PERFORMANCE OF TESTS FOR EQUAL VARIANCE

EMPIRICAL ASSESSMENT OF PERFORMANCE OF TESTS OF EQUAL VARIANCE IN THE PRESENCE OF WITHIN AND BETWEEN DEPENDENCE AND UNDER SMALL SAMPLE SIZE: APPLICATION TO CRANIOFACIAL VARIABILITY INDEX IN SMITH-LEMLI-OPITZ SYNDROME (SLOS)

By

GERMAN REANO-ALVAREZ, M.A, M.B.A

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University © Copyright by German Reano-Alvarez, January 2012

McMaster University MASTER OF SCIENCE (2012) Hamilton, Ontario (Statistics)

TITLE: Empirical Assessment of Performance of Tests of Equal Variance in the Presence of Within and Between Dependence and Under Small Sample Size: Application to Craniofacial Variability Index in Smith-Lemli-Opitz Syndrome (SLOS)

AUTHOR: German Reano-Alvarez, M.A (The Netherlands), M.B.A (Singapore)

SUPERVISOR: Dr. Jemila S Hamid

NUMBER OF PAGES: 1 TO 69

This thesis is dedicated to the memory of my dear mother, Grimaneza, who went to the presence of the Lord last year, to my devoted father, Cesar, to my loving wife Cecilia and, to my beloved children German Jr., Maria Grazia and Josue.

ABSTRACT

Craniofacial variability index (CVI), estimated by the standard deviation of z-scores obtained from craniofacial measurements of one or more individuals, is considered in the medical literature as a useful and relatively simple quantitative measure of the degree of dysmorphogenesis in the head and face.

CVI obtained from patients diagnosed with syndromes such as Smith-Lemli-Opitz Syndrome (SLOLS) is often compared with CVI for healthy individuals. Moreover, CVI is commonly used to compare degree of dysmorphogenesis among individuals and groups characterized by the presence or absence of certain syndromes and abnormalities. However, the type of comparison provided is often subjective with no statistical comparison of values provided to account for sample to sample variability.

We performed a simulation study to compare the performance of tests of variance under the presence of within and between individual dependencies. We compare empirical level and power obtained from 10,000 simulations to assess performance. We considered four variance comparison tests: F-test, Levene's Test, Fligner-Killeen test and permutation F test. We also provided a detailed analysis of a real data set to illustrate our results. Overall, results from our simulation indicate that the performances of the F and the permutation F tests are better than the other methods. However, for all the tests considered, power for detecting small differences in variance is very low when the sample size is small. An interesting finding in our simulation is

that the performance of the tests was actually enhanced by the presence of within individual or group dependence, where the power of the tests increased with increased correlation. On the other hand, between dependence and within and between individual dependence have the effect of lowering the power in comparison with the scenarios of independence and within dependence. It is also observed that the higher the correlation the lower the power. It should be noted that in the case of group comparison the simulation scenario of within and between dependence shows opposite results to those observed in the individual comparison. A higher correlation is associated to both a higher level and power.

On another note, the analysis of the Smith-Lemli-Opitz Syndrome (SLOS) dataset has shown that comparison of pattern profile between individuals is a useful tool to identify influential z-scores of craniofacial measurements that affect CVI and the subsequent results of hypothesis testing of equal variances using the classical F test versus Levene-median and Fligner-Killeen tests.

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Dr. Jemila Hamid for sharing her knowledge and providing programming training, guidance and financial support in the preparation of this manuscript.

A special word of recognition to Dr. Malgorzata Nowaczyk for authorizing the use of a dataset whose valuable information she collected with great personal effort in several countries.

TABLE OF CONTENTS

Descriptive note ii
Abstract iv
Acknowledgments vi
Table of Contents vii
List of Tables viii
List of Figures ix
Chapter I: Introduction 3
Chapter II: Motivating Data 7
Chapter III: Methods 10
3.1 Tests for variance comparison of two individuals 10
3.1.1 F test 10
3.1.2 Permutation F test 12
3.1.3 Levene test 12
3.1.4 Fligner-Killeen test 13
3.2 Tests for average variance comparison of two groups 15
3.2.1 Student's t- test 15
3.2.2 Permutation t- test 16
Chapter IV: Simulation 17
4.1 Description of the simulation 18
4.2 Implementation of the methods 18
4.3 Variance comparison between individuals 19
4.3.1 Scenario of independence 19
4.3.2 Scenario of within dependence 20
4.3.3 Scenario of between dependence 23
4.3.4 Scenario of within and between dependence 25
4.4 Variance comparison between groups 26
4.4.1 Scenario of independence 28
4.4.2 Scenario of within dependence 29
4.4.3 Scenario of within and between dependence 30
Chapter V: Application to Smith-Lemli-Opitz Syndrome 31
Chapter VI: Discussion 62
Bibliography 64
Appendix 67

LIST OF TABLES

Number *Page*

1. Empirical power of the variance tests under independence..... 20
2. Individual variances tests-within dependence..... 21
3. Individual variances tests-between dependence 24
4. Individual variances tests-within and between dependence..... 26
5. Group variance tests under independence 28
6. Group variance tests within dependence..... 29
7. Group variance tests within and between dependence..... 30
8. Characteristics of the study sample 32
9. CVI percentiles for the total sample 33
10. CVI percentiles for samples split by gender 36
11. CVI percentiles for samples split by age range 39
12. CVI percentiles for samples split by severity index..... 41
13. CVI percentiles for samples split by cholesterol index 44
14. CVI percentiles for samples split by severity-cholesterol..... 47
15. Craniofacial measurements used in the study 67
16. Characteristics of the sample study by individual..... 69

LIST OF FIGURES

Number *Page*

1. Power and level of individual variance tests 22
2. Power and within dependency in F test 23
3. Pattern profiles by gender 35
4. CVI distribution by gender 36
5. Pattern profiles by age 37
6. CVI distribution by age 38
7. Pattern profiles by severity 40
8. CVI distribution by severity 41
9. Pattern profiles by cholesterol 43
10. CVI distribution by cholesterol 43
11. Pattern profiles by severity-cholesterol 45
12. Distribution of CVI by severity 46
13. Pattern profiles of two siblings 48
14. Distribution of two relatives 50
15. Pattern profiles of two siblings 51
16. Distribution of two relatives 52
17. Pattern profiles of two siblings 53
18. Distribution of two relatives 54
19. Pattern profiles of two relatives 55
20. Distribution of two relatives 56
21. Pattern profiles of cousins 58
22. Distribution of two relatives 58
23. Pattern profiles of two teens 60
24. Distribution of two teens 61

Chapter 1

INTRODUCTION

Several studies have shown that comparison of individual facial features with statistical norms calculated from reference populations can be used as an important tool in diagnosis of several syndromes including Smith-Magenis Syndrome (SMS), which is associated with a deletion of chromosome 17p11.2 and Smith-Lemli-Opitz Syndrome (SLOS), caused by a deficiency of 7-dehydrocholesterol reductase which is encoded for by the *DHCR7* gene (Allanson et al, 1999). These measurements have also been used to compare the degree of dysmorphogenesis among individuals/groups within patients with specific syndrome or compare different malformation syndromes. To this effect, several studies have attempted to systematically compare a set of linear and angular morphological measurements between two individuals/groups characterized by the presence or absence of certain syndromes and abnormalities (Garn S. 1984, 1985, 1987; Laurencikas E. 2005, 2006).

Based on a conceptual framework of pattern profile comparison, which is oriented to compute indexes of similarity or congruence, researchers use a measure of correlation called Craniofacial similarity index (CSI) to address the question of "syndrome identification" and, an index to determine the "excess of dimensional variability" or pattern deviation in morphological measurements, called Craniofacial

Variability Index (CVI). Both indexes are computed from averaged Z-scores of craniofacial measurements standardized for age and gender using the mean and standard deviation of a reference population. Each of these Z-scores indicates how many standard deviations an individual observation for a certain characteristic is above or below the norms for age and gender in the reference population (Garn, 1984, 1987).

At the individual level CVI is defined as the standard deviation of a set of Z-scores from morphological measurements of an individual. Garn (1984) demonstrated that this index had a relatively normal distribution among clinically normal individuals. Using data from patients with known craniofacial syndromes, they indicated that the CVI in such cases consistently falls outside of the normal range of the index, and its value is highest in individuals with the greatest subjective degree of facial involvement (Garn, 1984). For group comparison, the average CVI and percentiles, calculated from the empirical distribution of CVI, are used. This type of comparisons arises when one is interested in comparing groups of individuals with syndromes and between groups of individuals with different types of syndromes. For the motivating example described in the next section, for instance, we are interested in comparisons between groups with low and high cholesterol level, between low and high severity score as well as difference between age groups and genders. Individual comparison of CVI, for instance, arises when one is interested in comparing CVIs between twins,

other family members or when comparing two unrelated individuals with the same diagnosis.

Application of pattern variability indices are not limited to study of facial measurements. Such measures are also applied in anthropology (anthropometric pattern profile analysis) and in cardio-vascular studies (heart rate variability index). Despite a well-defined measure of craniofacial variability and its wide spread application in biomedical sciences, most of the studies rely on subjective assessment of the distributions without formal statistical significance tests performed.

From statistical perspective, comparison of two CVIs is nothing but comparison of variances for two samples. However, most of distribution assumptions for tests of variances are violated. This is because the facial measurements are taken from the same individuals and hence the within sample independence (the iid assumption) is violated. Moreover, the two samples are often correlated (between sample independence is violated) since the individuals are often family members or are patients with the same diagnosis. For group comparison of CVIs, a subjective assessment of average CVIs as well as other percentiles of the distribution is used to make comparison between two or more groups. In general, however, statistical significance of these differences has not been investigated, although the distribution

of CVIs for healthy reference population has been shown to follow a normal distribution.

The objective of this thesis is to perform a simulation study to investigate and compare: 1) performance of different tests for variance comparison, including non-parametric permutation test, under within and between dependencies similar to those arising in the study of craniofacial pattern profiles, 2) performance of the t-test and permuted t-test in comparing average CVIs under the assumption of dependent measurements. We perform the simulation for different sample sizes to assess performance under small sample size. Different levels of correlation between and within the samples are also considered. The empirical level and power of the tests are used to measure performance. We also use data from patients with Smith-Lemli-Optiz Syndrome (SLOS) to illustrate our results. A brief description of the SLOS data set is provided in the next section.

Chapter 2

MOTIVATING DATA

This thesis is motivated by data collected from patients with a rare genetic disorder known as Smith-Lemli-Opitz Syndrome. Dependent measurements from patients with the condition were obtained in an observational study. A brief description about the data is provided in this section. For more details about the data, we refer the reader to the original study (Nowaczyk et al., 2011).

Smith-Lemli-Opitz Syndrome (SLOS) is an inborn error of cholesterol synthesis with protean manifestations. It is a rare syndrome with an estimated incidence of 1:20,000 – 1:70,000. Although present in other ethnic groups, SLOS appears to be most frequent in Caucasians of northern and central European origin (Nowaczyk et al., 2011). SLOS was described in 1964 as an autosomal recessive condition characterized by multiple malformations, distinctive facial appearance, developmental delay, intellectual disability, and behaviour abnormalities (Nowaczyk et al., 2011). It is caused by a deficiency of 7-dehydrocholesterol reductase which is encoded by the *DHCR7* gene (Nowaczyk et al., 2011). The severe form of SLOS presents in the newborn period with abnormal facial appearance and multiple internal malformations. In contrast, milder forms of SLOS may present later in life

with isolated developmental delay or autism spectrum disorder and may be accompanied by a much more subtle facial dysmorphology.

Demonstration of an elevated precursor molecule, 7-dehydrocholesterol (7DHC), in plasma, or of two mutations in *DHCR7* in subjects with borderline 7DHC levels, is required for the diagnosis (Nowaczyk et al., 2011). Individuals with severe cholesterol deficiency at the time of diagnosis have a higher incidence of congenital malformations, both minor and major, more severe growth retardation, and what is said to be a more unusual facial appearance (Nowaczyk et al., 2011). Individuals with less severe cholesterol deficiency, atypical or near-normal facial features, and minimal physical findings have been reported (Nowaczyk et al., 2011), raising the possibility that mildly affected subjects may be missed. Behaviour abnormalities occur even in those individuals with the fewest physical findings, and include self-injury, aggressiveness, autistic behaviours, severe sleep cycle disturbances, and hyperactivity (Nowaczyk et al., 2011).

Our data comes from an observational study consisting of 51 individuals with biochemically-proven SLOS (Nowaczyk et al., 2011). Of the 51 patients who were recruited in the observational study, 42 of them participated in the anthropometric study. For these 42 patients, 28 standardized cranial and facial measurements (Z scores) were obtained. The 28 dimensions were chosen to represent craniofacial widths, lengths, depths, circumferences as well as ear, eye, nose and mouth

structures. For each measurement, age and sex matched normal standards from a reference population were used to calculate the Z-scores. The reference population consists of measurements of head and face of 2326 healthy North American white children and young adults. Please refer to the original study for more details about the data (Nowaczyk et al., 2011). Other variables that are considered in the study are age, sex, height, weight, physical severity score and plasma cholesterol concentrations.

Chapter 3

METHODS

In this section we describe the methods that are considered in our simulation study. For variance comparison of two related individuals, we will assess the performance of four tests under violation of assumptions of within and between independence in normal distributions. The uniformly most powerful test under normality assumptions, the classical F-test and the permutation F-test will be used in the comparison as well as two of the most robust tests under departures of normality conditions, the Levene and the Fligner-Killeen tests. For average variance comparison (average CVIs) between two groups of individuals with certain characteristics, we will only use the t-test and permutation t-test since the average CVIs display a near normal distribution pattern.

3.1 Tests for variance comparison of two individuals

3.1.1 F-test

Suppose X_1, X_2, …, X_n and Y_1, Y_2, .., Y_m are identically and independently distributed as $N(\mu_1,\sigma^2)$ and $N(\mu_2,\tau^2)$, respectively. The F statistic for comparing variances of the two distributions is given by

(1) $$F = \frac{{S_x}^2}{{S_y}^2}, \quad \text{where, } {S_x}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ and } {S_y}^2$$

$$= \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$$

Under the null hypothesis, the test statistic is distributed as an F distribution with *n*-1 and *m-1* degrees of freedom. The F distribution is asymmetric and has two degrees of freedom, one for the numerator and one for the denominator. Each combination of degrees of freedom gives a different F distribution.

The null hypothesis of equal variances is rejected if F > F (α/2,n-1,m-1), where α is the significance level.

Different studies have shown that the F test performs the most powerful test with low probability of type I error- under the normality assumption, under unequal sample sizes as well as for distributions with different means. A generalization of the F test for more than two samples was proposed by Bartlett (1937).

Box (1953) showed that the probability distribution of the Bartlett´s statistic is dependent on the kurtosis of the variable´s distribution, which suggests sensitivity to departure from the assumption of normality.

This limitation prompted the development of alternative methods, including methods that use correction factors to the degrees of freedom of the F distribution of ratio of variances to reduce sensitivity to kurtosis (Shoemaker 2003).

