Uniwersytet Przyrodniczy we Wrocławiu
Wydział Biologii i Hodowli Zwierząt
Kierunek: Bioinformatyka
Studia stacjonarne drugiego stopnia
Specjalność: Biostatystyka i programowanie bioinformatyczne

Aneta Sowińska
115962

Analysis of selected indicators for tonsillectomy

Praca wykonana pod kierunkiem:
dr Anna Mucha
Katedra Genetyki

Wrocław, 2020

Wrocław, dnia ....

Oświadczenie opiekuna pracy

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Podpis opiekuna pracy.........................

Oświadczenie opiekuna pracy

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Podpis autora pracy ....

Abstract

Analysis of selected indicators for tonsillectomy

Bootstrapping, and other resampling methods, has considerable advantages when it is applied to non-Gaussian data or data with small sample.

Most of the published literature is concerned with the mathematical aspects of the bootstrap, however, the use of this technique in medicine and other fields, including biological sciences, is becoming more common.

In the following work the bootstrap was used for hypothesis testing and the estimation of standard errors, bias and confidence intervals for parameters. The results were compared and contrasted with standard statistical inference outcomes. The standard methods included t-tests, Wilcoxon tests and summary measures.

Overall, in the studied data, bootstrap method produce results similar to conventional statistical methods.

Keywords: tonsillectomy, halitosis, bootstrap, PCA.

Contents

Chapter 1

Introduction

The analysis of medical data is characterized by the difficulty of statistical verification of their results, but the health care industry also struggle with issues around data storage and access, data quality, pipeline reliability, privacy and security. With increasing computational power, resampling methods such as bootstraping are becoming more popular for investigating problems arising during data-driven model building.

Moreover, these methods can be used in order to overcome the problem of datasets with insufficient sizes. This approach helps in building a probability distribution based on much larger number of observations, and thus statistical inference. In general, resampling consists of multiple repetitions of a procedure of drawing samples from the original data to draw certain conclusions about the distribution of a population [1].

The main purpose of this article is to present one of the possible resampling methods - the bootstrap method - and its application in the analysis of medical data. Also, Principal Component Analysis is being considered as an effective method to reduce the dimensionality of the data to a few principal variables which can explain most of the variance present in the original data, as well as to characterize the mutual position of observation under given conditions. The research hypothesis that will be put forward here is the assumption that halitosis can be an indication for tonsillectomy.

1.1 Chronic tonsillitis

Chronic tonsillitis most often develops as a result of absence or incorrect treatment of angina [2]. Food debris, exfoliated epithelial cells and bacteria remaining in the expanded crypts might form repentant plugs over time. As a consequence, of a prolonged inflammatory process ulceration or fibrosis of the tonsil tissue may occur. These conditions favor the development of aerobic and anaerobic bacteria, as well as fungi. Chronic tonsillitis is about the only reason for removal of palatine tonsils.

1.2 Halitosis

The concept of halitosis is defined as a bad breath. Although it can be caused by many factors, most of them are located within the mouth. The unpleasant odor may be related to cavities, periodontal disease, poor oral hygiene or other dental problems. Also, otolaryngological factors and respiratory diseases, gastrointestinal diseases, hepatic, pancreatic and nephritic insufficiencies may be the source of a problem. In addition, certain foods or drinks sometimes result in temporary malodorous.

The presence of various substances in the exhaled air is directly responsible for unpleasant odors. These substances arise as products of bacterial metabolism and are formed by the decomposition of organic material, such as food debris, exfoliated epithelial cells, dead leukocytes, and others. The degradation of proteins, peptides and amino acids containing thiol groups leads to the formation of compounds such as indols, skatols, volatile sulfur compounds (VSC) including hydrogen sulfide, methyl mercaptan and dimethyl sulfide responsible for the odor.

Studies have shown that halitosis reflects complex interactions between several oral Gram-negative, most of which are anaerobes, however no particular bacterial strain is responsible for the condition [3].

