ZESZYTY NAUKOWE UNIWERSYTETU RZESZOWSKIEGO

SERIA FILOLOGICZNA

ZESZYT 115/2020 STUDIA ANGLICA RESOVIENSIA 17

DOI: 10.15584/sar.2020.17.3

Ewa KUSZ

University of Rzeszow, Poland

ekusz@ur.edu.pl

STATISTICS FOR LINGUISTS REVISITED: THE REVIEW OF SOME BASIC STATISTICAL TOOLS IN LINGUISTIC RESEARCH AND DATA ANALYSIS

Abstract: The major aim of this paper is to emphasise the importance of implementing statistical tools in the field of linguistic research, as well as to acquaint the reader with the basic statistical methods that can be used while conducting linguistic studies. The article introduces the idea of five steps in data analysis that any researcher of applied linguistics can take in order to carry out relevant studies. The steps include choosing statistical programmes, eliciting data, selecting some visual methods and applying normality tests, as well as choosing applicable parametric or non-parametric tests, all of which requires appropriate planning, designing, analysing and interpreting data. The theoretical part is an interlude to the practical realisation of the above-mentioned five steps, which is based on the part of linguistic research conducted on the students of English Philology. The major purpose of it was to prove (or refute) that there is a positive correlation between participants' level of musical intelligence and their L2 pronunciation skills. The practical use of statistical methods enables the readers to familiarise themselves with one of the patterns of statistical analysis in the field of applied linguistics.

Key words: statistics, data analysis, quantitative methods, parametric and non-parametric tests

Introduction

There are numerous studies which prove that statistical literacy in the field of applied linguistics, including Second Language Acquisition (SLA) require more attention and linguists who attempt to conduct research often struggle with the use of statistical tools. On the one hand, statistics is in most cases linked to science and it attempts to present more systematic data, whereas linguistics, on the other, is perceived as the field within which quantitative analyses are implemented. It should not come as a surprise that quantitative analyses appear more frequent in linguistics modern literature, yet according to Eddington

(2015), there is still a large number of linguists who do not possess sufficient knowledge and experience in statistics due to their lack of training. The issue that appears among linguists is that they find it difficult to adjust existing guidelines to linguistic reality. Hence, it is heartening that there is a growing number of works on statistics written for the field of applied linguistics (Baayen 2008, Rasinger 2010, Larson-Hall 2010, Cantos Gómez 2013, Gries 2013, Eddington 2015, etc.)

The following paper aims to prove that quantitative methods should not be neglected in the field of applied linguistics. Some selected quantitative methods will be presented and subsequently they will be based on the part of research conducted between 2016-2017 on the group of students of English Studies from the University of Rzeszow in Poland. The statistical analyses used to present research results can be a guide for other linguists who struggle with the issue of more sophisticated and advanced quantitative methods.

From the linguists perspective, the most important factor about statistics is to know how to apply statistical tools and methods correctly to their studies. This paper is to address this issue. It includes linguistic data sets which on the one hand may suggest ways in which some other (similar or not) linguistic questions may be answered using the same statistical research design methodology. On the other, due to an overwhelming number of statistical methods, only the most common used statistical analyses are included.

Literature Review

There is a growing number of academics who claim that statistical analysis in linguistics, including SLA (Second Language Acquisition) studies are of crucial importance (e.g. Lazaraton (2000; 2005), Loewen and Grass (2009)). According to Loewen and Gass (2009:181), ‘SLA is not an innovator but an increasingly knowledgeable borrower and adapter of statistical procedures’. Moreover, Loewen et al. (2014) assessed statistical knowledge in the field of applied linguistics and SLA. The major claim of their work is that ‘statistics [is] an important and necessary component of applied linguistics/SLA research, although more so by quantitatively oriented participants than by qualitatively oriented ones’ (2014:20). They also emphasise the fact that researchers need ‘detailed guidelines […] and examples of good practice’ (2014:22).

The study of the quality of quantitative research in the field of applied linguistics and SLA conducted by Plonsky and Grass (2011) has shown that among 174 interaction studies published between 1981 and 2009, there is some improvement research methodology and presentation of outcomes. Yet, there is still a strong need to enhance the quality of using statistical methods in this area.