3.1.2 Permutation F-test

In the permutation test, the observed sample values of X and Y are considered as a group of *n*+*m* measurements that can be permuted in $\binom{n+m}{n}$ ways. In each of these permuted samples, the F-statistic given in equation (1) is computed and compared with the observed F-statistic in the original sample. The p-value is then computed as the number of times the permuted F-statistic is greater than the observed F-statistic, divided by the total number of permuted samples.

3.1.3 Levene Test

This test was originally developed based on the Analysis of variance (ANOVA) of transformed data (deviations from the mean), which resembles a test of equal variances while using an ANOVA F test of equal means to the centered data (Levene, 1960).

In this thesis we use the median instead of the mean to transform the data because of the strong empirical evidence indicating that the median-centered Levene test is robust to the normality assumption (Marozzi, 2011).

Suppose we have two groups with observations in each group identically and independently distributed as the normal distribution. Let $Z_{ij}=|X_{ij}$ - $\overline{X}_i|$ where *i*=1, 2 and *j*= 1, 2,…n_i , then Levene's test for testing homogeneity of variances is given by the following ANOVA-type statistic.

(2) $$W_o = \frac{n_1 n_2 \left(\widetilde{Z_1} - \widetilde{Z_2}\right)^2 / ((n_1 + n_2)}{(\sum_{j=1}^{n_1}(Z_{1j} - \widetilde{Z_1})^2 + \sum_{j=1}^{n_2}(Z_{2j} - \widetilde{Z_2})^2)/(n_1 + n_2 - 2)}$$

Where n_1 and n_2 are the sample sizes of group 1 and 2; Z_{1j} and Z_{2j} represent the individuals in each group and, $\widetilde{Z_1}$ and $\widetilde{Z_2}$ correspond to the median of the first and second group, respectively. Under the null hypothesis, W_o is distributed as an F distribution with 1 and $n_1 + n_2 - 2$ degrees of freedom.

3.1.4 Fligner-Killeen

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, .., Y_m$ be two independent random samples of size *n* and *m* respectively, from populations identically and independently distributed with unspecified continuous distributions which have unknown but equal medians. The Fligner-Killeen non-parametric test for testing equality of variances converts each X_i and Y_j to its absolute deviation from the respective sample median using U= $|X_i - \hat{X}_i|$ and V=$|Y_j - \hat{Y}_j|$. It then computes the ranks of the combined U and V vectors and assigns them increasing z-scores using the following standard normal distribution function.

(3) $$a_{N,i} = \varphi^{-1}[\tfrac{1}{2} + \frac{i}{2(N+1)}] \quad , \quad \text{for rank i and N=}m+n$$

Where i corresponds to the ranks and φ^{-1}denotes the quantile function or inverse of the cumulative standard normal distribution function. This function provides the values $a_{N,i}$ -called scores- for which the probability is less or equal than $[\frac{1}{2} + \frac{i}{2(N+1)}]$.

The test is based on the statistic

$$T^2 = \frac{(m)(n)(\overline{A_1} - \overline{A_2})^2}{(m+n)V^2} \tag{4}$$

Where $\overline{A_1}$ and $\overline{A_2}$ are the means of $a_{N,i}$ values for the two samples and,

$$V^2 = \sum_{i=1}^{N} \frac{(a_{N,i} - \bar{a})^2}{(N-1)} \quad , \quad \bar{a} = \frac{m\overline{A_1} + n\overline{A_2}}{m+n}$$

Under the null hypothesis and for large sample size, the Fligner-Killeen statistic is approximately distributed as chi-square with one degree of freedom. Several simulation studies show good robustness properties of this method under conditions of non-normality.

3.2 Tests for variance comparison of two groups

When studying pattern analysis of individuals with craniofacial syndromes, comparison of the distribution of craniofacial variability index (CVI) is often compared between individuals in certain groups. In our motivating data, for instance, we are interested in comparing CVI between groups of patients with high and low cholesterol levels, between patients with high and low severity score, etc. Investigators often calculate the average, the 5th and 95th percentile of the distribution of CVI and make subjective comparison without performing statistical significance of the differences. In this study we will empirically assess the performance of two tests (t test and permutation t test) in comparing average CVIs between two groups.

3.2.1 Student's t test

Let X_1, X_2, X_n and Y_1, Y_2, Y_m be two independent random samples of size *n* and *m* respectively, from populations identically and independently normally distributed. The t statistic for comparing means of the two distributions under unequal sample sizes and equal variances is given by

$$t = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{(n-1){S_x}^2 + (m-1){S_y}^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \tag{5}$$

Where $\bar{X}$ and $\bar{Y}$ are the sample means of the two samples X and Y and, N= *m*+*n*.

Where ${S_x}^2$ is the sample variance of X and ${S_y}^2$ is the sample variance of Y both computed with the following formulas.

$$S_x{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S_y{}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$$

Under the null hypothesis, the test statistic is distributed as a t distribution with *n+m-2* degrees of freedom. P-values are computed using the t probability distribution function.

3.2.2 Permutation t-test

In the permutation test, the observed sample values of X and Y are considered a group of *n+m* measurements that can be permuted in $\binom{n+m}{n}$ ways. In each of these permuted samples, the t-statistic is computed and compared with the observed t-statistic in original sample. Then, the p-value is computed as the number of times the permuted t-statistic is greater than the observed t-statistic, divided by the total number of permuted samples.

Chapter 4

SIMULATION

We performed a simulation study to compare the performance of tests of equal variance described in the previous section, under within and between sample dependency and for various sample sizes. We would like to highlight that these assumptions are based on the characteristics of the motivating data presented in section 2, where several facial measurements were collected from individuals (some of them siblings and cousins) with Smith-Lemli-Optiz Syndrome. In the context of the motivating example of this study, there are three types of variance (Craniofacial variability index (CVI), as it is called in the medical literature) comparisons that are of interest to researchers and health practitioners: a) variance comparison between two related individuals diagnosed with the syndrome and, b) variance comparison between two unrelated individuals and c) average variance comparison between two groups with the syndrome.

For comparison of individual variances, we considered four tests, the F test, Permutation F test, Levene test (median-centered) and Fligner-Killeen test. For the group variance comparison of two tests are considered, the t-test and the permutation t test.

4.1 Description of the Simulation

The simulated data is generated from a multivariate normal distribution with various sample sizes (n=15, 25, 35, 45, 55). Four different levels of correlation were considered, r=0 for the case of independence, and r=0.4, r=0.6 and r=0.9 for the case of dependence. We also considered four values for the ratio of the variances to be compared (rvar=1, rvar=1.5, rvar=2 and rvar=3). Based on these parameters and the nature of the motivating data considered in this study, the following four scenarios are investigated using 10,000 simulations : i) independent data, where measurements from the two samples to be considered are generated from two independent normal distributions; ii) within dependence, where the two samples are assumed to be independent, however, measurements within each sample are correlated; iii) between dependence, where the two samples are dependent, however, the measurements within each sample are assumed to be independent, and iv) within and between dependence. We use a multivariate normal distribution with a partitioned covariance matrix to generate data for these scenarios.

4.2 Implementation of the methods

Performance comparison of the four variance tests included in the study (F, Permutation, Levene and Fligner-Killeen) were implemented in all scenarios under the two-sided alternative with a level of significance equal to 0.05 (α= 0.05). For instance, in the F test, null hypothesis was expressed as variance ratio equal to a set value such as

1, 1.5,2 or 3 and, alternative hypothesis as variance ratio different to the set value. We used empirical power and level obtained from 10, 000 simulations to robustness of each of the tests with respect to within and between independence assumptions as well as to compare performance among the different tests under the different scenarios considered.

4.3 Variance comparison between individuals

Overall, the simulation results indicate that the F and permutation F tests perform better than the other two approaches (Tables 1-5). The power for the permutation F test is in general higher than that of the F test. However, the level of the permutation test is quite high under highly correlated data in the scenario of within dependence (Table 2).

4.3.1 Scenario of independence

Our simulation results show that the power to detect small differences (variance ratio=1.5, 2) is quite low (Table 1), especially for small sample sizes, indicating that these tests are not powerful enough to detect true but differences in studies such as the one presented in this thesis where measurements sample size ranges between 15 and 30. The F test and permutation F test performed relatively better than the other two approaches (Table 1). However, the performances for all of the tests are comparable as the sample size gets larger.

Table 1. *Empirical power of the tests for individual variance under independence, where rvar=ratio of variance and the power for rvar=1, is the level of the test.*

Methods	n	rvar=1	rvar=1.5	rvar=2	rvar=3
	15	0.05	0.11	0.24	0.52
	25	0.07	0.18	0.39	0.77
F test	35	0.05	0.23	0.49	0.87
	45	0.06	0.26	0.60	0.94
	55	0.04	0.33	0.75	0.98
	15	0.03	0.07	0.15	0.35
Levene test	25	0.04	0.11	0.28	0.62
median	35	0.04	0.17	0.42	0.80
	45	0.04	0.20	0.51	0.89
	55	0.04	0.27	0.65	0.97
	15	0.03	0.06	0.14	0.31
	25	0.04	0.11	0.26	0.60
Fligner	35	0.04	0.16	0.39	0.77
	45	0.04	0.18	0.48	0.87
	55	0.04	0.27	0.62	0.95
	15	0.11	0.27	0.41	0.64
Permutation	25	0.03	0.14	0.34	0.67
F test	35	0.02	0.19	0.45	0.82
	45	0.07	0.44	0.76	0.97
	55	0.05	0.34	0.70	0.97

4.3.2 Scenario of within dependence

In this scenario, the two samples are considered to be independent. However, the measurements within each sample are assumed to be dependent with three different levels of correlation (r=0.4, r=0.6 and r=0.9). As is the case in the independent scenario, the permutation test performs better than the other three approaches, but with a notorious instability in level when the correlation is 0.9 (Table 2, Figure 1). The performance of the F test is quite good in general.

Table 2. Individual variance tests-within dependence

		rvar=1			rvar=1.5			rvar=2			rvar=3		
Tests	n	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9
	15	0.06	0.06	0.06	0.15	0.23	0.70	0.34	0.48	0.95	0.68	0.81	1.00
	25	0.05	0.05	0.05	0.24	0.34	0.91	0.55	0.72	1.00	0.90	0.97	1.00
F	35	0.04	0.04	0.04	0.32	0.46	0.97	0.71	0.86	1.00	0.97	0.99	1.00
	45	0.05	0.05	0.05	0.40	0.58	1.00	0.82	0.94	1.00	0.99	1.00	1.00
	55	0.05	0.05	0.05	0.48	0.66	1.00	0.89	0.97	1.00	1.00	1.00	1.00
	15	0.04	0.03	0.03	0.09	0.14	0.52	0.23	0.32	0.82	0.49	0.63	0.97
Lev	25	0.03	0.03	0.03	0.18	0.26	0.82	0.43	0.59	0.98	0.80	0.91	1.00
	35	0.04	0.04	0.05	0.26	0.38	0.94	0.60	0.78	1.00	0.93	0.98	1.00
	45	0.04	0.04	0.04	0.33	0.49	0.99	0.73	0.88	1.00	0.98	1.00	1.00
	55	0.04	0.04	0.05	0.40	0.58	0.99	0.82	0.94	1.00	0.99	1.00	1.00
	15	0.03	0.03	0.04	0.08	0.14	0.48	0.20	0.28	0.76	0.45	0.57	0.94
	25	0.03	0.03	0.03	0.16	0.24	0.78	0.40	0.55	0.97	0.75	0.88	1.00
Flig	35	0.04	0.04	0.05	0.24	0.36	0.91	0.57	0.74	1.00	0.90	0.97	1.00
	45	0.04	0.05	0.05	0.31	0.46	0.98	0.70	0.86	1.00	0.96	0.99	1.00
	55	0.04	0.04	0.04	0.39	0.56	0.99	0.80	0.93	1.00	0.99	1.00	1.00
	15	0.06	0.06	0.23	0.34	0.44	0.89	0.67	0.62	0.98	0.87	0.97	1.00
Perm	25	0.06	0.07	0.17	0.37	0.61	0.98	0.59	0.83	1.00	0.83	0.97	1.00
	35	0.06	0.10	0.13	0.60	0.51	0.97	0.80	0.92	1.00	0.96	0.99	1.00
	45	0.04	0.04	0.07	0.56	0.65	1.00	0.91	0.97	1.00	1.00	1.00	1.00
	55	0.05	0.06	0.08	0.68	0.78	1.00	0.93	0.97	1.00	1.00	1.00	1.00

Lev=Levene median Flig=Fligner Perm=Permutation r=correlation

An interesting finding from our simulation is that the power of the tests increases with correlation (Table 2). For instance for the permutation test, the power to detect variance ratio of 1.5 (rvar=1.5) using a sample size of 15 increased from 0.11 (independence) to 0.34 (r=0.4). The power increased to 0.44 for a correlation of 0.6, and for r=0.9, the power to detect a ratio of 1.5 is 0.89 (Table 2).

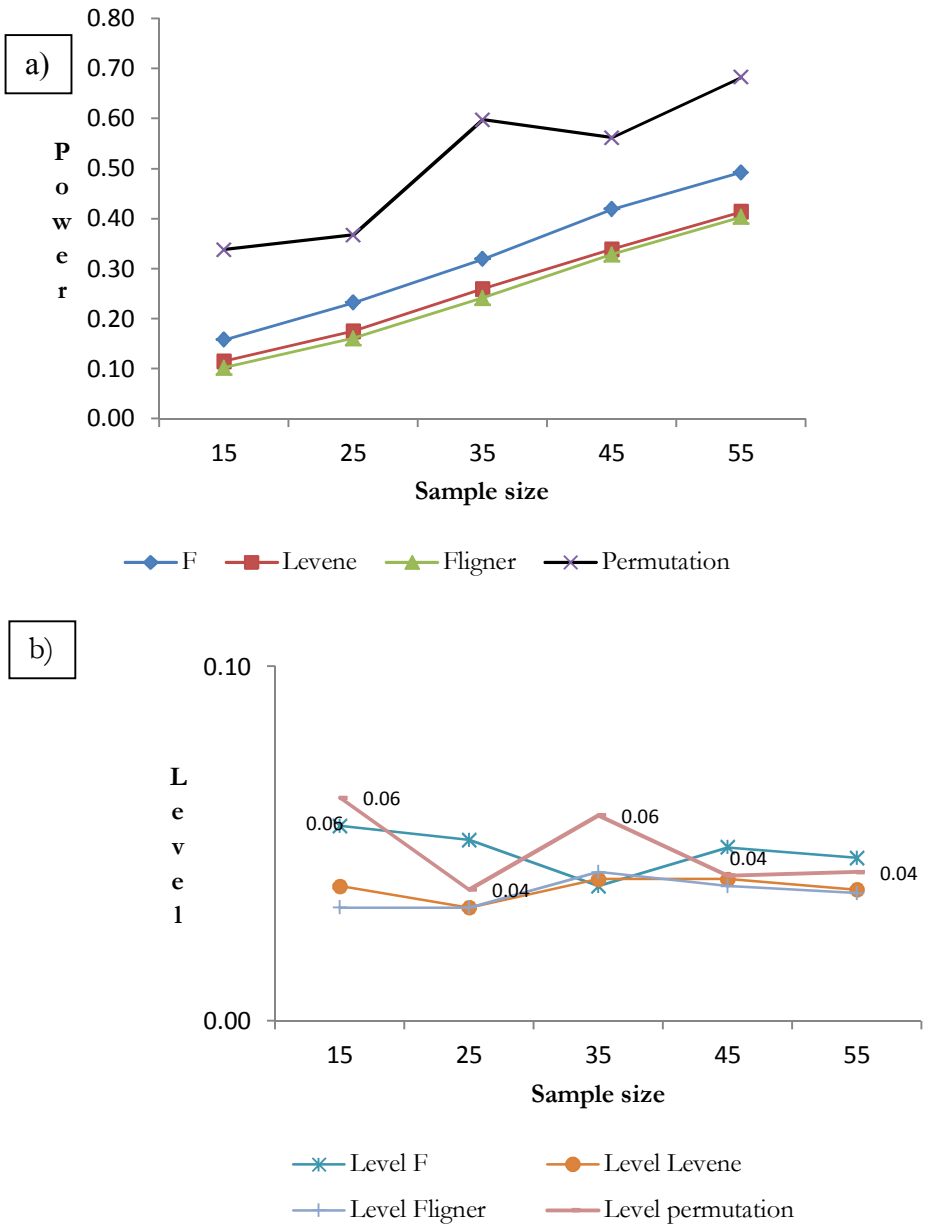

***Figure 1**. Performance of the tests to detect a variance ratio =1.5 where the two samples are assumed to be independent, but within correlation of 0.4 a) empirical power of tests b)level of the tests*

Figure 2 shows the performance of the F test for different sample sizes and for various levels of correlation. It can be seen from the figure that, under low and moderate correlation, increases in sample size makes a great difference in the power size. However, for high correlation these differences in power are considerably small for sample sizes of 25 and over.