Chapter 2

Material and methods

2.1 Material

41 patients including 29 women and 13 men, ranging age from 18 to 56 years (mean 27.02, SD 7.86) with chronic tonsillitis and halitosis were qualified for tonsillectomy and therefore selected for the study. A detailed anamnesis was conducted by otolaryngologists and dentists, and microbiological tests were carried out in the laboratory of the Academic Clinical Hospital in Wroclaw. Also, routine laboratory tests and two surveys were conducted in order to eliminate causes of halitosis other than those indicative of tonsillitis. Patients answered questions related to their medical history, food habits, smoking cigarettes and oral hygiene. Those who confirmed the occurrence of any gastrointestinal, pulmonary or other systemic metabolic disorders in the survey were eliminated from the research group. Another exclusion criteria included smokers, heavy alcoholic drinkers, improper oral hygiene and *Helicobacter pylori* carriers. Dental examination excluded also patients with carious lesions, exposed tooth pulps, peridonthal diseases and thick tongue coat.

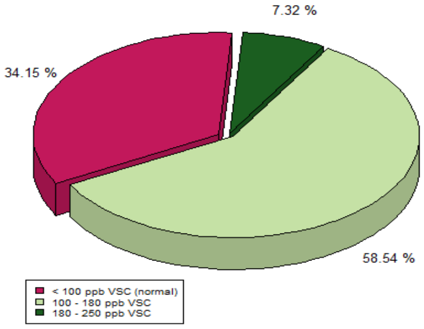
2.2 Methods

Patients underwent both subjective and objective evaluation of oral odor. For a subjective assessment of an unpleasant odor in the mouth each patient underwent organoleptic examination before tonsillectomy. Patients were neither allowed to take antibiotics for a period of three weeks, consume garlic, onion and spicy dishes for 24 hours nor apply perfume. Within 12 hours, it was also recommended to avoid food intake and drinks, brush teeth, use breath fresheners and smoke cigarettes. During the study, the smell of air exhaled through the mouth was compared with the smell of air exhaled through the nose (when mouth closed). The patient then exhaled into a transparent tube 10 cm long and 2.5 mm in diameter. In order to assess the intensity of the unpleasant smell from the mouth a six-grade Rosenberg scale was used which is the most widely used scale of halitosis [4]. The odor was classified between 0 and 5, where: 0 - absence of odor, 1 - barely noticeable odor, 2 - slight malodor, 3 - moderate malodor, 4 - strong malodor and 5 - severe malodor.

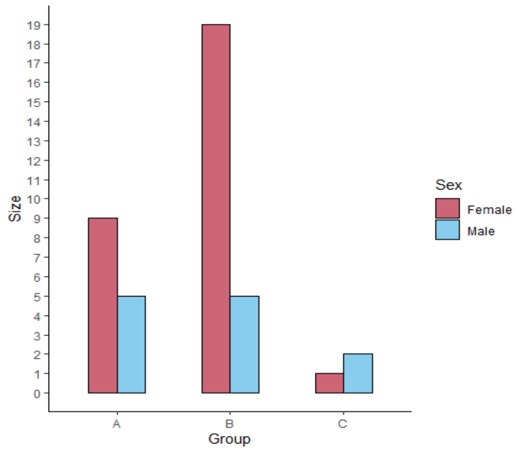
For objective assessment of halitosis each patient underwent a Man-hal-II Halimeter test. This sulfide monitor measures volatile sulfur compounds such as hydrogen sulfide, methyl mercaptan and dimethyl sulfide which are contributing factors for halitosis. Each sample was taken after 3 minute re-stabilization period when patients were breathing through the nose with mouths kept closed. During the test the device's straw was inserted into the subject's mouth and one was asked to exhale briefly for 30 seconds. The procedure was repeated in three trials for each patient and the peaks' values were recorded in parts per billion (ppb) sulfide equivalents [5]. In order to obtain reliable results, patients could not eat, drink, smoke, chew gum, brush their teeth, use the brush, fresheners or oral hygiene products for 12 hours prior to the halimetric test. It was also recommended not to use cosmetics such as perfumes, aftershave or lipstick.

Organoleptic and halimetric examination was repeated for each patient, at least 2-3 months after removal of the palatine tonsils and completion of the healing process.