The discrepancy in quality between research design in SLA studies and other fields (e.g. psychology, sociology, etc.) may be the consequence of the lack of training and acquiring statistical knowledge. Loewen et al. (2014:22) state that 'applied linguists are at the lower end of statistical training'. This statement is followed by research studies conducted by Aiken et al. (2008), Leech and Goodwin (2008), Brown and Bailey (2008), Plonsky and Gass (2011). For instance, Aiken et al. (2008) have shown in their research that the most traditional techniques (e.g. ANOVA or multiple regression) are at the high level among participants (graduate students) of research, however, more complex methods as structural equation modelling and logistic regression are not so frequently used. Woods, Fletcher and Hughes (1986:1-2) also highlight the importance of statistical analyses in the field of linguistics. '[…] [S]tatistics allows us to summarise complex numerical data and then, if desired, to draw interferences from them. […] [It] serves the purpose of reducing complex data to manageable proportions'.

Methodology

The following section attempts to present the most important stages that the researcher can follow in the process of statistical data analysis. The stages are divided into several steps, however, it should be emphasised that this path is only an example which may be helpful to conduct analytical part of any linguistic research. It is researcher's choice whether to follow it or not, yet, it could be the very start of a journey through professional statistical analysis.

Step one: choose the statistical programme

The following section is a short description of statistical programmes which can be used in the field of applied linguistics.

Before a researcher is able to generate descriptive statistics, they must first choose the statistical programme which will be the most suitable for the study. The data are obtained from the computer software, within which the most popular one is **SPSS** (Statistical Package for the Social Sciences). It was first launched in 1968 and subsequently acquired by IBM in 2009 (thus its official name is IBM SPSS). The major function of SPSS is to analyse and edit all sorts of data. It provides descriptive statistics, such as mean, median, mode, range, standard deviation and standard error. It contains all basic statistical tests as well as multivariate analyses, including correlation measures (Pearson and Spearman correlation coefficients), regression, parametric (t-test, analysis of variance) non-parametric tests (such as Mann-Whitney, Wilcoxon signed-ranks, sign tests), factor analysis, etc. Moreover, SPSS is equipped with a graphical user interface (GUI), which makes it user-friendly and intuitive.

R statistical programming language is a free statistical software developed in the University of Auckland, New Zealand. It is not only for data analysis, but also for data visualisation that is more powerful than SPSS, as it generates better graphics, however, it does not have a user-friendly interface (instead of a graphical user, it has a command line interface). Thus, in comparison to SPSS, it is considered more complicated for non-statisticians and beginners.

Minitab also inputs statistical data and it might be used for linguists to elicit data for their studies. It was developed at the Pennsylvania State University in 1972, and it mostly focuses on the analysis of statistical data, as well as the interpretation of the obtained results. Interestingly, this statistical software was originally intended as a tool for teaching statistics and for users with little statistical experience. Additionally, due to its user-friendly interface, as well as free online teaching resources, it is frequently used by the students of various faculties, including linguistics. Moreover, this software is still the leader among tools used for educational purposes and used by over 4,000 colleges and universities around the world.

Another popular statistical software is **SAS** (Statistical Analysis System), developed by SAS Institute. It is mostly used for advanced analytical studies, multivariate analysis, business intelligence, data management and predictive analytics, thus it is less frequently used for linguistic studies, however, still possible to be used in this area.

To conclude, it is a researcher's choice which statistical programme will be used to complete the research analysis. Each programme can be adjusted to the field of applied linguistics. Once the programme is selected, the next step will be to elicit data.

Step two: elicit data

The major aim of this section is to present the second statistical step which might be taken by the linguists to introduce the results of research. The variables which are usually taken into account in the scientific research (including the field of linguistics) can be grouped into three categories (measures), depending on their utility. It is possible to distinguish measures of central tendency, variability and shape. Measures of central tendency (or central location) include such variables as average (mean), median, lower and upper quartile, maximum and minimum value, and range (which might be also counted as a measure of variability). Their main aim is to identify a typical value for a probability distribution. Among all of the above-mentioned variables, there are lower and upper quartiles which require some explanation, as they are the least common values used in statistical analyses. Lower quartile value can be defined as the median of the lower half of the data, whereas the upper quartile value is the median of the upper half of the dataset.