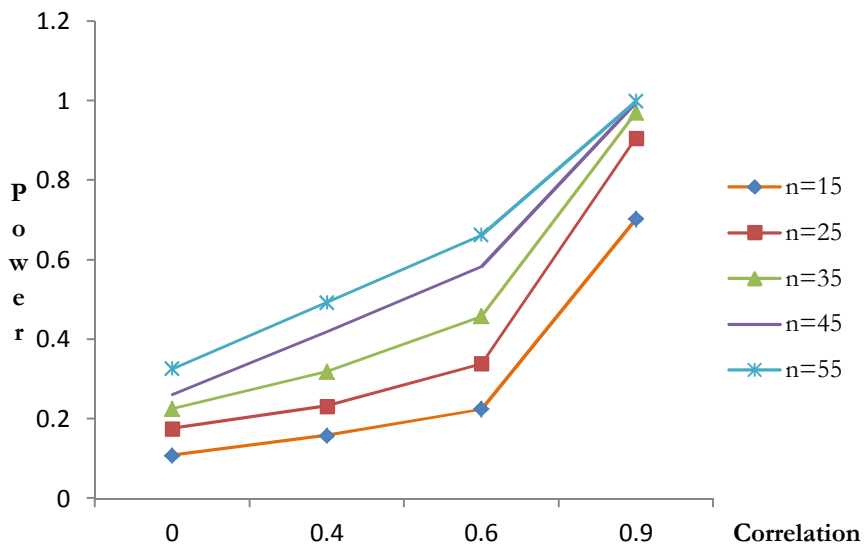

Figure 2: *Power and within dependence in F test*

4.3.3 Scenario of between dependence

For this scenario, we generated two dependent samples, where measurements within the two samples are generated to be independent. Our simulation results show that the performance of the tests is affected by between dependence where we obtained a much lower empirical power than the independent scenario (Table 3). This indicates

that the performance of the tests is not robust to the between independence assumption. Moreover, unlike the scenario of within dependence, there is a decrease in power as the correlation increases from 0.4 to 0.6 and 0.9.

As in the previous scenarios, the Permutation F test is clearly the most powerful test although it shows some instability (Table 3).

Table 3. Individual variance tests-between dependence

		rvar=1			rvar=1.5			rvar=2			rvar=3		
Tests	n	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9
	15	0.04	0.05	0.04	0.08	0.06	0.07	0.18	0.15	0.12	0.38	0.32	0.26
	25	0.04	0.04	0.04	0.12	0.10	0.10	0.28	0.24	0.22	0.60	0.52	0.42
F	35	0.03	0.05	0.04	0.15	0.14	0.11	0.37	0.34	0.28	0.74	0.68	0.60
	45	0.05	0.05	0.04	0.18	0.18	0.14	0.46	0.41	0.34	0.87	0.80	0.70
	55	0.06	0.04	0.04	0.24	0.18	0.16	0.55	0.48	0.44	0.91	0.88	0.79
	15	0.03	0.03	0.03	0.05	0.04	0.05	0.12	0.10	0.07	0.27	0.19	0.15
Lev	25	0.03	0.03	0.03	0.11	0.09	0.08	0.22	0.17	0.16	0.48	0.39	0.33
	35	0.04	0.04	0.04	0.12	0.11	0.08	0.30	0.28	0.22	0.65	0.55	0.50
	45	0.04	0.04	0.04	0.15	0.15	0.12	0.38	0.32	0.27	0.77	0.72	0.60
	55	0.05	0.04	0.04	0.20	0.15	0.12	0.46	0.42	0.36	0.86	0.81	0.70
	15	0.03	0.03	0.03	0.05	0.04	0.04	0.11	0.09	0.07	0.24	0.17	0.15
	25	0.03	0.03	0.03	0.10	0.08	0.08	0.19	0.17	0.16	0.44	0.36	0.32
Flig	35	0.04	0.04	0.03	0.11	0.11	0.08	0.27	0.27	0.20	0.62	0.52	0.45
	45	0.04	0.04	0.04	0.15	0.14	0.11	0.37	0.30	0.25	0.75	0.69	0.56
	55	0.04	0.04	0.04	0.19	0.15	0.11	0.43	0.40	0.35	0.82	0.79	0.67
	15	0.02	0.02	0.05	0.16	0.10	0.18	0.38	0.19	0.16	0.59	0.52	0.20
Perm	25	0.06	0.03	0.06	0.21	0.21	0.10	0.41	0.50	0.31	0.57	0.61	0.61
	35	0.08	0.02	0.07	0.35	0.19	0.26	0.50	0.32	0.23	0.78	0.68	0.72
	45	0.08	0.06	0.04	0.40	0.30	0.17	0.55	0.49	0.52	0.78	0.74	0.74
	55	0.06	0.03	0.04	0.44	0.31	0.22	0.63	0.64	0.60	0.94	0.90	0.90

Lev=Levene median Flig=Fligner Perm=Permutation r=correlation

4.3.4 Scenario of within and between dependence

In this scenario, we let both the within and between individual assumptions to be violated where we generated two normal samples that are dependent. We also generated the data such that the measurements within each of the samples are dependent. This scenario is actually the most common violation of assumptions in practice, including for our motivated data, where several measurements were taken from individuals with the same syndrome – some of the individuals in our data are also siblings and relatives.

The combined effect of within and between dependence (Table 4), in all tests, greatly reduced the level and power of the tests as the correlation increases. This finding indicates that when both the within and between assumptions are violated, which is the case in our real data, the power is very low. Researchers should, therefore, be cautions when making comparisons.

The results of our simulation for this scenario clearly show a pattern of lower power with higher correlation (Table 4). In general, the permutation test provided the highest power. The next powerful test, for this scenario, is the F test.

Table 4. Individual variance tests-within and between dependence

		rvar=1			rvar=1.5			rvar=2			rvar=3		
Tests	n	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9	r=.4	r= .6	r= .9
F	15	0.05	0.04	0.00	0.07	0.05	0.01	0.15	0.13	0.05	0.30	0.36	0.17
	25	0.04	0.03	0.00	0.09	0.09	0.03	0.19	0.24	0.08	0.51	0.57	0.32
	35	0.04	0.03	0.00	0.12	0.10	0.04	0.28	0.32	0.15	0.66	0.73	0.46
	45	0.05	0.03	0.00	0.13	0.14	0.07	0.36	0.43	0.22	0.79	0.86	0.59
	55	0.04	0.04	0.00	0.15	0.19	0.07	0.43	0.51	0.27	0.87	0.93	0.69
Lev	15	0.03	0.03	0.00	0.05	0.04	0.01	0.11	0.09	0.03	0.22	0.24	0.11
	25	0.04	0.03	0.00	0.06	0.07	0.02	0.15	0.19	0.07	0.39	0.46	0.20
	35	0.04	0.02	0.00	0.09	0.09	0.03	0.22	0.25	0.11	0.56	0.63	0.36
	45	0.04	0.03	0.00	0.10	0.11	0.05	0.30	0.36	0.17	0.68	0.75	0.49
	55	0.05	0.03	0.00	0.13	0.16	0.06	0.37	0.43	0.22	0.79	0.85	0.57
Flig	15	0.03	0.03	0.00	0.05	0.04	0.01	0.11	0.09	0.03	0.20	0.23	0.11
	25	0.03	0.03	0.01	0.06	0.07	0.02	0.14	0.17	0.07	0.37	0.42	0.19
	35	0.04	0.02	0.00	0.08	0.08	0.03	0.21	0.25	0.11	0.53	0.60	0.34
	45	0.04	0.03	0.00	0.10	0.11	0.06	0.28	0.33	0.16	0.65	0.73	0.45
	55	0.05	0.03	0.00	0.12	0.16	0.06	0.35	0.41	0.21	0.76	0.83	0.53
Perm	15	0.06	0.03	0.01	0.17	0.09	0.08	0.31	0.26	0.07	0.57	0.38	0.21
	25	0.03	0.04	0.01	0.25	0.18	0.10	0.35	0.47	0.15	0.46	0.74	0.54
	35	0.07	0.04	0.01	0.13	0.17	0.20	0.33	0.22	0.20	0.82	0.61	0.49
	45	0.05	0.02	0.00	0.11	0.21	0.18	0.41	0.47	0.56	0.75	0.81	0.60
	55	0.05	0.04	0.00	0.28	0.26	0.34	0.41	0.70	0.54	0.95	0.91	0.84

Lev=Levene median Flig=Fligner Perm=Permutation r=correlation

4.4 Average variance comparison between groups

At the individual level Craniofacial Variability Index (CVI) has been defined as the standard deviation of a set of Z-scored morphological measurements of an individual. Each of these Z-scores indicates how many standard deviations an individual observation for a certain characteristic is above or below the norms for age and gender given by the reference population mean or median.

At the group level CVI is computed as the average of CVIs from individuals who belong to the same group. For instance, for a matrix **A** of *n* individuals and *q* dimensions, belonging to the same group

$$\mathbf{A}=\begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nq} \end{bmatrix}$$

The algebraic formula for CVI can be expressed as the mean of *n* individual standard deviations.

$$CVI = \frac{\sum_{i=1}^{n}({S_{xi}}^2)^{\frac{1}{2}}}{n} \tag{7}$$

Where ${S_{xi}}^2 = \frac{1}{q-1}\sum_{j=1}^{q}(x_{ij} - \bar{x}_i)^2$ is the sample variance of *q* measurements for individual *i* and, $\bar{x}_i = \frac{\sum_{j=1}^{q} x_{ij}}{q}$ corresponds to the mean of *q* measurements for individual *i*.

Suppose we have two such matrices **A** and **B**. From matrix **A,** we compute a vector of individual variances $V_A = \begin{bmatrix} {S_{x1}}^2 \\ .. \\ {S_{xn}}^2 \end{bmatrix}$ whose mean will be compared with other sample of variance vector V_B from a Matrix B, using the t-test and permutation t-test.

Since our goal is to assess the performance of these tests under scenarios of dependent data similar to the motivating data, it is relevant to establish if the conditions of within and between dependence in the multivariate datasets, **A** and **B ,** also generate similar effects on the performance of the tests for comparing averages of the two variance vectors, V_A and V_B.

4.4.1 Scenario of independence

In the scenario of independence both tests, the t-test and the permutation t-test, exhibit low level (equal to the nominal level). Our results also show that the two tests are very powerful in detecting small differences where the power for detecting a difference greater than 0.15 was close to one for very small sample sizes, consequently, we considered differences smaller than 0.15 to make comparisons between the two methods. Table 5 shows that the permutation test performs better than the t-test.

Table 5. *Group variance tests under independence*

Methods	n	dvar=0	dvar=0.1	dvar=0.15
t test	15	0.04	0.10	0.18
	25	0.06	0.22	0.39
	35	0.05	0.38	0.67
	45	0.05	0.57	0.87
	55	0.05	0.73	0.96
Permutation t test	15	0.05	0.18	0.35
	25	0.06	0.30	0.69
	35	0.05	0.50	0.92
	45	0.06	0.67	0.99
	55	0.05	0.80	1.00

n= sample size dvar=difference between average variances

4.4.2 Scenario of within dependence

For both tests level is insensitive to violation of within independence assumption. The level remains close to the nominal level (α = 0.05) for all the correlation values considered (Table 6). However, the power of the tests increased considerably compared to the independence scenario. The power of the two tests also increased with an increased within dependence where the power of the F-test for detecting a difference of 0.1 between the two average variances, with sample size of 25, increased from 0.22 (independence) to 0.45 (r=0.4), for r=0.6, the power increased to 0.75 (Table 6). Overall, the permutation test performs better than the t-test. However, the difference is negligible for large samples and higher correlation values.

Table 6. *Empirical power and level of the t and permutation t tests for comparing average variances between two independent samples at different levels of within dependence*

		dvar=0			dvar=0.1			dvar=0.15		
Methods	n	r=.4	r=.6	r=.9	r=.4	r=.6	r=.9	r=.4	r=.6	r=.9
t test	15	0.05	0.05	0.05	0.20	0.36	1.00	0.36	0.59	1.00
	25	0.05	0.05	0.05	0.45	0.75	1.00	0.75	0.97	1.00
	35	0.05	0.05	0.05	0.76	0.97	1.00	0.97	1.00	1.00
	45	0.05	0.05	0.05	0.93	1.00	1.00	1.00	1.00	1.00
	55	0.05	0.05	0.05	0.99	1.00	1.00	1.00	1.00	1.00
Permutation t test	15	0.05	0.05	0.06	0.31	0.48	1.00	0.48	0.72	1.00
	25	0.04	0.06	0.05	0.57	0.85	1.00	0.85	0.99	1.00
	35	0.06	0.05	0.05	0.85	0.99	1.00	0.99	1.00	1.00
	45	0.05	0.04	0.06	0.97	1.00	1.00	1.00	1.00	1.00
	55	0.06	0.04	0.06	1.00	1.00	1.00	1.00	1.00	1.00

n=sample size dvar= difference between average variances r=correlation

4.4.3 Scenario of within and between dependence

The results of the simulation under this scenario presented in Table 7 clearly show that the power increases as the correlation gets higher. However, at the same time it is observed that the higher the correlation the higher the level. Besides the level increases as the sample size gets larger.

