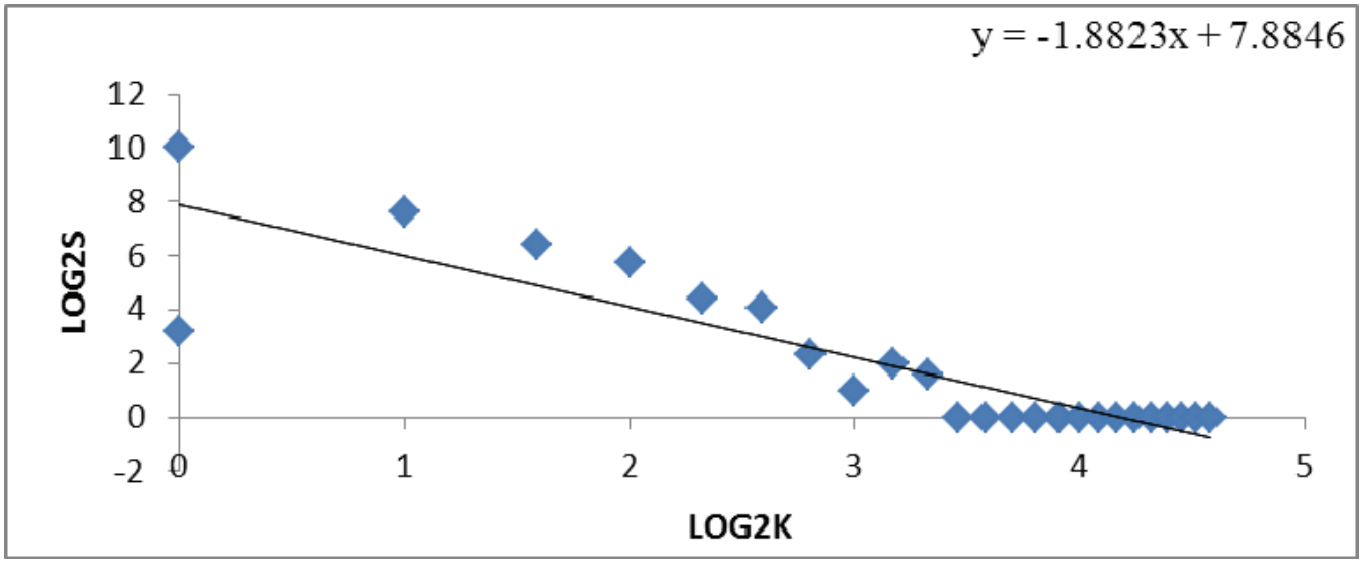

Fig 1-2: The slopes of the regression line in immigrants

The graphs in the figure 1.1 and 1.2 show that the slopes of the regression line change significantly with the origin: in immigrants it had the value b = - 1.8823, whereas in natives the value of the slope was b = - 1.753. The high number of surnames with unique frequency or with a small one among immigrants is a good indicator that migration occurs generally by single individuals or small family groups.

2. Indicators of temporal changes in the group structure:

Fisher's α is an indicator of the richness of surnames because of the accumulated migration, while Karlin-McGregor's v is an indicator of the immigration since it is correlated to the surnames that appear only once.