The value of the average of the three measurements of VSC was used for further statistical analysis. Depending on the chosen criterion of division, two or three groups of patients have been distinguished. First one considered 100 ppb as threshold dividing group into normal (when below) and abnormal (when above the threshold). Second, more detailed criterion included intermediate states as follows: the range up to 100 ppb considered as correct, 100 - 180 ppb as a light form of disease and above 250 ppb as a severe one [6]. However, the study group did not include patients belonging to the latter (Fig. 2.1).



(a) the ratio of groups with different forms of the disease



(b) the gender distribution within groups A: $<$ 100 ppb, B: 100 - 180 ppb, C: $>$ 180 ppb VSC

Figure 2.1: Groups of patients due to the level of VSC before tonsillectomy.

Principal Component Analysis (PCA) was used in order to validate the above classification. This method enables finding the directions of maximum variance in high-dimensional data and project it onto a smaller dimensional subspace while retaining most of the information. In the following work, a self-written script in the R programming language was used to conduct the analysis. PCA was performed using *FactoMineR* which is an R package dedicated to multivariate Exploratory Data Analysis, as well as *factoextra* to extract and visualize the results. There are also other, built-in R functions avaliable - *princomp()* and *prcomp()*. The preceding one uses the spectral decomposition approach, while the functions *prcomp()* and *FactoMineR::PCA()* use the singular value decomposition.

Comparisons between groups and within groups were made using appropriate statistical tests, after checking the assumptions authorizing to carry them out. The gender distribution was included in each group. Considered level of significance for all of the statistical tests was 5% (α = 0,05).

Chapter 3

Results

The first stage of the analysis involved examining the population of forty-one patients qualified for the study. The halitometry cut-off value divided patients into four groups: Group A - normal (below 100 ppb), Group B - minor halitosis (100 - 180 ppb), Group C - abnormal halitometry (180 - 250 ppb) and Group D - severe or chronic form of halitosis (above 250 ppb). None of the patients showed symptoms, classifying him or her to the latter, therefore Group D is not presented in the results. The first section summarizes the distribution of age, living place, education, profession and halitometry results for the observed data.

The second section presents the results of Principal Component Analysis examining the adequacy of the assignment of patients to groups.

Due to the relatively small research sample's size, the statistical inference procedure was repeated in conjunction with bootstrap technique. The results were posted at the end of this chapter.

3.1 Original sample results

Table 3.1 summarizes the comparison of features: age, living place, profession and education between men and women in the original data. Student's t test (*stats::t.test*) was used to test the significance of differences in means of age after checking the assumption of normality and homogeneity of variance. Categorical data - living place, profession, education - were tested with χ^2 test of independence. No statistically significant difference was observed in terms of analyzed features ($p > 0.05$).

	Female n = 29	Male n = 12	Total n = 41	Comparison (*p-value*)
Age				
mean	26.69	27.83	27.02	
sd	8.49	6.32	7.86	0.6391
me	24.00	27.00	25.00	
spread	18.00-56.00	19.00-36.00	18.00-56.00	
Living place				
village	5 (17.24)	0 (0.00)	5 (12.20)	
town	2 (6.90)	1 (8.33)	3 (17.85)	0.3078
city	22 (75.86)	11 (91.67)	33 (80.49)	
Profession				
school-age student	2 (6.90)	2 (16.67)	4 (9.76)	
student	6 (20.69)	1 (8.33)	7 (17.10)	0.1834
blue-collar worker	3 (10.35)	4 (33.33)	7 (17.10)	
white-collar worker	18 (62.07)	5 (41.67)	23 (56.10)	
Education				
elementary	1 (3.45)	2 (16.70)	3 (7.32)	
secondary	13 (44.80)	5 (41.70)	18 (43.90)	0.2699
higher	15 (51.70)	5 (41.70)	20 (48.80)	

Table 3.1: General statistics for studied sample and comparisons between female and male.