***Table 7**. Empirical power and level of the tests for comparing average variance under violation of within and between independence assumptions.*

		dvar=0			dvar=0.1			dvar=0.15		
Methods	n	r=.4	r=.6	r=.9	r=.4	r=.6	r=.9	r=.4	r=.6	r=.9
	15	0.15	0.31	0.70	0.14	0.28	0.64	0.13	0.26	0.60
	25	0.22	0.43	0.78	0.20	0.42	0.71	0.20	0.38	0.71
t test	35	0.28	0.50	0.82	0.25	0.48	0.76	0.24	0.47	0.75
	45	0.29	0.52	0.83	0.29	0.50	0.79	0.25	0.49	0.77
	55	0.35	0.60	0.84	0.32	0.57	0.82	0.32	0.55	0.80
	15	0.14	0.22	0.38	0.11	0.19	0.34	0.11	0.20	0.34
Permutation	25	0.15	0.24	0.39	0.14	0.25	0.37	0.13	0.23	0.37
t test	35	0.18	0.23	0.38	0.16	0.26	0.37	0.17	0.26	0.38
	45	0.17	0.28	0.40	0.16	0.27	0.37	0.15	0.27	0.36
	55	0.20	0.31	0.43	0.20	0.29	0.41	0.19	0.27	0.41

n=sample size dvar= difference between the average variances r=correlation

Chapter 5

APPLICATION TO SMITH-LEMLI-OPITZ SYNDROME

In this section we will describe more specific characteristics of the motivating data in terms of its multivariate structure and probability distribution. Examining these conditions in light of the simulation results will provide useful insights into the process of testing equality of individual and group variances.

As mentioned in chapter 2, the study sample comprises 42 individuals diagnosed with SLOS on the basis of plasma 7DHC level[1]. Data on 28 craniofacial measurements were collected for each individual and transformed to Z-scores using the mean and standard deviation of a reference population for comparable age and gender. A description of these 28 variables is provided in Table 15, in the Appendix.

The dataset also include information on gender, age, physical severity score and cholesterol deficiency at the time of diagnosis. These two latter variables were dichotomized in low and high levels using cut-off values determined by clinicians

The distribution of the study sample by gender, age range, severity and cholesterol (Table 8) shows that the overall majority of participants (60%) belongs either to the category of low severity or high cholesterol. This percentage is even considerably

[1] Demonstration of an elevated precursor molecule, 7-dehydrocholesterol (7DHC), in plasma, or of two mutations in DHCR7 in subjects with borderline 7DHC levels, is required for the diagnosis (Nowackzyk, 2011).

higher in the age ranges of 12 to 16 and 16 to 25. Only in the age range of 0 to 5 it is observed just the opposite for the severity scale, where 77% is in the high severity group.

Table 8. Characteristics of the study sample

		Severity index		**Cholesterol index**		**Severity-Cholesterol**			
	n (%)	Low	High	Low	High	L-L	L-H	H-L	H-H
Total	42 (100%)	25 (60%)	17 (40%)	17 (40%)	25 (60%)	5 (12%)	20 (48%)	12(29%)	5 (12%)
Gender									
Female	17 (100%)	11 (65%)	6 (35%)	9 (53%)	8 (47%)	4(24%)	7(41%)	5(29%)	1(6%)
Male	25 (100%)	14 (56%)	11(44%)	8 (32%)	17 (68%)	1(4%)	13(52%)	7(28%)	4(16%)
Age									
0-5	13 (100%)	3 (23%)	10 (77%)	6 (46%)	7 (54%)	0(0%)	3 (23%)	6(46%)	4(31%)
6 to 12	8 (100%)	5(63%)	3(37%)	4 (50%)	4 (50%)	1(13%)	4(50%)	3(38%)	0(0%)
12 to 16	13 (100%)	10(77%)	3(23%)	4(31%)	9(69%)	2(15%)	8(62%)	2(15%)	1(8%)
16 to 25	8(100%)	7(88%)	1(12%)	3(38%)	5(62%)	2(25%)	5(63%)	1(13%)	0 (0%)

L-L: Low severity and low cholesterol L-H:Low severity and high cholesterol
H-L: High severity and low cholesterol H-H:High severity and high cholesterol

Based on the finding that high severity and low cholesterol are correlated (Nowaczyk, 2011), we have included four categories that result from the combination of severity and cholesterol scales: low severity-low cholesterol (L-L), low severity-high cholesterol (L-H), high severity-low cholesterol (H-L) and high severity-high cholesterol (H-H). It can be observed that 48% of the sample is in the group mildly affected by the syndrome according to the two criteria (L-H) while 29% is in the opposite group (H-L). The comparison between these two groups is relevant in the context of identifying the craniofacial variability in the group that presents near-normal facial features. Table

9 shows that craniofacial variability is greater in the group diagnosed with SLOS versus the normative reference.

Table 9. *CVI percentiles for SLOS patients versus healthy sample*

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Smith Syndrome	42	0.61	1.30	1.62	2.08	2.72	2.79
Normative reference(*)	1248	na	0.52	0.62	0.81	1.04	1.18

(*) Garn (1985)

Another objective of the data analysis is to apply hypothesis testing of equal variances in the context of craniofacial variability index (CVI) comparison between individuals. It should be noted that the literature reports examples of comparison between individuals with CVI equal to and greater than 1.2 (the 95th CVI percentile of the normative reference in Table 9), noting that in these cases a CVI greater than 1.2 corresponds to an observable dysmorphogenesis (Garn, 1985). However, since p-values have not been reported for individual variance comparison, it is relevant to fill that gap in the literature.

5.1 Comparison of two group variance vectors

Important groups to be compared are: gender, cholesterol deficit (low vs. high), severity (low vs. high), and age (< 12 (below puberty) and >12 (above puberty)) and the combination of severity and cholesterol levels. For each group we w analyzed the pattern profile between groups in order to obtain the correlation and visually identify those craniofacial measurements in which average z-scores might differ.

We will verify if the variance vectors of the corresponding groups are normally distributed and have unequal sample sizes and equal variances. These conditions are relevant when analyzing the level and power of the test to be used for testing equality of means of the variance vectors.

CVI by Gender

The pattern profiles by gender exhibit a high correlation of 0.95. Figure 6 clearly shows a very close match between average z-scores for most of the craniofacial measurements. There are a few variables where the z-scores differ and the greatest difference is observed in the head length (g-op). Taking the normative reference for age and gender, on average women with SLOS seem to have a greater deviation in head length than men diagnosed with the same syndrome.

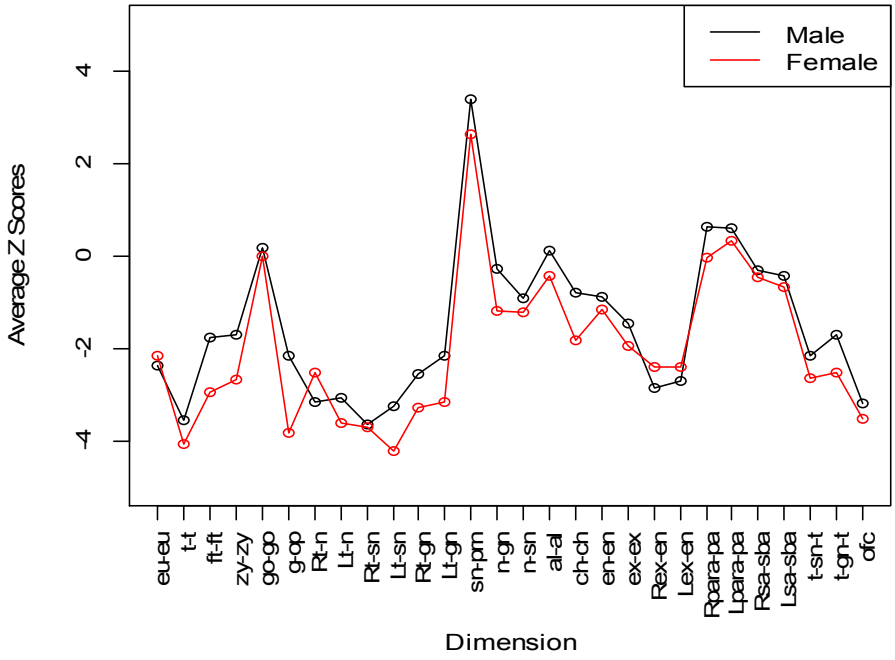

Figure 3*.Pattern profiles by gender (r=.95, p <.001)*

CVI distributions for males and females have the characteristics of a normal distribution (similar skewness of 0.5 and kurtosis of -0.5), as it can observed in Figure 3 and verified by Shapiro-Wilk tests (p-values of 0.98 and 0.77 for males and females).

To compare the CVI means for males and females we use the Student t test for unequal sample sizes (25 and 17) and equal variances (F = 0.7143, p-value= 0.4443).

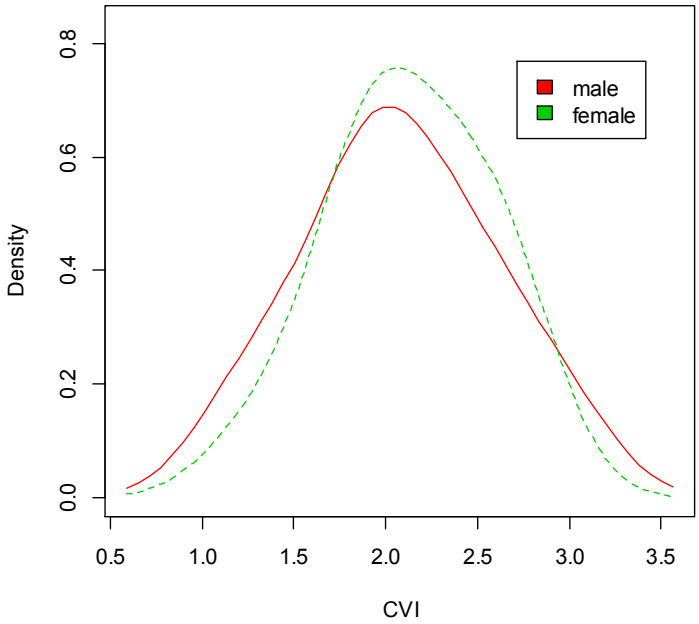

Figure 4: *CVI distribution by gender (p=.71 t-test)*

The results of the t-test give a p-value of 0.71 indicating that there is no evidence to reject the hypothesis of equal means between groups of male and female. The CVI medians for males and females are very close, 2.07 and 2.09 respectively (Table 10).

Table 10. CVI percentiles for samples split by gender

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Male	25	0.59	1.31	1.55	2.07	2.73	3.07
Female	17	0.61	1.48	1.80	2.09	2.63	2.78

CVI by Age (below vs. above poverty)

The pattern profiles by the age range of under and over 12 years old exhibit a high correlation of 0.89. Figure 5 shows a very close match between average z-scores for most of the craniofacial measurements. There are a few variables where the z-scores differ and the greater differences are observed in the forehead width (ft-ft) and in the right side of the palpebral fissure length (Rex-en). Taking the normative reference for age and gender, on average individuals age 12 and older with SLOS seem to have a greater deviation in forehead width than their younger ones diagnosed with the same syndrome. In the variable Rex-en the deviations are greater in the group younger than 12 with respect to the older ones.

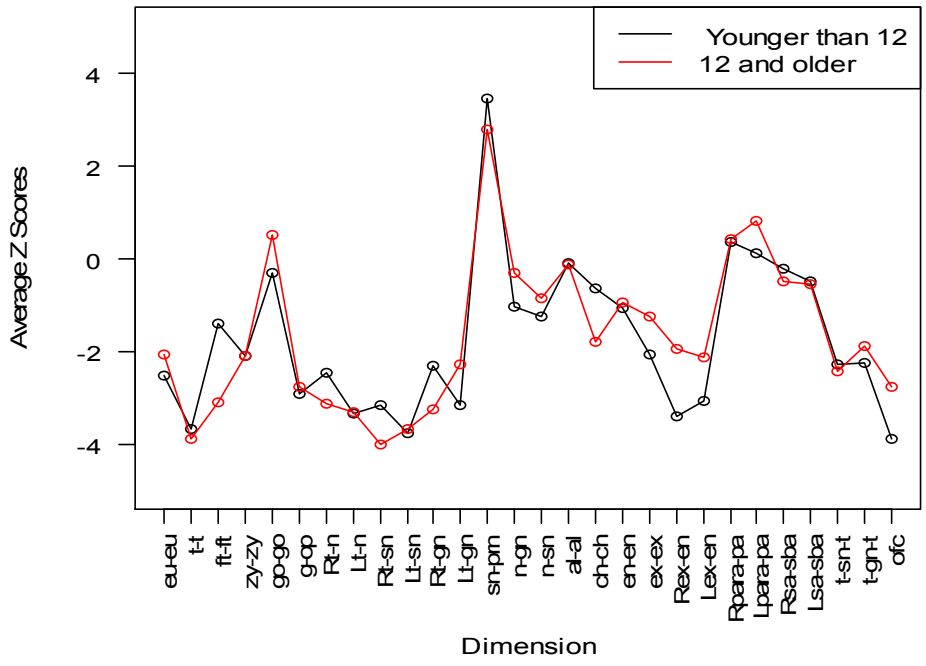

Figure 5: *Pattern profiles by age (r=.89, p<.001)*

CVI distributions for younger than 12 and 12 and older have the characteristics of a normal distribution (similar skewness of 0 and kurtosis of -0.9 and -1), as it can observed in Figure 6 and verified by Shapiro-Wilk tests (p-values of 0.97 in both cases).

To compare the CVI means for males and females we use the Student t test for equal sample sizes (21 and 21) and equal variances (F = 1.979, p-value= 0.1353). It should be noted that correlation between the two CVI vectors is 0.113 and thus the p-value of a paired t test is similar to the standard student t test.

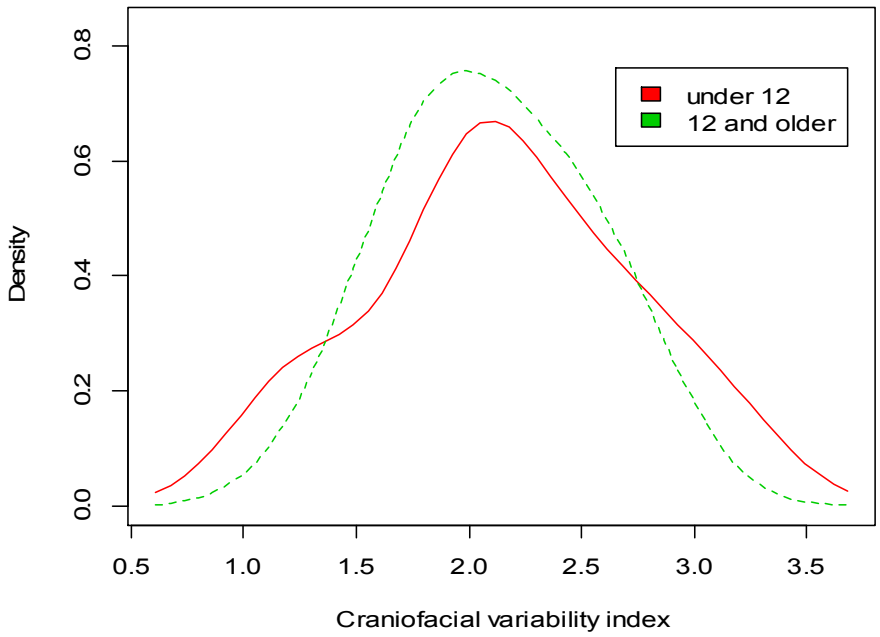

Figure 6: *CVI distribution by age (p=.7 t-test)*

The results of the t-test give a p-value of 0.74 and thus we cannot reject the hypothesis of equal means between groups of age under and over 12 years. CVI medians for younger and older than 12 are very close to each other, 2.09 and 2.07, respectively (Table 11).