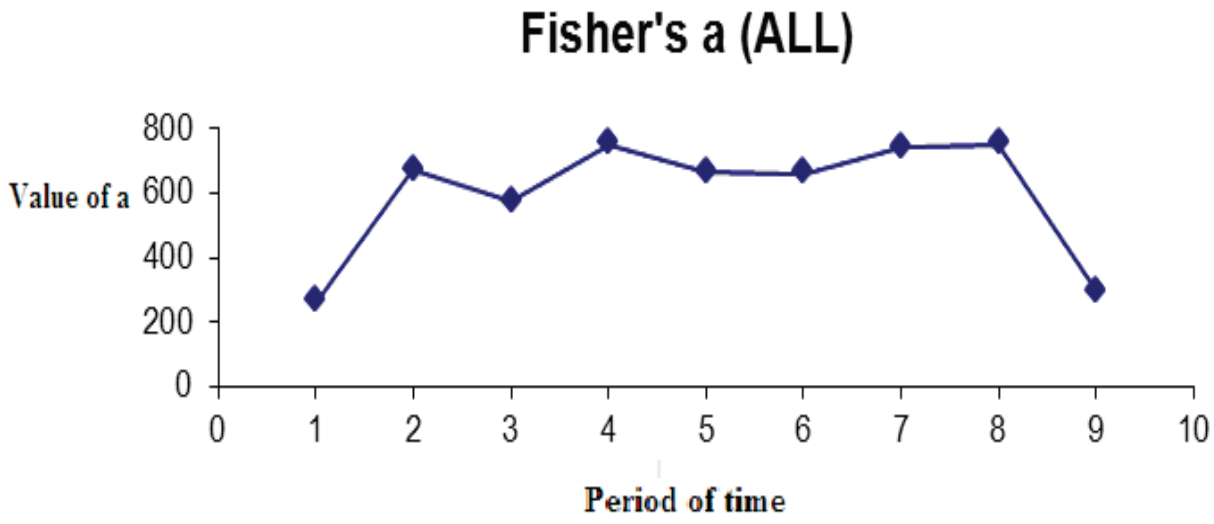

Fig 2-1: Fisher's α in different decades for both natives and immigrants (All)

The figure 2.1 show that the highest values of α corresponded to the second (1911-1920), fourth (1931-1940) and to the eighth (1971-1980) periods, while the lowest value is reached during the third (1921-1930) and ninth period (1981-1992)

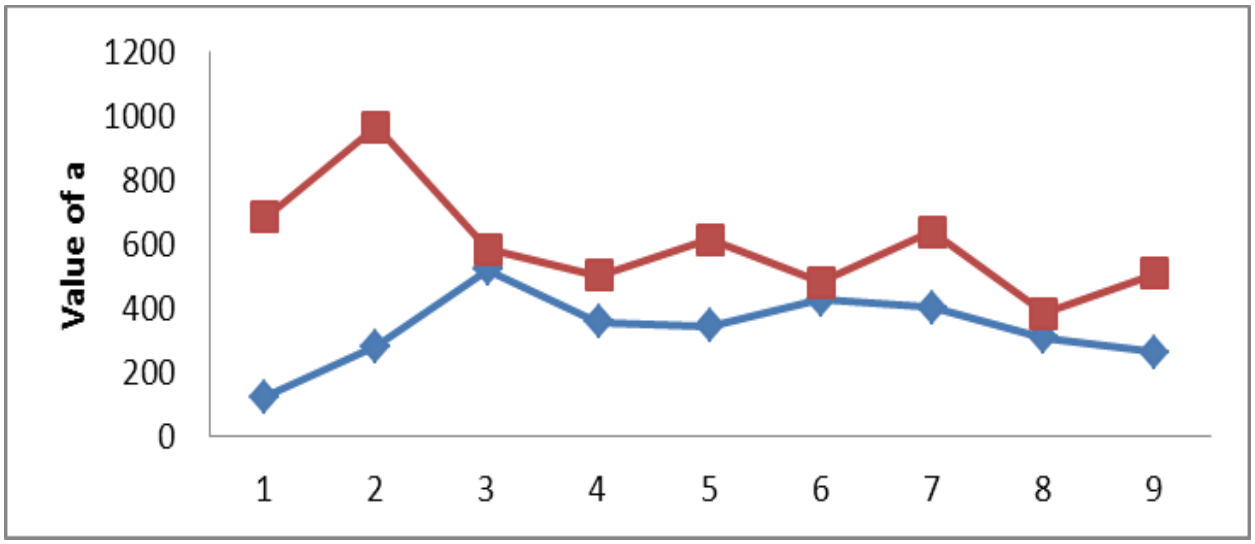

Fig 2-2: Comparison of Fisher's α between immigrants and natives in Fier

Similar values of α for both residents and immigrants (red line in figure 2.2) groups were obtained for the first three periods of time, while the values of α were higher in immigrants in the consecutive periods. Since migration generally moves single persons or small family groups, the richness of surnames and consequently surname variability was higher among immigrants than in natives.