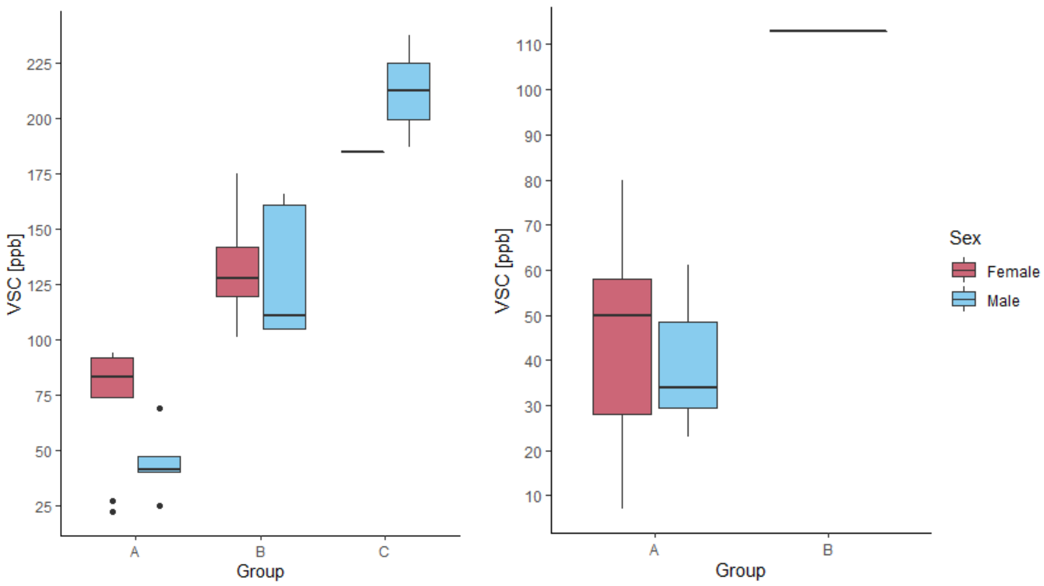
The comparisons of VSC levels between men and women were made with Wilcoxon test (*stats::wilcox.test*) due to heterogeneity of variance between groups. According to *p-values* 0.5378 and 0.8185, the null hypothesis saying that the mean VSC levels for women and men were equal both before and after tonsillectomy failed to reject (Tab. 3.2).

Levels of VSC before and after tonsillectomy (for whole sample) were compared using a Paired t-test (*stats::t.test*) after bringing the prior to a normal distribution. Originally, the average level of VSC had right-skewed distribution with coefficient = 0.787. Taking a square root of this variable turned out to be enought to decrease the coefficient of skewness to 0.018 which is close to zero and meet the assumptions of normality in Shapiro test [7]. According to *p-value* < 0.05, the null hypothesis saying that the average VSC levels in both groups are equal was rejected, thus the distributions of those variables differed significantly. The level of VSC in the exhaled air was effectively lowered as a result of tonsillectomy.

	Female n = 29	**Male** n = 12	**Total** n = 41	**Comparison** (*p-value*)
Before				
mean	115.03	107.92	113.00	
sd	37.89	67.81	47.76	0.5378
me	120.00	105.00	117.00	
spread	22.00-186.00	25.00-238.00	22.00-238.00	
After				
mean	43.52	45.08	43.98	
sd	19.20	24.81	20.66	0.8185
me	50.00	36.00	40.00	
spread	7.00-80.00	23.00-113.00	7.00-113.00	

Table 3.2: Halitometry results: levels of VSC before and after tonsillectomy.

There was only one patient left with minor halitosis after the removal of palatine tonsils - a 29-year old man with 113 ppb of VSC. In general, the avarage level of VSC dropped by 106.50 ppb compared to pre-tonsillectomy measurements (Fig. 3.1).



(a) before tonsillectomy (b) after tonsillectomy

Figure 3.1: Levels of VSC in groups with split on gender. Group A: < 100 ppb, Group B: 100 - 180 ppb, Group C: > 180 ppb VSC.

3.2 PCA results

While Principal Component Analysis can be only performed on numeric variables, a subset of data was created. The analysis concerned the following features: age, three VSC measurements before and three measurements after tonsillectomy, two averages of these measurements, height and weight (41 individuals and 11 variables in total).

In order to determine the minimum number of principal components (PCs) that account for most of the variation in analyzed data two methods were used. The first one was based on the proportion of variance that the components explain. The acceptance level of variance was set to 80%, thus PC3 might be taken as the cut-off point (Tab. 3.3).