Table 11. CVI percentiles for samples split by age range

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Under age of 12 yrs.	21	0.60	1.23	1.49	2.09	2.77	3.15
Over age of 12 yrs.	21	0.59	1.54	1.75	2.07	2.51	2.74

CVI by Severity

The pattern profiles by severity exhibit a high correlation of 0.97. Figure 7 shows a close match between average z-scores for most of the craniofacial measurements. There are some variables where the z-scores differ and the greater differences are observed in the head length (g-op), in the right and left sides of the palpebral fissure length (Rex-en and Lex-en) and also in the head circumference (ofc). Taking the normative reference for age and gender, SLOS individuals with high severity seem to have on average a greater deviation in these specific craniofacial measurements than those with low severity.

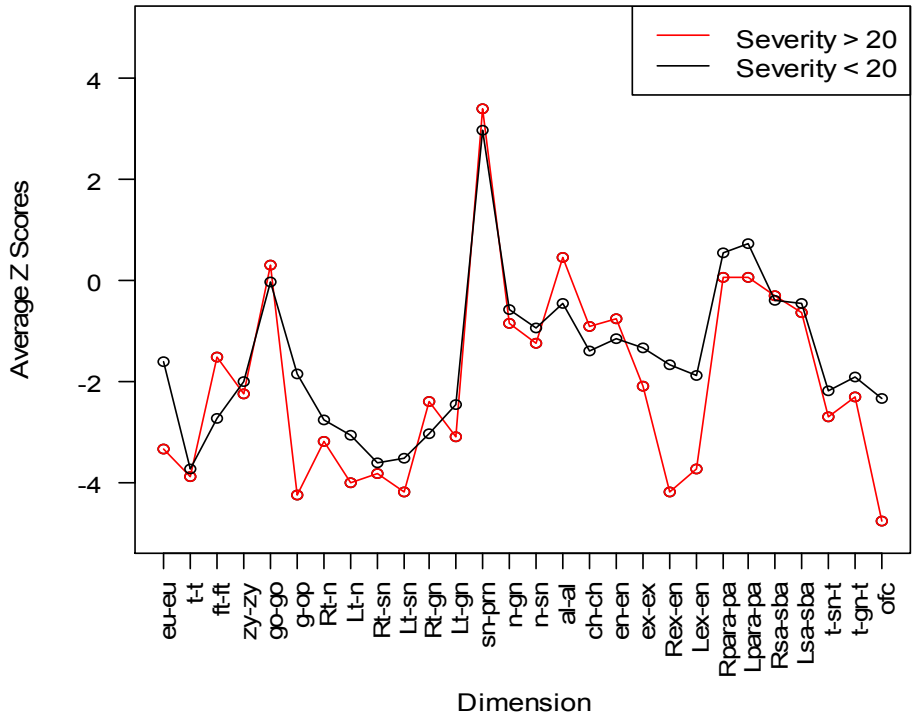

Figure 7: *Pattern profiles by severity (r=.97,p<.0001)*

CVI distributions for low and high severity groups have the characteristics of a normal distribution (similar skewness of 0.2 and 0 and kurtosis of -0.7 and -0.9, respectively), as it can observed in Figure 8 and verified by Shapiro-Wilk tests (p-values of 0.88 and 0.96 for low and high severity groups).

To compare the CVI means for low and high severity groups we use the Student t test for unequal sample sizes (25 and 17) and equal variances (F test p-value= 0.2956).

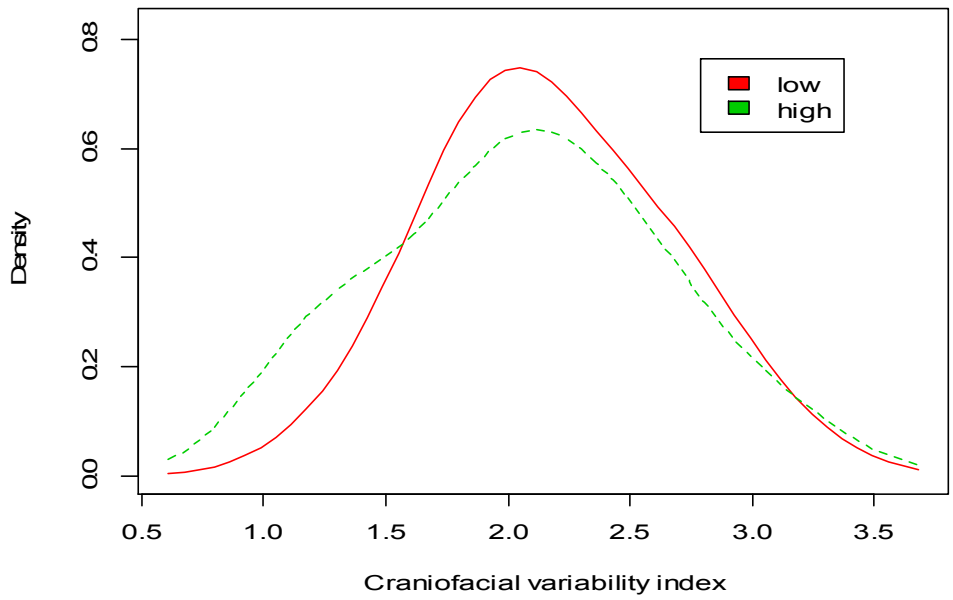

Figure 8: *CVI distribution by severity (p=.5 t-test)*

To test if the means of two CVIs were equal we used the t-test given that the assumptions of normality and equal variances are met. The results of this test give a p-value of 0.5091 and thus we cannot reject the hypothesis of equal means between the groups with low and high severity. CVI medians for low and severity groups are very close to each other, 2.07 and 2.09, respectively (Table 12).

Table 12. CVI percentiles for samples split by severity index

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Low severity index	25	0.63	1.55	1.80	2.07	2.73	2.79
High severity index	17	0.64	1.21	1.40	2.09	2.62	2.85

CVI by Cholesterol level

The pattern profiles by cholesterol exhibit a high correlation of 0.9. Figure 9 shows a close match between average z-scores for most of the craniofacial measurements. There are some variables where the z-scores differ and the greater differences are observed in the head length (g-op), in the right and left sides of the palpebral fissure length (Rex-en and Lex-en) and also in the head circumference (ofc). Taking the normative reference for age and gender, SLOS individuals with low cholesterol seem to have on average a greater deviation in these specific craniofacial measurements than those with high cholesterol. It should be noted that same pattern was observed when comparing high and low severity groups.

CVI distributions for low and high cholesterol groups have the characteristics of a normal distribution (skewness of 0.1 and -0.1 and kurtosis of -0.99 and -0.77, respectively), as it can observed in Figure 10 and verified by Shapiro-Wilk tests (p-values of 0.9 and 0.65 for low and high cholesterol groups).

To compare the CVI means for low and high cholesterol groups we use the Student t test for unequal sample sizes (17 and 25) and equal variances (F test p-value=0. 57).

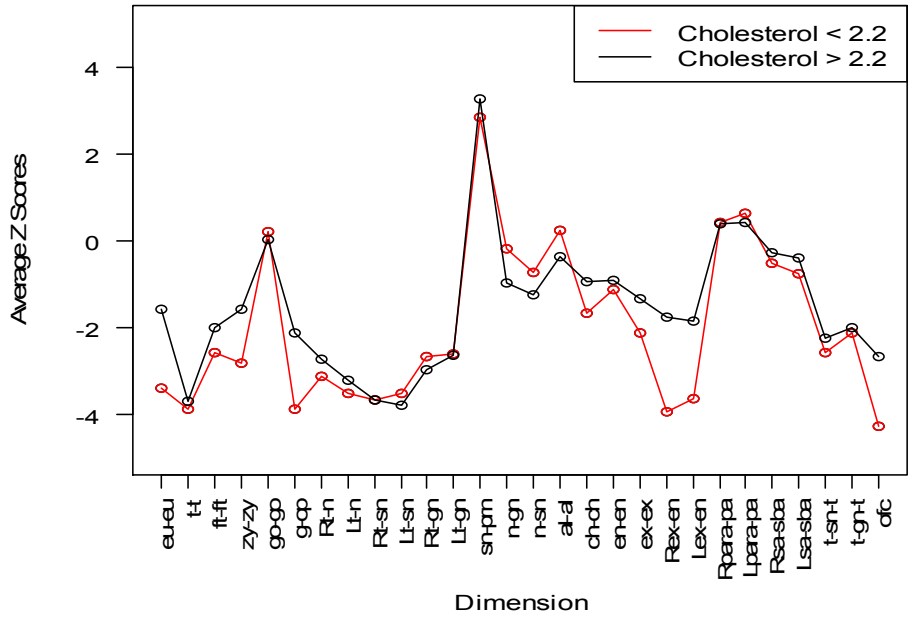

Figure 9: *Pattern profiles by cholesterol (r=.9,p<.001)*

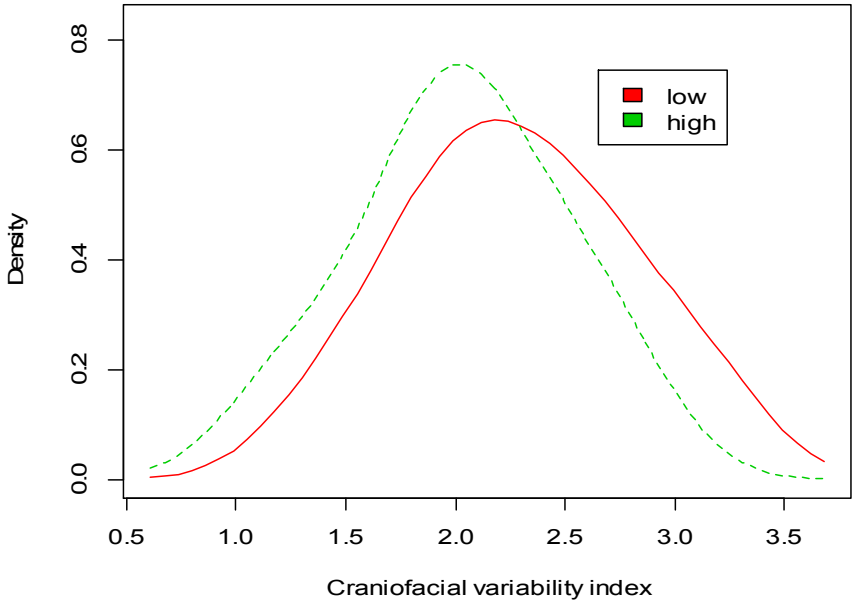

Figure 10: *CVI distribution by cholesterol (p=.12 t-test)*

The results of the t-test give a p-value of 0.1154 and thus we cannot reject the hypothesis of equal means between groups with low and high cholesterol. CVI median for the low cholesterol group is somewhat higher (2.24) than the high cholesterol group (1.99) as it can be observed in Table 13.

Table 13. CVI percentiles for samples split by cholesterol index

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Low cholesterol index	17	0.79	1.50	1.82	2.24	2.75	3.15
High cholesterol index	25	0.52	1.25	1.55	1.99	2.51	2.76

CVI by Severity-Cholesterol

The pattern profiles by cholesterol exhibit a high correlation of 0.88. Figure 11 shows a close match between average z-scores for most of the craniofacial measurements. There are some variables where the z-scores differ and the greater differences are observed in the head length (g-op), in the right and left sides of the palpebral fissure length (Rex-en and Lex-en) and also in the head circumference (ofc). Taking the normative reference for age and gender, SLOS individuals in the group with high severity-low cholesterol seem to have on average a greater deviation in these specific craniofacial measurements than those with low severity-high cholesterol. It should be

noted that same pattern was observed when comparing high and low severity groups as well as low and high cholesterol.

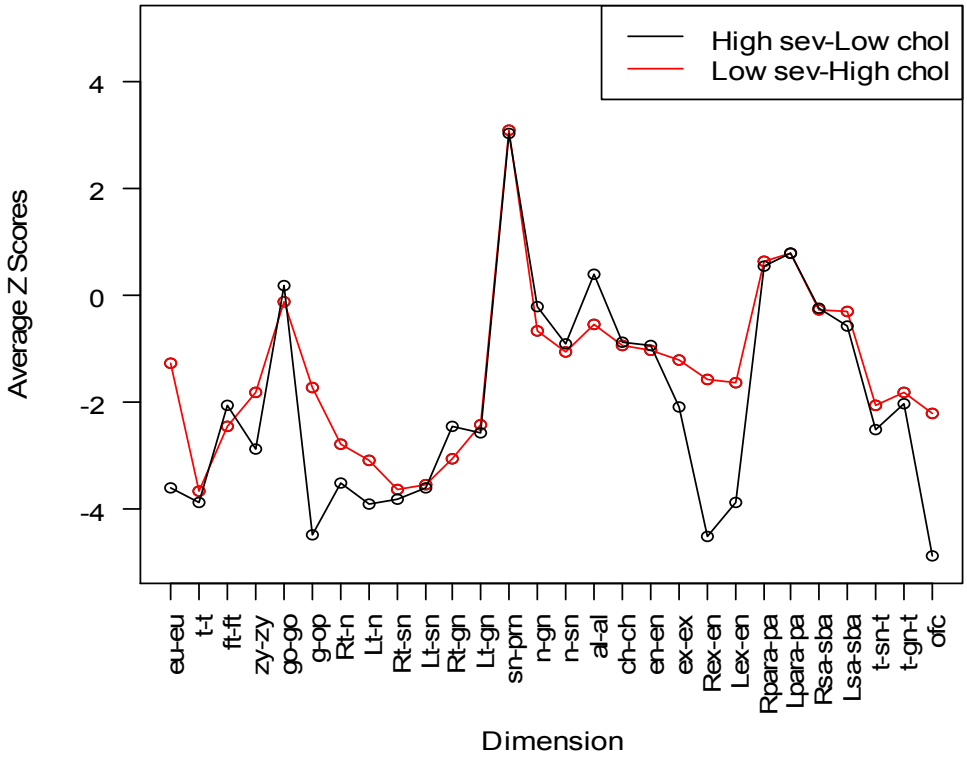

Figure 11: *Pattern profiles by severity-cholesterol (correlation=.88,p<.001)*

Both groups subject to comparison of mean CVI, low severity-high cholesterol and, high severity-low cholesterol, exhibit a normal distribution of CVI (p-values of Shapiro-Wilk test, 0.9 and 0.5, respectively). However, skewness (-0.9) and kurtosis (-1.5) in the high severity-low cholesterol group diverges from the near-zero values observed in the other group. These differences are graphically presented in the CVI density figure.