Measures of variability include variance, standard deviation and coefficient of variation. They describe the data spread. The variance tells the researcher how spread out the data are, whereas standard deviation indicates how far, on average, each value is from the mean. Finally, coefficient of variation (also known as CV or RSD – relative standard deviation) measures the dispersion of frequency distribution.

Measures of shape are skewness, kurtosis, as well as standardised skewness and standardised kurtosis. All of the above identify the distribution of the data within a dataset. Skewness indicates departure from horizontal symmetry of the probability distribution of a real-valued random variable about its mean. It shows which values are more frequent around the high or low ends of the distribution. On the basis of that, it is possible to distinguish positively or negatively skewed distribution. The first type appears when the right part of a tail of the histogram is longer than the left side. Accordingly, left-side tail longer than the right sight indicates negatively skewed distribution. Kurtosis also refers to the shape of a random variable's probability distribution, however, the main difference between this and former is that it implies the 'pointedness' of a peak in the distribution centre. In other words, skewness indicates the degree of symmetry, whereas kurtosis shows the degree of 'peakedness'. Last but not least, standardised skewness, as well as standardised kurtosis determine if the sample set of data is normally distributed. The values outside the range of -2 to +2 imply departures from normal distribution, which refutes the possibility to use any statistical tests in terms of standard deviation.

On the basis of the above, the last two variables, namely standardised skewness and standardised kurtosis enable the researcher to provide information whether the set of data is normally distributed or not. This is extremely important in terms of the use of statistical tests (parametric or non-parametric), which will be carefully described in the next subsection.

Step three: choose visual methods and apply normality tests

Graphical (or visual) method is one of the options that can be taken into consideration while doing statistical analysis. It may be used as a first, initial step to assess normality, however, it must be highlighted that this approach is not as reliable as regular statistical tests and it does not guarantee interpretation that will be free from error. According to Chambers et al. (1983:56), '[g]raphical methods provide powerful diagnostic tools for confirming assumptions, or […] for suggesting corrective actions.' Moreover, the visual presentation of research data enable readers (and researchers) to judge its distribution by themselves. There is a large number of variants of probability plots, including histograms, box plots, stem-and-leaf plots, percent-percent plots (P-P), quantile-quantile plots (Q-Q), etc. However, due to space limitations, only the most common plots, namely histograms and box plots will be carefully described in terms of applied linguistics research.

Histogram

Due to its simplicity, histogram is considered to be the easiest and the most common plot to present and interpret statistical data. It enables researchers to inspect the data not only for its distribution but also skewness, kurtosis or outliers. The ideal bell-shaped curve of presented data indicates its normality, however, the lack of normal distribution can be also easily noticed with the use of this visual method. The example below presents non-symmetric data of linguistic research[1], in which leptokurtosis occurs (this can be seen in the shape of distribution, which is more peaked). The data are not normally distributed, although most values centre around the average score (which is almost 9).

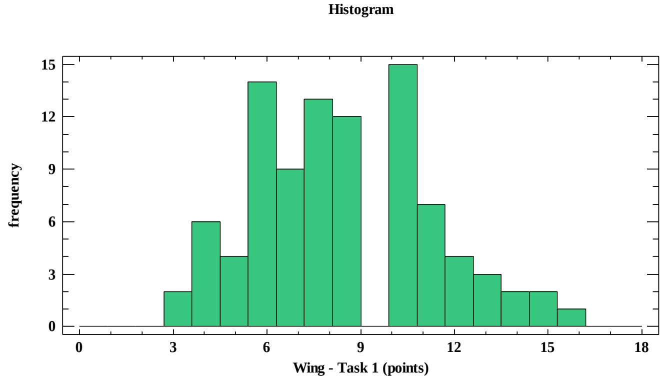

Figure 1. An example of histogram.

Box-and-Whisker plot

Also known as a Box Plot is another example of presenting the distribution of dataset, which was first introduced by John W. Turkey. Each box consists of five-

[1] Please note that the detailed description of the research can be found in Kusz, E. 2019. *Musical intelligence and its impact on English pronunciation skills in the process of second language acquisition*. Frankfurt am Mann, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang. DOI: https://doi.org/10.3726/b16166

number summaries, within which there are five values: the most extreme (minimum and maximum), the upper and lower quartiles, and the median. The example below shows how this type of visual method attempts to plot data in a box.