PC	**eigenvalue**	**pct of variance**	**cumulative pct of variance**
comp 1	5.3278	48.434	48.4342
comp 2	2.0581	18.7102	67.1444
comp 3	1.7014	15.4671	82.6115
comp 4	0.9906	9.0054	91.6170
comp 5	0.3051	2.7735	94.3905
comp 6	0.2083	1.8938	96.2843
comp 7	0.1968	1.7896	98.0738
comp 8	0.0944	0.8585	98.9323
comp 9	0.0705	0.6406	99.5729
comp 10	0.0467	0.4245	99.9974
comp 11	0.0003	0.0026	100.0000

Table 3.3: The eigenvalues and percentages of explained variance for corresponding principal components.

An alternative method of estimating the number of principal components is a scree plot (Fig. 3.2), which sorts the eigenvalues descending order. The number of PCs is selected at the point where there is a relative decrease in the amount of variance explained by given component. In that case, both methods indicated the same cut-off point which was the third principal component (PC3) meaning that first three principal components explain 82.61% of variation in the data.

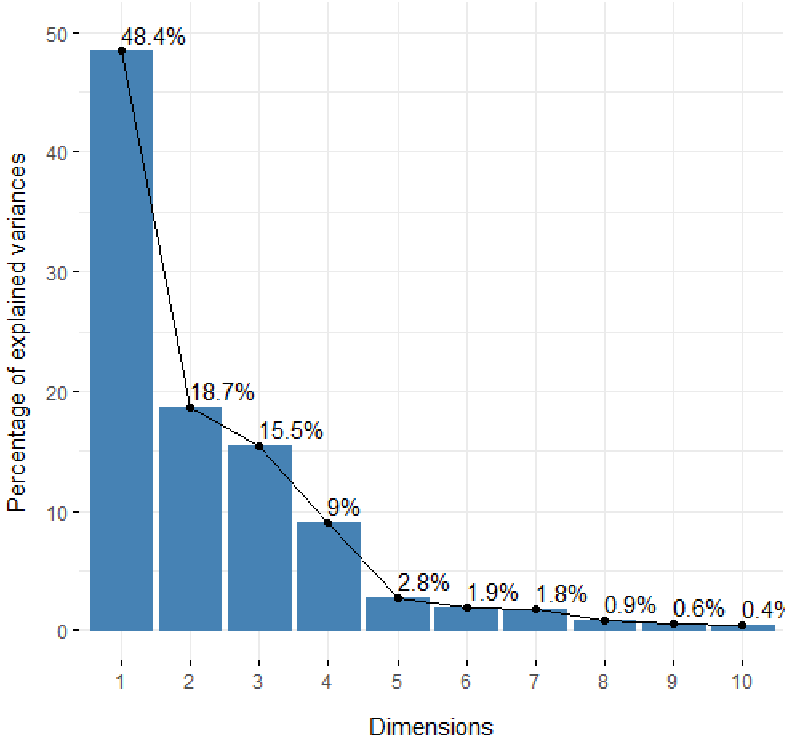


Figure 3.2: Scree plot.

Examining the magnitude and direction of the coefficients for the original variables allows to interpret each principal component. In these results, first and second principal component had positive associations with the average levels of VSC (before and after tonsillectomy), so this component primarily measured surgery effectiveness. The second component had positive associations with individual levels of VSC, so this component primarily measured repeatability of results. The third component had large positive associations with height and weight, so this component primarily measured the patient's body mass index.

A PCA biplot (Fig. 3.3) visualize both principal component scores of samples (here already grouped into three) and loadings of variables (so called loading plot). In general, a biplot is a projection of high dimension space onto a two-dimension plane. Each observation is displayed in the figure, keeping their original axes [8]. The further away vectors (variables) are from the origin, the more influence they have on that principal component. From loading plots we can notice how well are the variables correlated with each other: a small angle implies positive correlation, a large one suggests negative correlation. A right angle indicates no correlation between two features.