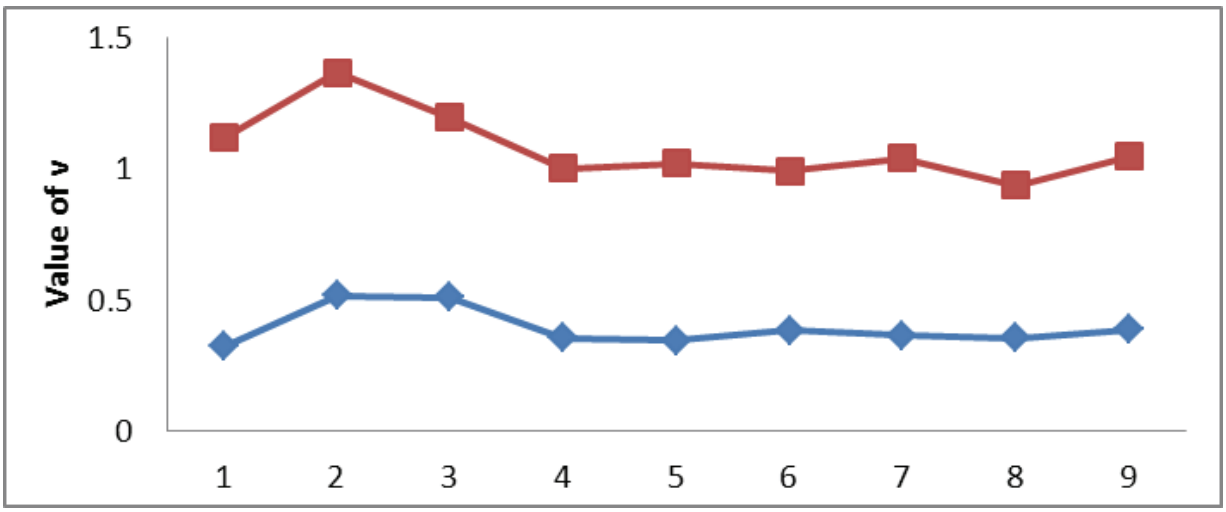

Figure 2-3: Comparison of Karlin-McGregor's v between immigrants and natives in Fier

The values of Karlin-McGregor's v were higher among immigrants (red line in figure 2.3) compared to natives (blu line in figure 2.3). This is a consequence of the fact that v is correlated to the surnames of unique frequency.

3. Relationship between cohorts as measured by Lasker's coefficient and Euclidean distances:

The distances were calculated separately for natives, immigrants and for all together. If immigration through different decades was random regarding the surname, Lasker's distance between cohorts was expected to be nil, while Euclidean distance should be one. Lasker's distance values tended to be lower among immigrants compared to natives, while for Euclidean distance the relationship is inverted. It was observed that the relationship between the cohorts is decreasing in the following decades for both Lasker's and Euclidean distance.
Rodriguez-Larralde *et al*. (1993) estimated Lasker's and Euclidean distances in relationship with time periods using data from Municipality of Perugia, Italy, for the period 1900-1990. They observed that the relationship among the cohorts decreased with time, showing that the population became more heterogeneous with time.