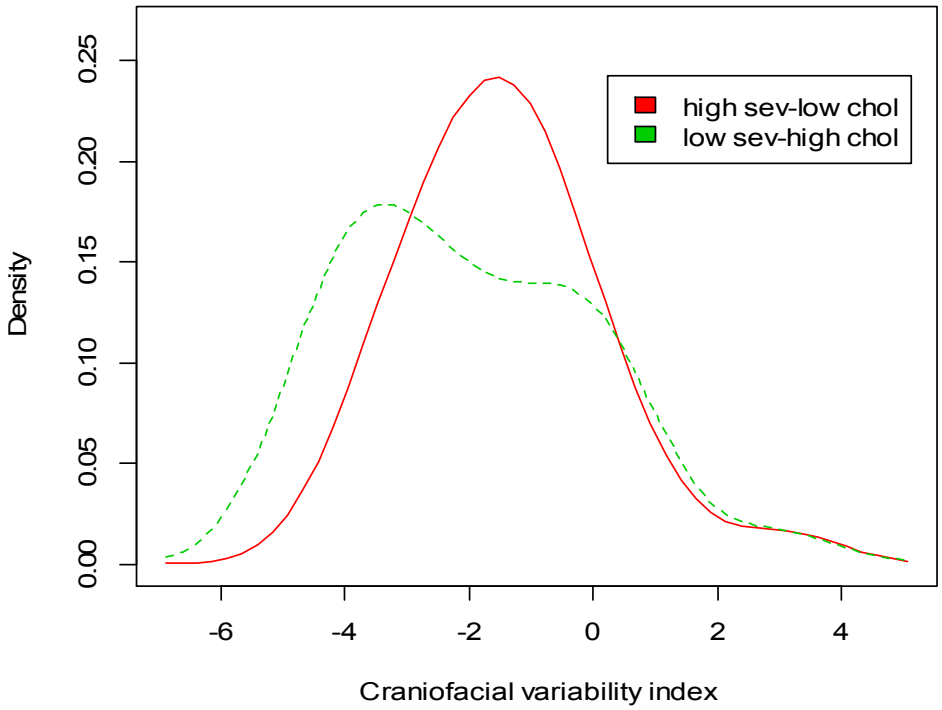

Figure 12: *Distribution of CVI by severity-cholesterol*

Unlike the previous group comparisons, in this case it is noted a greater difference between the CVI medians (Table 14). If we express these median differences in terms of a ratio, we would have a ratio of 1.2. It should be noted that in the comparison of low and high cholesterol groups this ratio was 1.1 and in gender, severity and age groups the ratio was 1.0. The t-test p-value is less than 0.01 and thus we reject the null hypothesis of equal CVI means between these groups.

Table 14. CVI percentiles for samples split by severity-cholesterol

			Percentiles				
Group	Sample size	IQR	5	15	50	85	95
Low severity-high cholesterol	20	0.55	1.23	1.46	1.95	2.43	2.52
High severity-low cholesterol	12	0.69	1.84	1.94	2.37	2.90	3.16

5.2 Individual variance comparisons

In this section we test the hypothesis of equal variances between two related individuals. There are 6 pairs of related individuals whose ratios of CVI ranges from 1.07 to 1.5. The CVI and other characteristics of the data for all the 42 patients are provided in Table 16, in the Appendix.

Individuals with ID 2 and 3 (Siblings)

The pattern profiles of this pair of siblings exhibit a relatively low correlation of 0.4. Figure 13 shows how well match z-scores for the 28 craniofacial measurements of the

two siblings. There are some variables where the z-scores markedly differ and the greater differences are observed in the nasal protrusion (sn-prn) and nasal width (al-al). Taking the normative reference for age and gender, individual with ID 2 seems to have a greater deviation in these specific craniofacial measurements than sibling with ID 3. It should be noted that CVI of ID 2 (1.3) is higher than his sibling (1.12). Both individuals are boys and belong to the same group of low severity-high cholesterol. The age of ID 2 is 6 years and his sibling is 4 years old.

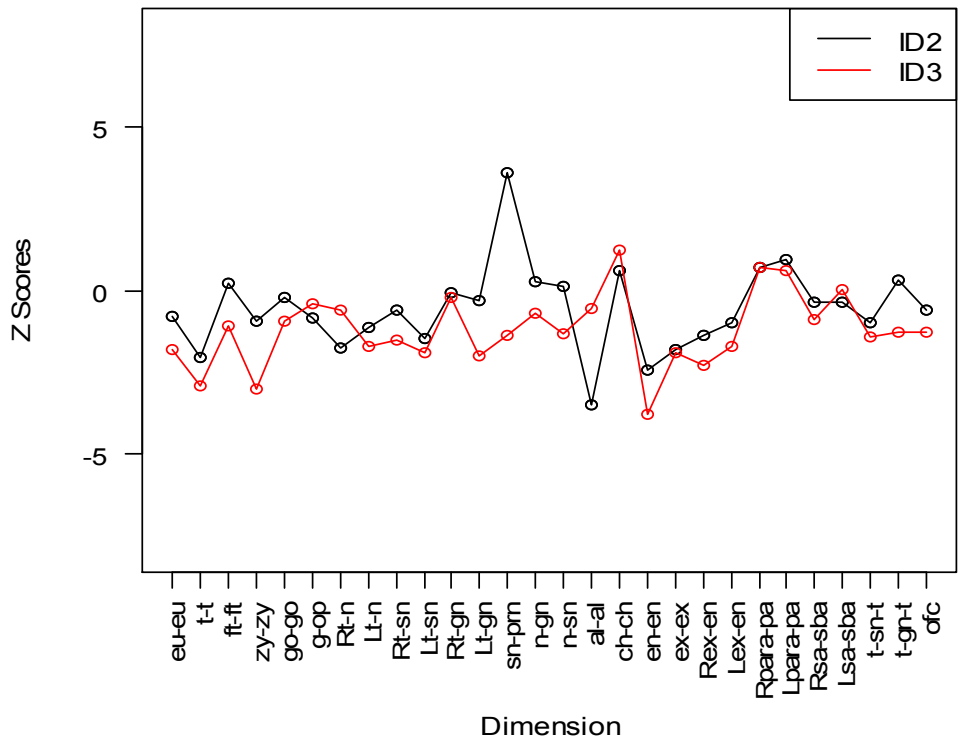

Figure 13: *Pattern profiles of two siblings (r=.4,p=.02)*

With regard to the characteristics of their distributions, we will examine for each individuals whether or not his observations are independent and normally distributed.

Firstly, in relation to randomness the runs test provides p-values greater than 0.05 and thus the null hypothesis of randomness is accepted.

Secondly, the Shapiro-Wilk test gives p-values greater than 0.05 for both vectors and therefore the null hypothesis of normality can be accepted. However, it should be noted that vector of individual 2 has a kurtosis of 2.1 which is reflected in the below distribution graph as a sharper peak and longer tails. On the other hand, distribution of individual 3 has a kurtosis of zero and in the graph it is seen with a more rounded peak and one shorter tail. Skewness is close to zero in both samples.

In short these distributions have equal sample sizes, a normal distribution but with different kurtosis, randomness in each sample and low correlation between samples.

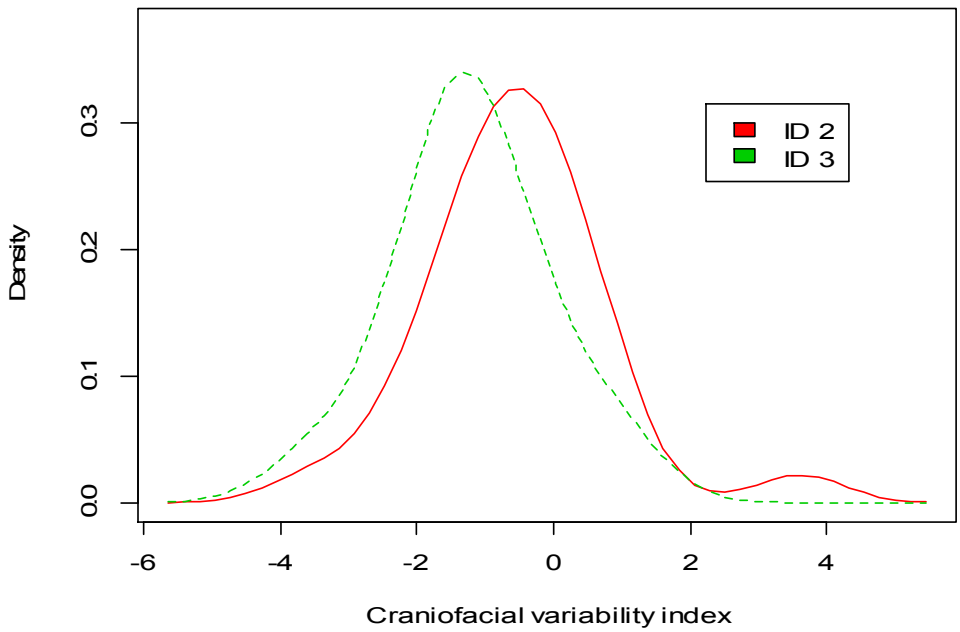

Figure 14: *Distribution of two relatives (p=.45, F test)*

Keeping in mind the above characteristics we proceed to test the equality of variances using three tests: F, Levene-median and Fligner. The F test of equality of variances gives a p-value of 0.4525 (F ratio=1.3) indicating that the null hypothesis of variance ratio equal to 1 cannot be rejected. Similar results are obtained with the Levene test- median (p-value=0.72) and Fligner tests (p-value=0.85).

<u>Individual 7 versus 8 (Siblings)</u>

The pattern profiles of this pair of siblings exhibit a high correlation of 0.9. Figure 15 shows how well match z-scores for the 28 craniofacial measurements of the two

siblings. There are some variables where the z-scores markedly differ and the greater differences are observed in the inner and outer canthal distance (en-en; ex-ex).

Taking the normative reference for age and gender, individual with ID 7 seems to have a greater deviation in these specific craniofacial measurements than sibling with ID 8. It should be noted that CVI of ID 7 (2.51) is higher than his sibling (2.35). Both individuals are men and belong to the same group of low severity-high cholesterol. The age of ID 7 is 25 years and his sibling is 23 years old.

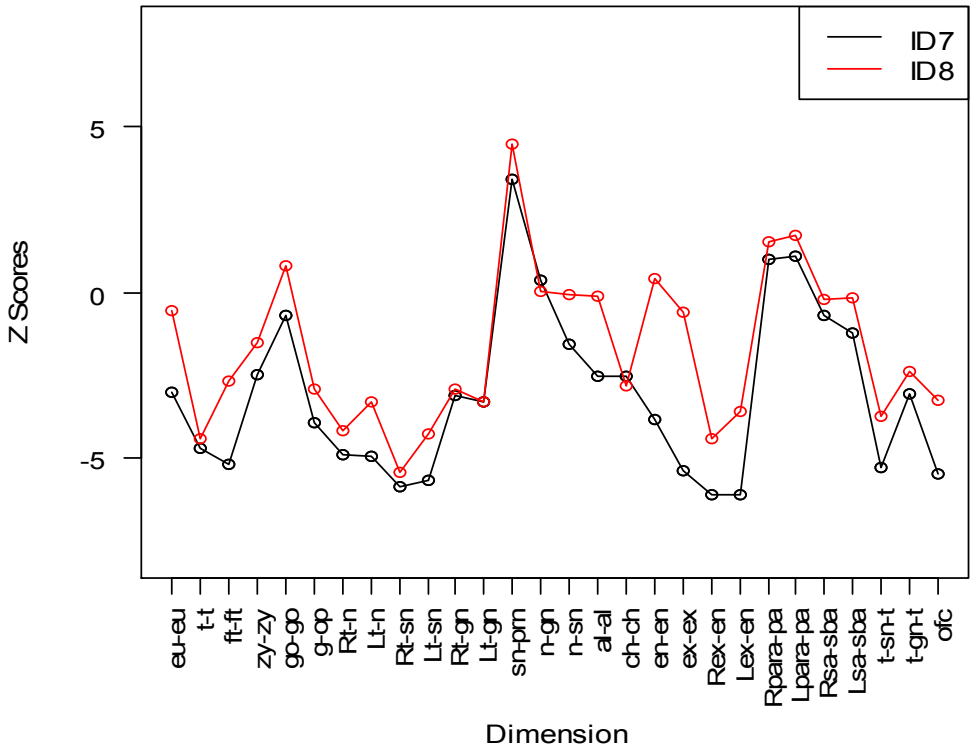

Figure 15: *Pattern profiles of two siblings (correlation=.9,p=.001)*

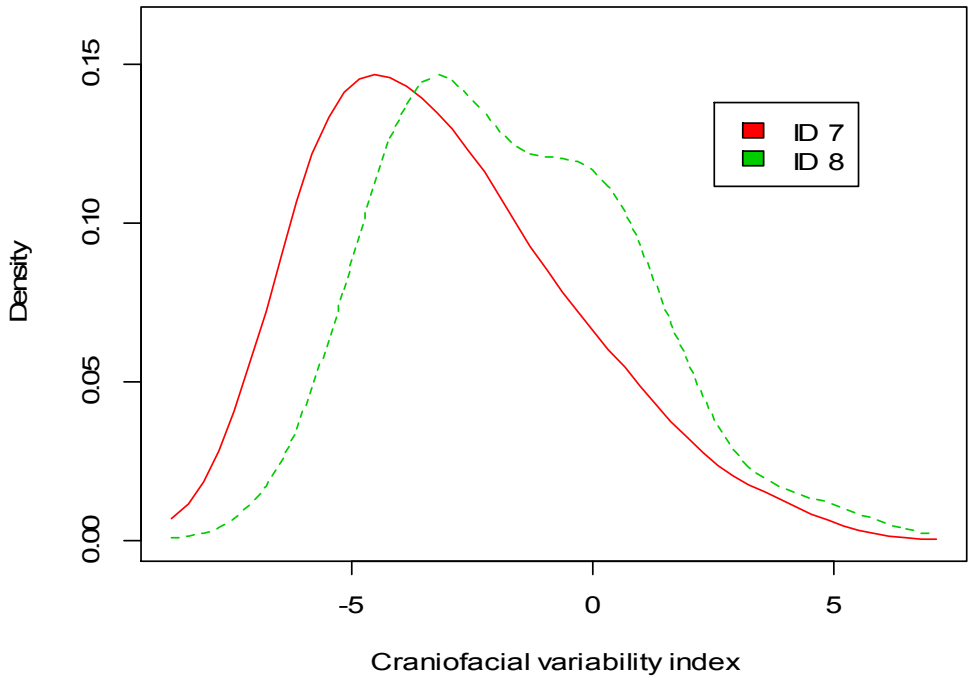

Figure 16: *Distribution of two relatives (p=.74, F-test)*

There are two characteristics of the distributions of z-scores for these individuals that differ from the previous case. A similar skewness and kurtosis in the normally distributed samples -close to zero in both cases- and, a high correlation of 0.9 between samples.

In this case, three tests of equality of variances (F, Levene-median and Fligner) give p-values greater than 0.05 supporting the null hypothesis of equal variances.

Individual 39 versus 40 (Siblings)

The pattern profiles of this pair of siblings exhibit a correlation of 0.7. Figure 17 shows how well match z-scores for the 28 craniofacial measurements of the two siblings. There are some variables where the z-scores markedly differ and the greatest difference is observed in the nasal protrusion (sn-prn).