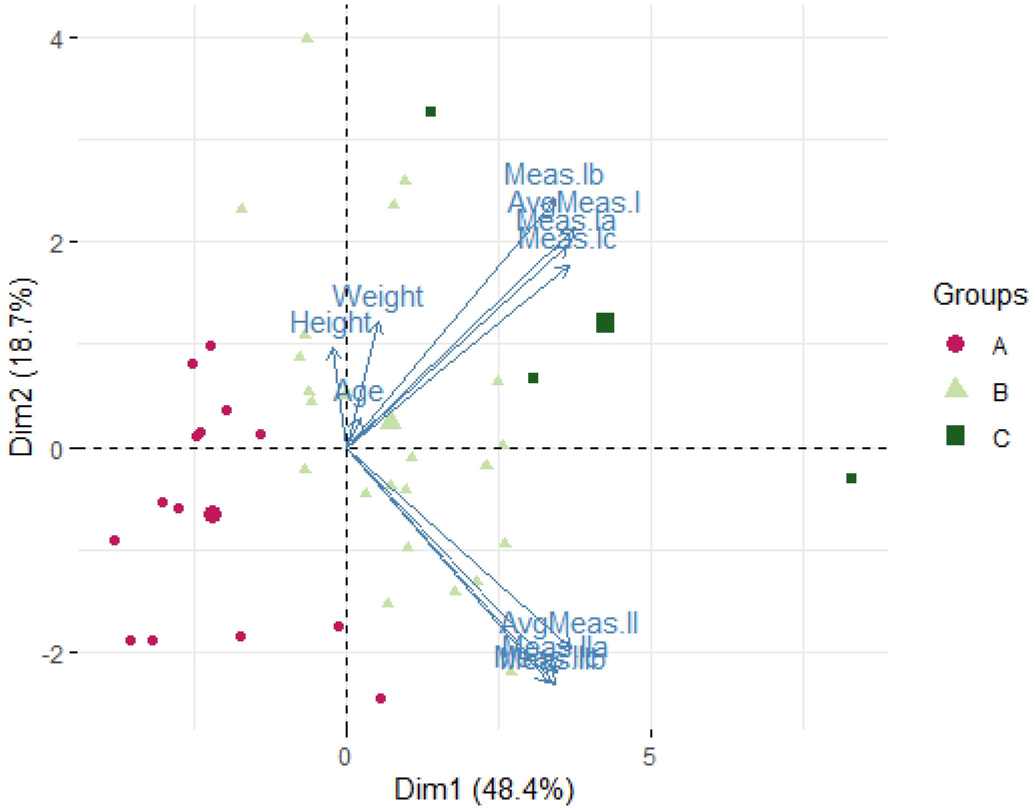


Figure 3.3: A biplot showing both principal component scores for individuals categorized into three groups and loadings of variables

3.3 Bootstrap results

Based on the replications of the original sample, distributions' parameters of features describing the studied group of patients were estimated. The number of replications used in the bootstrap method was equal 1000 [9].

Table 3.4 summarizes the results of the procedure. The χ^2 goodness of fit test was performed for the categorical data, while the "age" variable was compared with the Wilcoxon test due to the lack of normal distribution.

At 0.05 significance level, the null hypothesis saying that the patients (either within one sex or between both sexes) are equally distributed across education, profession and living place, was rejected. Also, the null hypothesis saying that the age of female and male come from identical distribution was not accepted.

	Female B = 1000	**Male** B = 1000	**Total** B = 1000	**Comparison** (*p-value*)
Age				
mean	26.68	27.87	27.02	
sd	8.11	6.00	7.63	< 0.05
me	24.44	27.83	25.09	
spread	18.39-51.86	19.42-35.56	18.35-51.87	
Living place				
village	161 (16.10)	0 (0.00)	113 (11.30)	
town	68 (6.80)	90 (9.00)	79 (7.90)	< 0.05
city	771 (77.10)	910 (91.00)	808 (80.80)	
Profession				
school-age student	61 (6.10)	166 (16.60)	92 (9.20)	
student	194 (19.40)	104 (10.40)	181 (18.10)	< 0.05
blue-collar worker	110 (11.00)	313 (31.30)	147 (14.70)	
white-collar worker	635 (63.50)	417 (41.70)	580 (58.00)	
Education				
elementary	27 (2.70)	171 (17.10)	72 (7.20)	
secondary	432 (43.20)	417 (41.70)	444 (44.40)	< 0.05
higher	541 (54.10)	412 (41.20)	484 (48.40)	

Table 3.4: General statistics for bootstrap population.