Table 3-1: Euclidean distance

	1X2	1X3	1X4	1X5	1X6	1X7	1X8	1X9
1-cos θ	0.762788	0.704481	0.748473	0.781681	0.777937	0.845525	0.841184	0.81895
		0.781299	0.736238	0.80027	0.791874	0.829872	0.864291	0.832217
			0.65754	0.695241	0.727886	0.726384	0.792284	0.786029
				0.624121	0.570214	0.656241	0.700181	0.723387
					0.604987	0.608381	0.701061	0.71827
						0.57881	0.644466	0.694316
							0.613892	0.60872
								0.637453

Table 3- 2: Matrix of Lasker's coefficient

1X2	1X3	1X4	1X5	1X6	1X7	1X8	1X9
0.00119	0.001521	0.000733	0.000372	0.000948	0.000815	0.000401	0.000333
	0.000708	0.000874	0.000379	0.000435	0.00027	0.000258	0.000249
		0.000586	0.000752	0.001257	0.001048	0.000741	0.000549
			0.000577	0.000885	0.00044	0.00037	0.000362
				0.000643	0.000734	0.00068	0.000508
					0.001044	0.001021	0.000915
						0.001197	0.001185
							0.00146

4. Temporal trends in marital structure:

Results on temporal changes in marital structure could give us the possibility to estimate the trends of genetic isolation of the population under investigation. The values of endogamy are especially high during 1976 – 1986 when the mobility of the population was under total control according to the laws of the past regime. Lower values of endogamy are observed after the1996, which follows the social and political changes when the population mobility was much higher compared to the past regime. However, for the entire period of investigation still high values of endogamy were observed. This is related to the isolation due to the geographic location of Fier and the limited mobility of the population compared to the other regions of the country.

Table 4-1: *Temporal trends in the frequency of endogamous and exogamous marriages*

MARRIAGES		PER1(-56)	PER2(-66)	PER3(-76)	PER4(-86)	PER5(-96)	PER6(-06)
GROUP 1	NR	160	523	361	201	320	92
	%	0.415584	0.519364449	0.50419	0.436009	0.47976	0.393162
GROUP 2	NR	125	404	270	185	264	80
	%	0.324675	0.401191658	0.377095	0.401302	0.395802	0.34188
GROUP 3	NR	100	80	85	75	83	62
	%	0.25974	0.079443893	0.118715	0.16269	0.124438	0.264957
ENDOGAME	NR	160	523	361	201	320	92
	%	0.415584	0.519364449	0.50419	0.436009	0.47976	0.393162
ENDOGAME	NR	225	484	355	260	347	142
	%	0.584416	0.480635551	0.49581	0.563991	0.52024	0.606838