Taking the normative reference for age and gender, individual with ID 39 seems to have a greater deviation in this specific craniofacial measurement than sibling with ID 40. It should be noted that CVI of ID 39 (1.99) is higher than his sibling (1.59). Both individuals are men and belong to the same group of low severity-high cholesterol. The age of ID 39 is 16 years and his sibling is 14 years old.

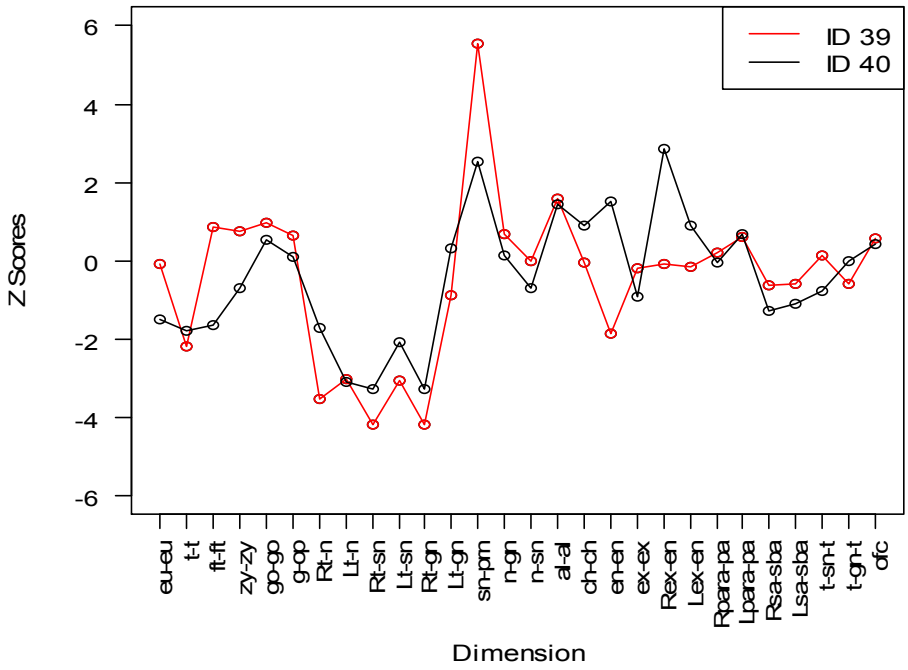

Figure 17: *Pattern profiles of two siblings (correlation=.7, p=.001)*

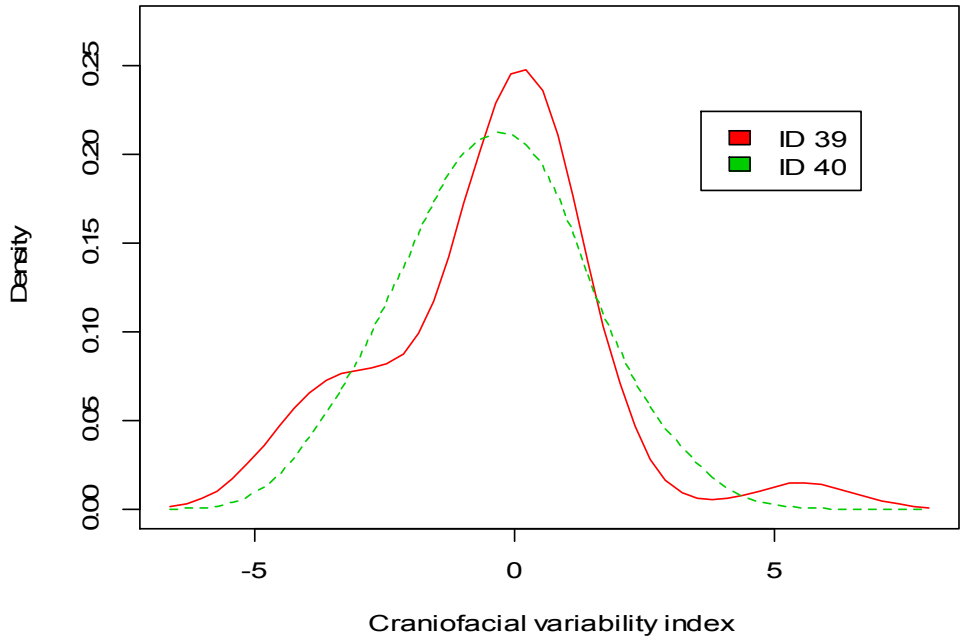

Figure 18: *Distribution of two relatives (p=.25, F test)*

The non-normal distribution of individual 39 has a higher kurtosis than normal distribution of individual 40 (1.27 versus 0.62). Skewness is very close to zero in both individual distributions.

In this case, three tests of equality of variances (F, Levene-median and Fligner) also give p-values greater than 0.05 supporting the null hypothesis of equal variances.

<u>Individual 18 versus 35(Relatives)</u>

The pattern profiles of this pair of relatives exhibit a correlation of 0.7. Figure 19 shows how well match z-scores for the 28 craniofacial measurements of the two

siblings. There are some variables where the z-scores markedly differ and the greater differences are observed in the mouth width (ch-ch) and head length (g-op).

Taking the normative reference for age and gender, individual with ID 18 seems to have a greater deviation in this specific craniofacial measurement than relative with ID 35. It should be noted that CVI of ID 18 (2.48) is higher than his relative (1.86). Both individuals are women and belong to the same group of low severity but to a different group of cholesterol (low and high, respectively). The age of ID 18 is 13 years and her relative is 14 years old.

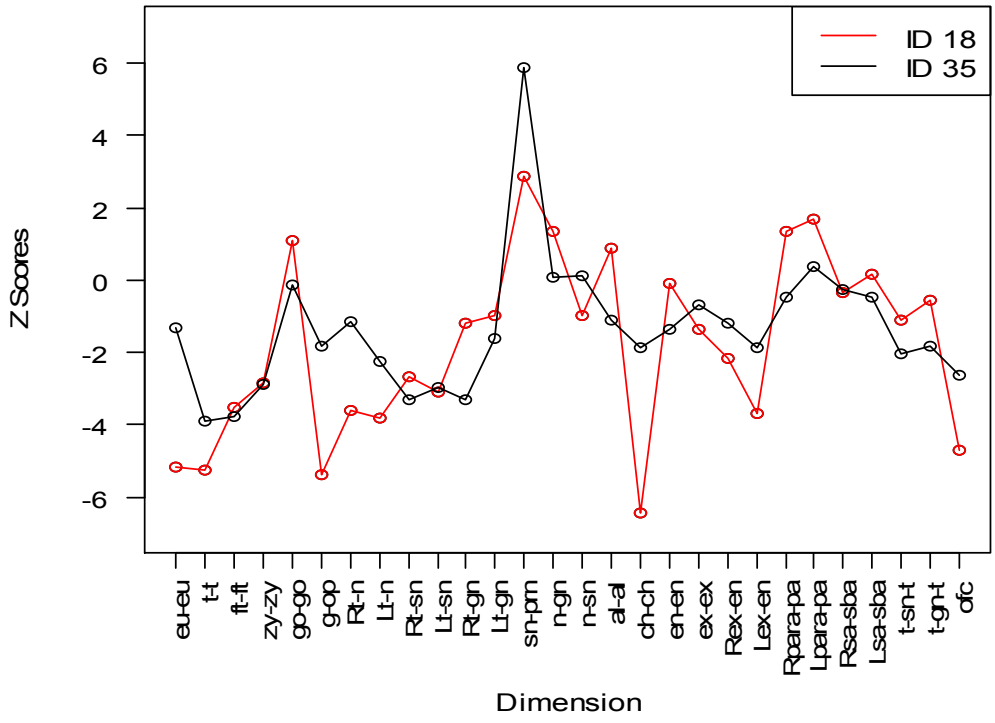

Figure 19*: Pattern profiles of two relatives (correlation=.7, p=.001)*

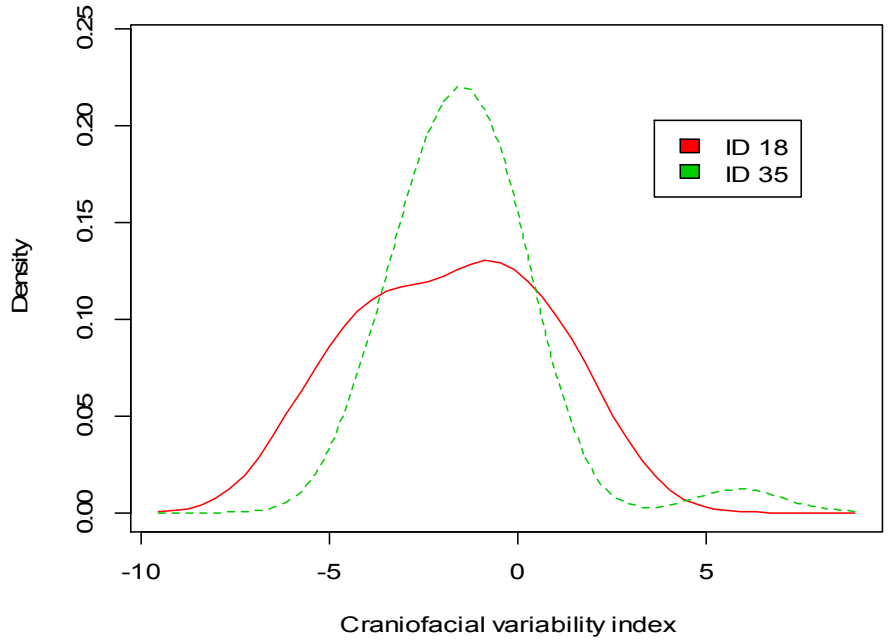

Figure 20: *Distribution of two relatives (p=.14, F test)*

The non-normal distribution of individual 35 has a considerably higher kurtosis than normal distribution of individual 18 (5.6 versus 1.1). Skewness is very close to zero in individual 18 and near 2 in individual 35.

In this case the F test of equality of variances gives a p-value of 0.1392 indicating that the null hypothesis of variance ratio equal to 1 could not be rejected. In contrast, the Levene test (median) and Fligner tests provide p-values smaller than 0.05 supporting the rejection of the null hypothesis of equality of variances. These apparent contradictory results may reflect a greater sensitivity of F test to influential observations or outliers.

Individual 26 versus 41(Cousins)

The pattern profiles of this pair of cousins exhibit a correlation of 0.5. Figure 21 shows how well match z-scores for the 27 craniofacial measurements common to the two cousins. There are some variables where the z-scores markedly differ and the greatest difference is observed in the head circumference (ofc).

Taking the normative reference for age and gender, individual with ID 26 seems to have a greater deviation in this specific craniofacial measurement than relative with ID 41. It should be noted that CVI of ID 26 (2.07) is higher than his relative (1.83). Both individuals are men and belong to opposite groups of severity-cholesterol. ID 26 is in the category of high severity-low cholesterol. The age of ID 26 is 14 years and his relative is 15 years old.

Both individuals have normal distributions with similar skewness (around zero) and kurtosis (close to -1). Due to missing data the number of observations in individual 26 is one less than individual 41. With regard to randomness, runs test for individual 26 rejects the null hypothesis of randomness. In this case, three tests of equality of variances (F, Levene-median and Fligner) also give p-values greater than 0.05 supporting the null hypothesis of equal variances.

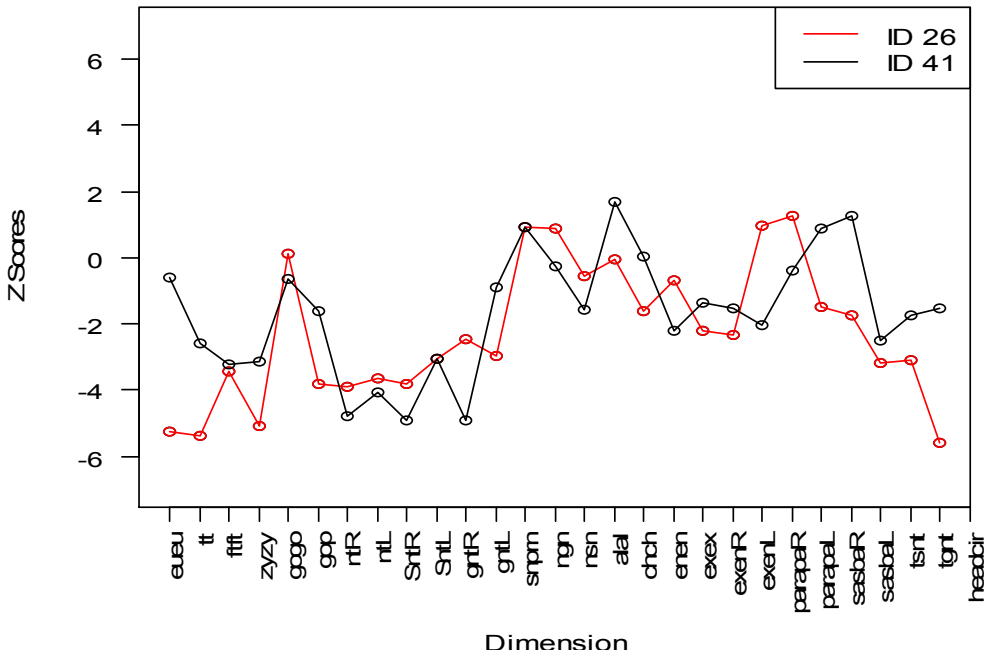

Figure 21: *Pattern profiles of cousins (correlation=.5, p=.007)*

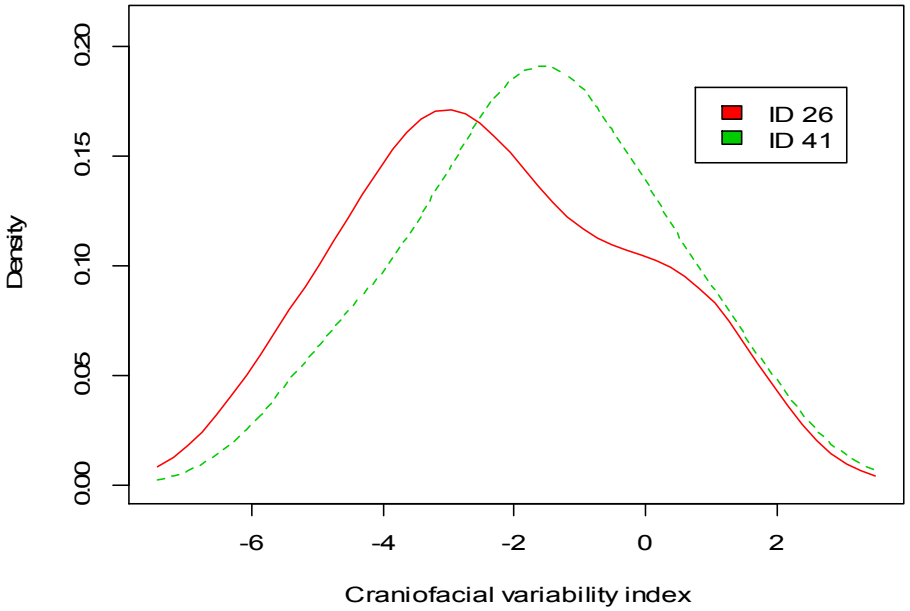

Figure 22: *Distribution of two relatives (p-value=.5, F test)*

Individual 6 versus 41(unrelated teens)

We have included this case of comparison of two unrelated teenagers from opposite sex (ID 6= female, ID 41=male) with a sizeable CVI difference (2.79 versus 1.83) to test if their variances are significantly different.