Subsequently, the distribution parameters were estimated for the average VSC measurements taken both before and after the tonsillectomy. All samples differed in terms of distribution; therefore Wilcoxon test was used for comparisons between sexes and the halitometric effects of the surgery (Tab. 3.5).

	Female n = 29	Male n = 12	Total n = 41	Comparison (*p-value*)
Before				
mean	114.71	108.01	113.06	
sd	36.89	64.08	46.68	4.340e-13
me	118.86	100.19	115.40	
spread	30.02-179.59	30.74-217.15	24.17-217.99	
After				
mean	43.40	45.17	43.69	
sd	18.82	22.67	20.07	0.052
me	45.21	38.35	42.07	
spread	9.93-75.27	24.15-95.53	9.96-98.21	

Table 3.5: Halitometry results for bootstrap population: levels of VSC before and after tonsillectomy.

The accuracies for the estimates of selected mean values are presented in table below (Tab. 3.6).

	Mean	Standard error	Bias	Confidence intervals
Female				
Age	26.68	1.490	0.004	(23.965; 29.759)
VSC before	114.71	7.320	0.333	(98.894; 128.693)
VSC after	43.40	3.428	0.113	(36.584; 50.277)
Male				
Age	27.87	1.777	0.402	(24.417; 31.417)
VSC before	108.01	18.930	0.096	(73.147; 145.096)
VSC after	45.17	6.804	0.092	(33.417; 59.417)
Total				
Age	27.02	1.185	0.003	(24.877; 29.538)
VSC before	113.06	7.462	0.113	(98.830; 128.722)
VSC after	43.69	3.136	0.288	(37.790; 49.881)

Table 3.6: The accuracy of the mean estimates for age and halitometry results.

The concentration of VSC after removal of palatine tonsils decreased on average by 68.976 ppb, falling within confidence intervals (-84.591; -52.434). The demonstrated decrease of 61% was statistically significant ($p < 0.05$) due to Wilcoxon signed rank test with continuity correction outcome.

Tutaj dołożylabym jeszcze histogramy porównujące rozkłady sprzed bootstrapa i po bootstrapie.

Chapter 4

Discussion

Halitosis affects about 25–30% of world's population. Tonsillectomy is one of the most frequently performed surgeries in the world in the practice of laryngologists [10]. About 400,000 are performed annually in the US these treatments. The study aims was to investigate connection between chronic tonsilitis, halitosis and other potential indices. Therefore, a restrictive selection of patients was necessary and the final sample size was equal 41.

The results of the empirical study seem to confirm the hypothesis saying that halitosis can be an indication for tonsillectomy and that this problem affects a population regardless of gender, age or living status.

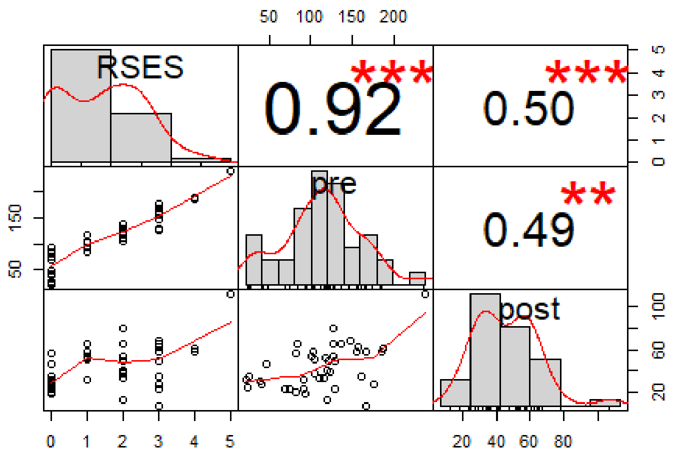


Figure 4.1: Correlation between Rosenberg scores and halitometry outcomes. RSES - Rosenberg scale, pre - VSC level before tonsillectomy, post - VSC level after tonsillectomy.