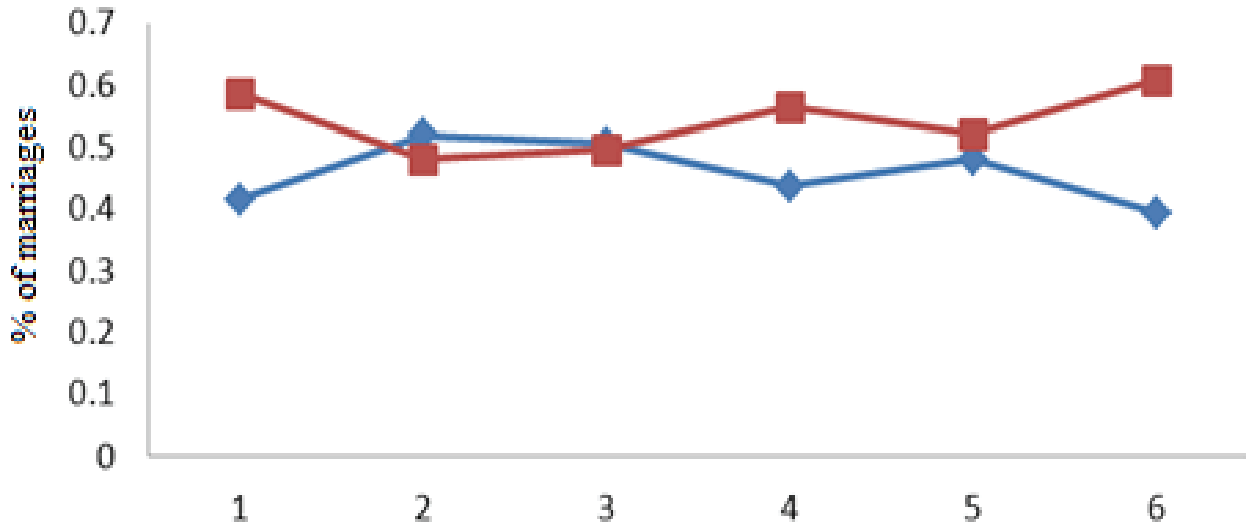

Fig 4-1: Frequency of endogamous and exogamous marriages

The value of endogamous (blu line in figure 4.1) marriages is still high despite the social and economic factors that have influenced the gradual decrease of inbreeding coefficient measured by isonymy.

5. Temporal trends in Isonymy and Inbreeding: Inbreeding has been widely considered as a good indicator of genetic isolation. Fisher's α, that is an indicator of the richness of surnames, could give additional information on the trends of the genetic isolation.