Both individuals have complete information for the set of 28 craniofacial measurements and they have similar age (14.9 years) and equivalent scale of severity and cholesterol (both belong to the category of low severity-high cholesterol).

Their pattern profile exhibits a correlation of 0.46 at the significance level of 0.01. It should be noted that z-scores graph (Figure 26) shows a remarkable variation for ID 6 in two variables (Rex-en and Lex-en), which measure the length between the outer corners of the right and left eye, respectively.

The influence of these variables in the CVI can be verified by comparing the sample standard deviation for both individuals with and without Rex-en and Lex-en. Without these variables the CVI would be 1.79 for ID 6 and 1.89 for ID 41.

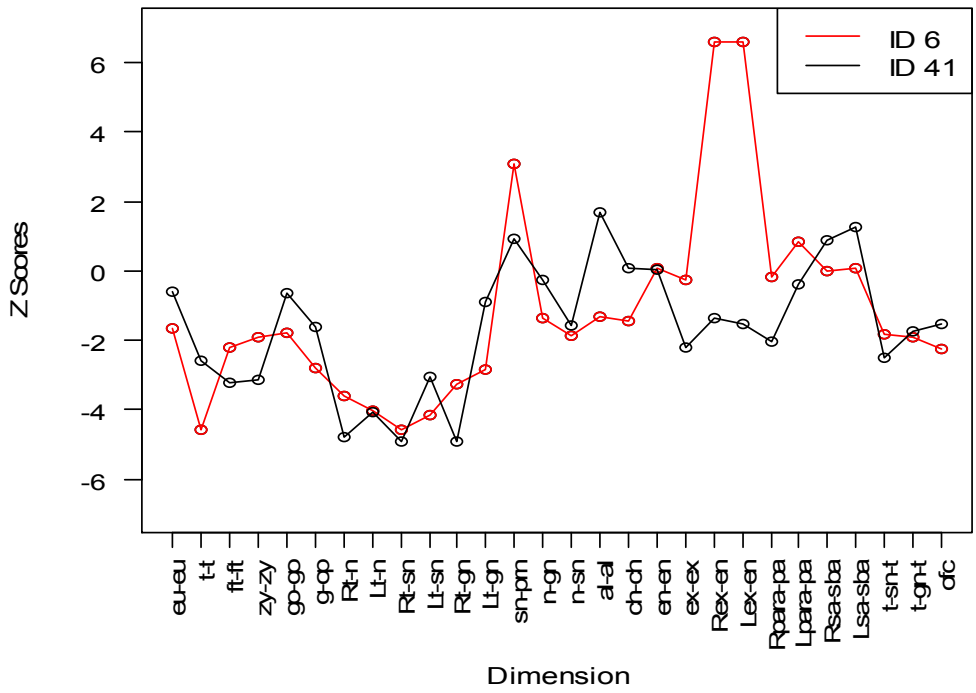

Figure 23: *Pattern profiles of two teens (correlation=.46, p=.01)*

The probability distribution of z-scores for both individuals (Figure 24) also shows that the above mentioned variables cause an abnormality in the distribution of ID 6. The p-value of Shapiro-Wilk test supports the rejection of the null hypothesis of normality for ID 6 distribution at the significance level of 0.01. In contrast, a p-value of 0.6 supports the hypothesis of normality in the case of ID 41. Skewness (-0.2) and kurtosis (-0.8) for the distribution of ID 41 diverges from ID 6 (1.4 and 1.8, respectively).

Hypothesis testing of equal variances give different results with the classical F test and with Levene-median and Fligner-Killeen tests. While the F test provides a p-value of 0.03 that supports the rejection of the null hypothesis of variance ratio equal to 1, Levene and Fligner tests give p-values of 0.7 and 0.8 that fail to reject the null hypothesis of equal variances. The low p-value of the F test can be explained by its loss of power under mixture of normal and non-normal distributions, as it is shown in the simulation study of Balakrishnan (1990).

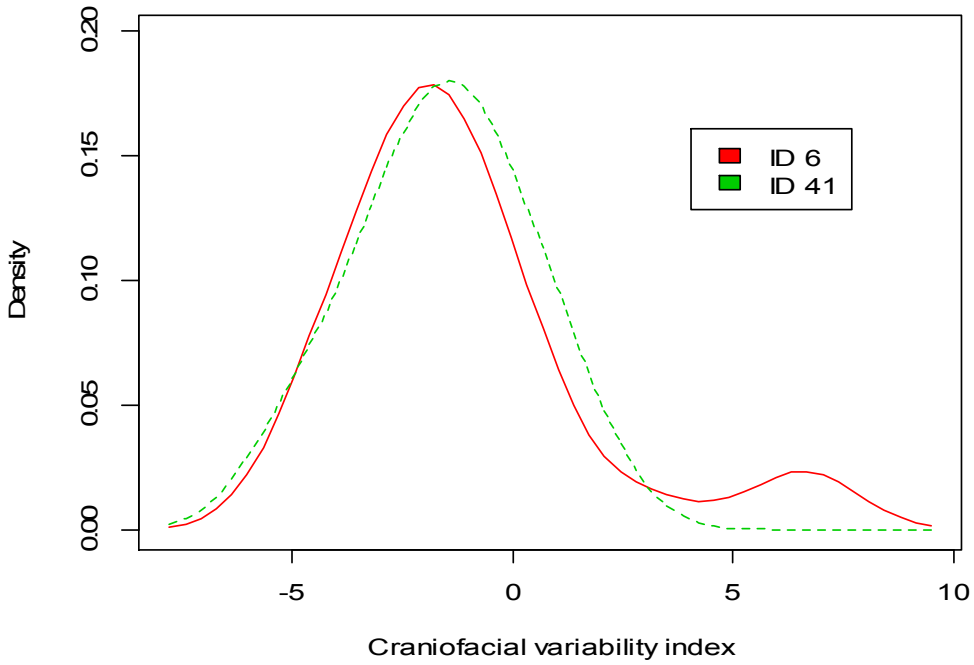

Figure 24: *Distribution of two teens (p=.03, F test)*

Chapter 6

DISCUSSION

Craniofacial variability index (CVI), or the standard deviation of z-scores for craniofacial measurements of one or more individuals, is considered in the medical literature as a useful and relatively simple quantitative measure of the degree of dysmorphogenesis in the head and face. Pioneering studies made by Garn (1985) suggested that the greater CVI the greater likelihood of a malformation syndrome. Moreover, Garn et.al set a CVI threshold of 1.2 as a guideline for medical practice to identify possible cases of malformation syndromes.

Distribution of CVI is used to compare degree of facial variability of individual patients or group of patients with syndromes with that of healthy population. Comparisons have also been done between individuals and groups diagnosed with the same or similar craniofacial syndromes. Moreover, several studies use average CVI to compare degree of dysmorphogenesis among groups of patients. However, all the comparisons considered in previous studies have been subjective where no statistical testing of the values was provided.

In this context, the contribution of this thesis is to provide alternative ways of testing the significance of the differences for CVI: both at individual and group levels. We also performed extensive simulations to examine the level and power of equality of the tests under assumptions that characterize datasets of craniofacial measurements, and

for various sample sizes. Our main findings are that the higher the correlation among variables (within individual and group dependence) the higher the power of variance tests with a level or type I error of the nominal value, α=0.05. On the other hand, between individual dependence reduces both the level and power as correlation increases. The combination of within and between individual dependence has the effect of lowering the power in comparison with the scenario of independence. It is also observed that the higher the correlation the lower the power. It should be noted that in the case of group comparison the simulation scenario of within and between dependence shows opposite results to those observed in the individual comparison. A higher correlation is associated to both a higher level and power.

On another note, SLOS dataset analysis has shown that comparison of pattern profile between individuals is a useful tool to identify influential z-scores of craniofacial measurements that affect CVI and the subsequent results of hypothesis testing of equal variances using the classical F test versus Levene-median and Fligner-Killeen tests.

BIBLIOGRAPHY

Allanson Judith E., Greenberg Frank., and Smith Ann C.M. (1999). The face of Smith-Magenis syndrome: a subjective and objective study. *Journal of Medical Genetics 36:394-397*.

Balakrishnan N. and Ma C.W (1990). A Comparative Study of Various Tests for the Equality of Two Population Variances. *Journal of Statistical Computation and Simulation 35:1-2, 41-89.*

Bartlett M.S (1937). Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London. Series A Mathematical and Physical Sciences, Vol.160,No.901, pp.268-282.*

Box G.E.P(1953). Non-Normality and Tests on Variances. *Biometrika Vol.40,No.3/4 pp.318-335.*

Garn, Stanley M., Lavelle, Marquisa and Smith, B. Holly (1984). Applications of Pattern Profile Analysis to Malformations of the Head and Face. *Radiology 150:683-690.*

Garn, Stanley M., Lavelle, Marquisa and Smith, B. Holly (1985). Quantification of Dysmorphogenesis: Pattern Variability Index. *American Journal of Radiology 144:365-69.*

Garn, Stanley M., Leonard William R., and Poznanski Andrew K.(1987). Applications of the Pattern Variability Index to the Quantification of Dysmorphogenesis in the Hand. *American Journal of Medical Genetics 27:143-152.*

Laurencikas E., Savendahl L. & Jorulf H.(2006). Metacarpophalangeal Pattern Profile Analysis: Useful Diagnostic Tool for Differentiating between Dyschondrosteosis, Turner Syndrome, and Hyppochondroplasia. *Acta Radiologica vol.47 no.5, 518-524.*

Laurencikas E., Soderman E., Grigelioniene G., Hagenas L. & Jorulf H.(2005). Metacarpophalangeal Pattern Profile Analysis in Leri-Weill Dyschondrosteosis. *Acta Radiologica vol.46 no.2, 200-207.*

Laurencikas E., Soderman E., Davenport M., Horulf H. & Savendahl L. (2005). Metacarpophalangeal Pattern Profile Analysis as a Tool for Early Diagnosis of Turner Syndrome. *Acta Radiologica vol.46 no.4, 424-430.*

Levene, Howard (1960) Robust Tests for Equality of Variances. In Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling.

Marozzi, Marco (2011) Levene type tests for the ratio of two scales. *Journal of Statistical Computation and Simulation, Vol.81, No.7, 815-826.*

Nowaczyk M.J.M, Tan M., Hamid J., Allanson J. (2011). Smith-Lemli-Opitz Syndrome: Subjective and Objective Assessment of Facial Phenotype. *American Journal of Medical Genetics*, In press.

Shoemaker, Lewis H.(2003). Fixing the F tests for Equal Variances. *The American Statistician, Vol.57, No.2,pp.105-114.*

Zimmermann, Donald W. (1994). A Note on the *F* test for Equal Variances Under Violation of -Random Sampling. *The Journal of General Psychology*, 121 (1), 77-83.

APPENDIX

Table 15. Craniofacial measurements used in the study

No.	Variable	Description	Latin name
1	eu.eu	Head width	eurion to eurion
2	t.t	Skull base width	tragion to tragion
3	ft.ft	Forehead width	frontotemporale to frontotemporale
4	zy.zy	Upper facial width	zygion to zygion
5	go.go	Mandible width	genion to genion
6	g.op	Head length	glabella to opisthocranion
7	n.t.r	Upper facial depth (right)	nasion to tragion (right)
8	n.t.l	Upper facial depth (left)	nasion to tragion (left)
9	sn.t.r	Middle facial depth (right)	subnasale to tragion (right)
10	sn.t.l	Middle facial depth (left)	subnasale to tragion (left)
11	gn.t.r	Lower facial depth (right)	gnathion to tragion (right)
12	gn.t.l	Lower facial depth (left)	gnathion to tragion (left)
13	sn.prn	Nasal protrusion	subnasale to pronasale
14	n.gn	Total facial height	nasion to gnathion
15	n.sn	Upper facial height	nasion to subnasale
16	al.al	Nasal width	alare to alare
17	ch.ch	Mouth width	cheilion to cheilion
18	en.en	Inner canthal distance	endocanthion to endocanthion
19	ex.ex	Outer canthal distance	exocanthion to exocanthion
20	ex.en.r	Palpebral fissure length (right)	exocanthion to endocanthion (right)
21	ex.en.l	Palpebral fissure length (left)	exocanthion to endocanthion (left)
22	para.pa.r	ear width (right)	preaurale to postaurale (right)
23	para.pa.l	ear width (left)	preaurale to postaurale (left)
24	sa.sba.r	ear length (right)	superaurale to subaurale (right)
25	sa.sba.l	ear length (left)	superaurale to subaurale (left)
26	t.sn.t	Maxillary arc	tragion to subnasale to tragion
27	t.gn.t	Mandibular arc	tragion to gnathion to tragion
28	ofc	Head circumference	ofisthocranion

Table 16. Characteristics of the sample study by individual

ID	Gender	Age	CVI	No.variables	Severity	Cholesterol
1	Female	2.54	2.45	22	High	Low
2	Male	6.00	1.30	28	Low	High
3	Male	4.34	1.12	28	Low	High
4	Female	10.64	1.91	27	High	Low
5	Male	2.97	3.15	27	High	Low
6	Female	14.94	2.79	28	Low	High
7	Male	25.02	2.51	28	Low	High
8	Male	22.69	2.35	28	Low	High
9	Male	4.40	2.24	10	High	Low
10	Male	0.78	3.17	16	High	Low
11	Male	2.76	2.19	18	High	High
12	Male	16.33	1.77	28	Low	High
13	Male	2.43	1.96	20	High	Low
14	Male	3.18	1.88	28	High	High
15	Male	13.85	2.74	10	High	High
16	Male	19.30	2.41	27	Low	High
17	Male	6.63	2.72	28	High	Low
18	Female	12.58	2.48	28	Low	Low
19	Female	13.68	2.72	25	High	Low
20	Female	2.53	2.09	27	High	High
21	Male	11.79	2.30	22	High	Low

Table 16. Characteristics of the sample study by individual (continuation)

ID	Gender	Age	CVI	No.variables	Severity	Cholesterol
22	Male	10.94	1.99	28	Low	High
23	Male	7.99	1.49	28	Low	High
24	Female	0.65	2.77	15	High	Low
25	Female	6.53	1.23	28	Low	High
26	Male	13.84	2.07	27	High	Low
27	Female	4.54	2.51	23	Low	High
28	Female	12.96	1.54	23	Low	Low
29	Female	9.46	1.93	27	Low	Low
30	Female	1.94	2.06	27	Low	High
31	Male	12.62	2.12	27	Low	High
32	Male	3.76	2.77	16	High	High
33	Female	17.83	1.75	24	High	Low
34	Female	20.17	2.24	28	Low	Low
35	Female	13.94	1.86	28	Low	High
36	Female	15.86	1.88	28	Low	High
37	Male	15.16	1.90	28	Low	High
38	Female	22.91	2.25	28	Low	High
39	Male	15.98	1.99	28	Low	High
40	Male	14.31	1.59	28	Low	High
41	Male	14.93	1.83	28	Low	High
42	Male	16.64	1.35	28	Low	Low