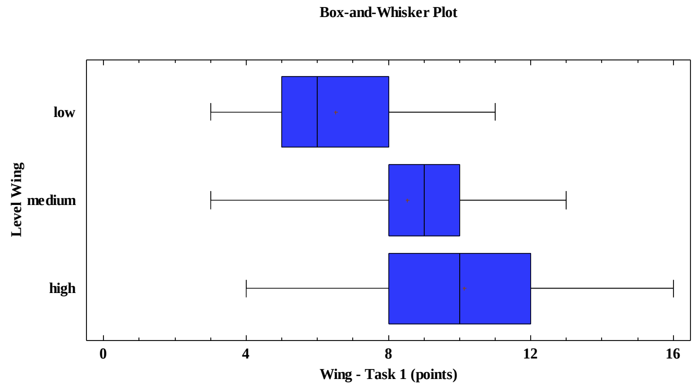

Figure 2. An example of box-and-whisker plot.

Step four: choose relevant statistical tests

In order to ascertain the validity of the obtained data elicited from the first and the second steps of statistical analysis, further statistical tests ought to be conducted. The major aim of it is to check the compliance with the distribution of dataset. For confirmation, tests of normality should be implemented. There are several main tests of normality, including Kolmogorov-Smirnov (K-S) test, Shapiro-Wilk test, Anderson-Darling test, Cramer-von Mises test, D'Agostino skewness test, Anscombe-Glynn kurtosis test, D'Agostino-Pearson omnibus test, and the Jarque-Bera test within which Kolmogorov-Smirnov test, Shapiro-Wilk test and Jarque-Bera are the most popular among researchers in the field of linguistics (Ghasemi and Zahediasl 2012). All of the above compare the sample of data to the associated and normally distributed values with the same mean and standard deviation. Hence, the null hypothesis is always that 'sample distribution is normal'. For the purpose of this paper, only two (Kolmogorov-Smirnov and Shapiro-Wilk) tests will be described, as the practice section focuses on these two examples.

Kolmogorov-Smirnov test

This test of normality is used to determine whether the distribution of the scores of data is close to normal. It was first introduced by Kolmogorov (1933) and subsequently modified and presented by Smirnov (1948). Despite its popularity, the Kolmogorov-Smirnov test has its limitations, namely it is highly sensitive to extreme values which results in lack of recommendation, regardless of sample size.

Shapiro-Wilk test

According to Mendes and Pala (2003), this test of normality is considered to be one of the most preferable and common ones due to its good properties of power. It was derived by Shapiro and Wilk (1965), however, it was originally restricted to the sample size of less than 50. The null hypothesis is rejected when the p-value equals (or is less than) 0.05. The test enables to state with 95% confidence that the data are (or are not) normally distributed and it provides better power than the above-mentioned Kolmogorov-Smirnov test. Additionally, there is a large number of researchers who recommend this type of normality test as the best option for statistical analysis (e.g. Thode 2002).

Step five: choose parametric or non-parametric statistical test

The previous steps allow the researcher to decide which type of test, parametric or non-parametric, should be applied regarding normal distribution of data set. If sample groups are normally distributed the parametric test should be applied. Following that, non-normally distributed data require application of non-parametric tests.

Kruskal-Wallis test

This type of test is also known as the 'one-way ANOVA on ranks', as it attempts to ascertain the validity of data which are not normally distributed, whereas ANOVA goes with parametric samples. According to McKnight and Najab (2010:4) it is a non-parametric test of statistics which major task is to assess the differences between three or more independently sampled groups on a single, non-normally distributed continuous variable. The Kruskal-Wallis test is also an extension of the two-group Mann-Whitney U test.

Bartlett's test

To decide which test, parametric or non-parametric, should be used in the statistical analysis, it is of crucial importance to check the equality (or in other words homogeneity) of variances. This will also allow to verify the null hypothesis, which assumes that each sample of groups is of the same variance. Unequal groups sizes indicate that the homogeneity is violated. In that case, Bartlett's test must be applied. This test introduced by Snedecor and Cochran (1989) checks the validity of the assumption of equal variances. Due to its sensitivity to departures from normality, it is highly recommended to implement this type of test if one attempts to test non-normality of dataset (especially when samples do not come from normal distributions).