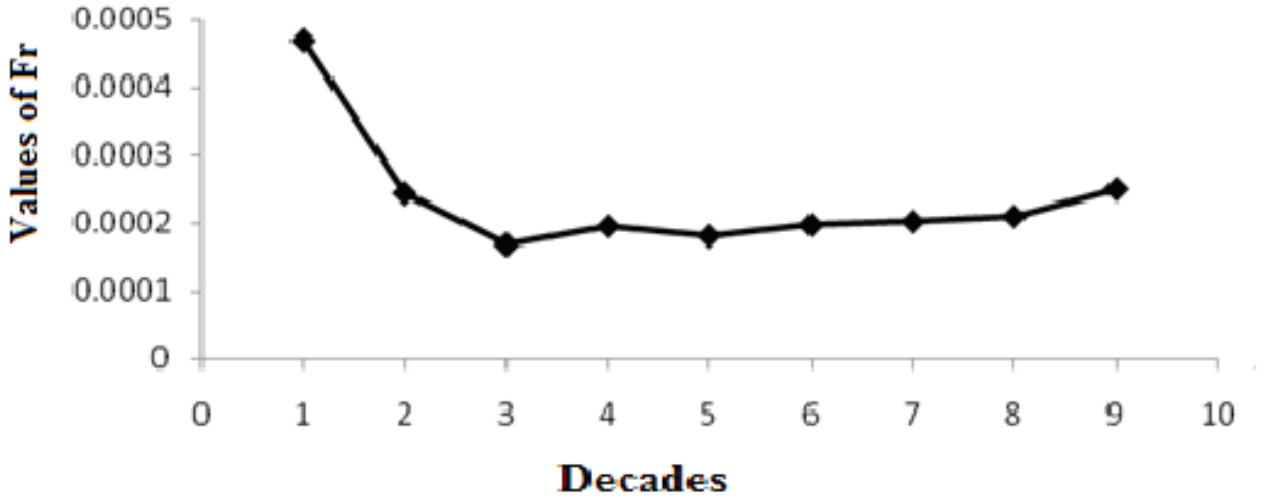

Figure 5-1: Temporal trend of random inbreeding Fr

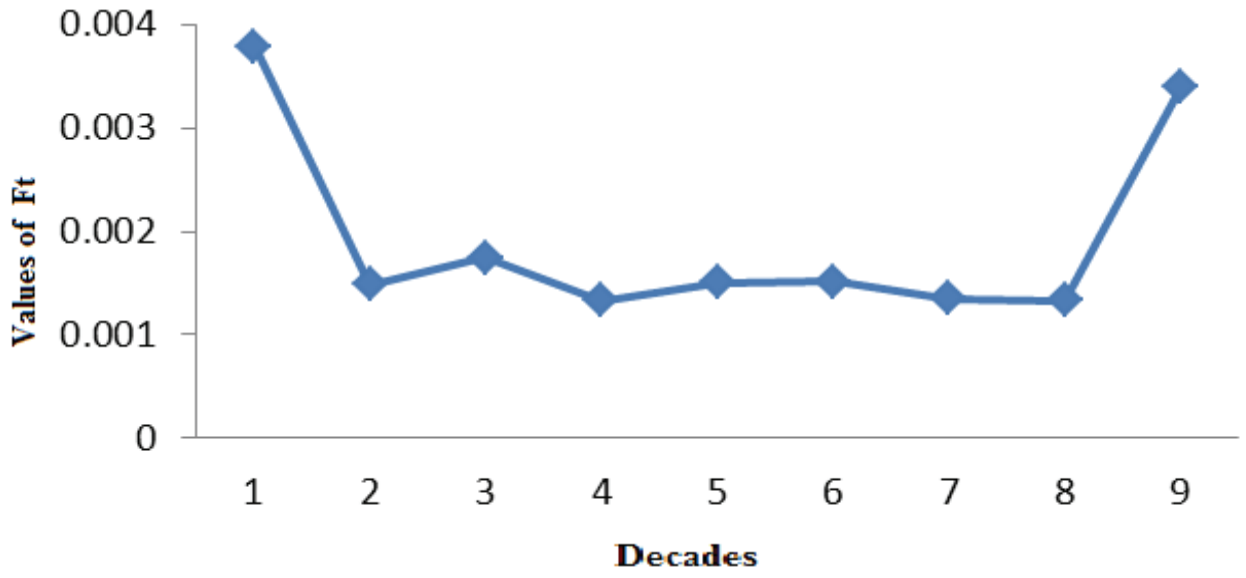

Figure 5-2: Temporal trend of total inbreeding Ft

In Figure 5.1 is shown the continuous increase of the inbreeding level in the population due to both geographic location and the gradual decrease of the population's mobility, especially during the period of last political regime. The highest value is observed during the first decade. It follows a slight decrease of the inbreeding level starting from the 6-th decade, due to the economic development of the region in this period of the past regime. However, in the last decade the decrease of inbreeding was higher because of free movement of the population following the political and social changes. As the result of the recent of migracion any change of Ft values in figure 5.2 is followed by an inverted change trend of Fisher's α value. For comparision reasons these changes are shown in this figure.(fig.5.2)

Bibliography

Barrai I, Barbujani G, Beretta G. et al. (1987) Surnames in Ferrara: Distribution, isonymy and levels of inbreeding. Anaquo;. Hum. Biol. 14:415-423.

Barrai I, Formica G, Barale R, Beretta M (1989) Isonymy and migration distance. Ann. Hum. Genet. 53:249-262.

Caravelle G, Tasso M (1999) An analysis of the spatial distribution of surnames in the Lecco area (Lombardy, Italy). Am. J. Hum. Biol. 11:305-315.

Cavalli-Sforza LL., Edwards A.W.F. (1967) Phylogenetic analysis models and estimation procedures. American Journal of Human Genetics, 19: 233-257.

Crow JF (1980): The estimation of inbreeding from isonymy. Human Biol., 52: 1-4

Crow JF, Mange AP (1965) Measurements of inbreeding from the frequency of marriages between persons of the same surnames. Eugenics Quarterly ,12: 199-2003.

Gjika I. (2005): "Fieri ndër shekuj", Fier

Rodriguez-Larralde A, Formica G, Scapoli C, Beretta M, Mamolini E, Barrai I (1993) Microevolution in Perugia: isonymy 1890-1990. An. Hum. Biol., 20: 261-274.

Zei G, Matessi RG, Siri E, Moroni A, Cavalli-Sforza LL (1983) Surnames in Sardinia, I. Fit of frequency distributions for neutral alleles and genetic population structure. Annals of Human Genetics, 47: 329-352.