A positive correlation was observed between Rosenberg's score and VSC concentration both before and after surgery. According to the Pearson coefficients obtained, we can conclude that the effectiveness of the organoleptic method depends on the level of VSC (Fig. 4.1). The more intense the problems with the unpleasant odor, the better the coverage of organoleptic and halimetric test results. Although the organoleptic evalutaion with six-level scale seem to give better results when compare to three-level scale [11], the methods should be used complementarily, not separately.

Based on the results obtained, we can also state clearly that the presented bootstrap method gives the possibility to obtain satisfactory results in the analysis of medical data. The applied resampling procedure gives the possibility of statistical inference without making assumptions about the analytical form of the distribution of the examined feature. Furthermore, this method allows to obtain empirical distributions of very large numbers with small numbers of research samples. In the case of statistical inference with respect to medical data, this is very important, because in many cases, obtaining the right amount of data necessary to confirm the effectiveness of therapy or inferring the frequency of a rare disease entity, is difficult.

Davison and Kuonen, believe that considering the bootstrap method as an alternative for classical statistics is misguided. The authors warn before using such a powerful tools as resampling technics without deep understanding why it works. They emphasize the importance of monitoring the output in order to ensure that the results make sense and understand how the output will be used. Therefore, bootstrap should always be preceded by population, samples parameters, and fundamental statistical notions examinations [12].

In order to reduce the number of dimensions of data as well as to check the relevance of thresholds dividing into groups of observations (A, B and C), a Principal Component Analysis was performed. PCA analysis reports as many PCs as the number of variables included in the analyses. However, we can observe that the first three PCs are responsible for the majority of the variance in the data which is 82.61%. The group allocation also seems to be consistent, as can be seen from the biplot in which group B is the most numerous and group C is the least numerous.

Results of PCA for other data sets are often be more challenging. Bigger amount of predictors included in the PCA, might result in bigger number of PCs selected, and thus increase the model size. In general, it aims to select the principal components with the largest variance in order to keep only those explaining enough variance to make epidemiological and clinical sense [8].

To sum up, described techniques may complement current nonparametric statistical methods and should, where appropriate, be included in data processing methodologies. The methods have proved to be useful especially in the analysis of medical data.

Bibliography

[1] Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap*, Chapman & Hall.

[2] Bochnia M, Dziewiszek W, Rostkowska-Nadolska (2005) *Tonsylektomia w XXI wieku*, Family Medicine & Primary Care Review, 7, 874–881.

[3] Kotti AB, Subramanyam RV (2015) *Oral malodor: A review of etiology and pathogenesis*, Journal of Dr. NTR University of Health Sciences, 4(1) 1-7.

[4] Rosenberg M, McCulloch CAG (1992) *Measurement of Oral Malodor: Current Methods and Future Prospects*, Journal of Periodontology, 63, 776–782.

[5] Alasqah M, Khan S, Elqomsan MA, Gufran K, Kola Z, Bin Hamza MO (2016) *Assessment of halitosis using the organoleptic method and volatile sulfur compounds monitoring*, Journal of Dental Research and Review 3(3), 94-98.

[6] Lee PP, Mak WY, Newsome P (2004) *The aetiology and treatment of oral halitosis: an update*, Hong Kong Medical Journal 10, 414–418.

[7] Wilcox R, Peterson TJ, McNitt-Gray JL (2018) *Data Analyses When Sample Sizes Are Small: Modern Advances for Dealing With Outliers, Skewed Distributions, and Heteroscedasticity*, J Appl Biomech. 34(4), 258-261.

[8] Zhang Z, Castello A (2017)*Principal components analysis in clinical studies*, Ann Transl Med, 5(17), 351.

[9] Davison AC, Hinkley DV (1997) *Bootstrap Methods and Their Application*, Cambridge, UK: Cambridge University Press.

[10] Blustoniec C (1997) *Status of tonsillectomy and adenoidectomy*, Laryngoscope, 8, 1233.

[11] de Mesquita-Guimaraes, et al. (2017) *Organoleptic reproducibility for halitosis*, European Journal of General Dentistry, 6(1), 9-13.

[12] Davison AC, Kuonen D (2003) *An Introduction to the Bootstrap with Applications in R*, Statistical Computing & Statistical Graphics Newsletter, 13(1), 6-11.