ANOVA

ANOVA stands for analysis of variance and it is a parametric test which determines if there are any statistically significant differences among the means of independent samples. This statistical technique was introduced by Ronald Fisher – a statistician and evolutionary biologist in 1925. There are several requirements that each dataset must meet in order to implement the ANOVA test. First, each group sample must be normally distributed. Secondly, the variance of data must be homogeneous. Third, the data must be numeric, and finally the results obtained from the study cannot influence each other, which means that they are sampled randomly and independently.

Multiple range tests

Before multiple range tests (also known as post hoc tests or multiple comparison analysis tests) are carefully described, it is of utmost importance to highlight the fact that they support ANOVA to fully analyse and understand group differences. Thus, researchers who decide to use this parametric test, must also conduct tests that enable them to notice the differences between particular control groups or subgroups ("pair-wise" differences). Multiple range tests are used for comparing means in an analysis of variance in order to determine which means of the collected dataset are significantly different. The most popular multiple comparison analysis tests are: Tukey, Newman-Keuls, Scheffée, Bonferroni and Dunnett (McHugh 2011:203). Due to the space limitations, only the Tukey method will be described, as it is the test used in the practical part of this paper.

Tukey's multiple comparison analysis method (also known as Tukey's honestly significant difference test or Tukey's HSD) is preferred for unequal group sizes and it is applied to test experimental group against each control group. In order to do the correct analysis, first ANOVA must be applied to evaluate whether there are any bases to claim that the means of the sample research groups differ. Also, both ANOVA and Tukey's HSD test assume that the data from the different groups come from the populations that are normally distributed and the standard deviation is the same for each group (Olleveant 1999:304). Tukey's multiple comparison analysis method allows the researcher to determine which means among various sets of means differ from each other. Its major aim is to determine if group differences are statistically significant. Thus, it starts with testing the largest means pair-wise difference. If the difference is significant, the next pair is tested. It is continued until the obtained value is not significant.

Results

Steps in practice: an example of statistical analysis on the basis of linguistics research

The following example of statistical analysis of the linguistic research may be helpful in the process of analysis of any other linguistic data. The procedure is based on the steps presented in the previous sections. Thus, it is of utmost importance to begin with the review of standard statistical measures, including average, median, standardised kurtosis and standardised skewness. The last two values enable researchers to verify if the sample is normally distributed. For normally distributed data parametric tests will be applied (ANOVA), whereas for those who are found not normally distributed the non-parametric tests must be used (Kruskal-Wallis test).

Among 94 students of English Studies three groups (high, medium, low) were distinguished, within which high group stands for the best scores achieved in Wing's musical intelligence test and low for the worst achievements. Their musical intelligence level was subsequently juxtaposed with the results of pronunciation test measured by three English native speakers and Praat software. The major aim of statistical analysis was to find if there is a positive correlation between the above-mentioned variables. For the purpose of this study, we shall focus on the interdependence between musical intelligence level and Praat results, including two variables: F_0 (fundamental frequency) and speech rate. Moreover, it is important to note that the statistical analysis is based on the absolute value, which is the margin between the variables (F_0 and speech rate (SR)) and their model values obtained from Praat. Thus, they are called F_0 range difference (F_0D) and speech rate difference (SRD). The less significant difference between the model value and the outcome obtained from the research, the better final results participants achieved.

Statistical analysis for F0 Range Difference

Table 1. F_0 range difference – selected statistical results.

	Wing Level			Total
Statistic	**Low**	**Medium**	**High**	
Count	27	36	31	94
Average	120.848	93.9244	70.6968	93.9977
Median	120	92.25	61.4	91.2
Standardised Skewness	**2.1255**	1.5738	**2.4828**	**3.2149**
Standardised Kurtosis	**3.1148**	0.54810	1.5862	**2.2515**

As seen above, the group with the high level of musical intelligence achieved the best results regarding average and median values. The analysis of standardised skewness and kurtosis indicate that there might be some deviations from normal distribution. In order to verify this assumption, step two must be taken, namely the application of Shapiro-Wilk test.

Table 2. Test of normality for F_0 Range Difference (94 participants).

Test	***Statistic***	***P-Value***
Shapiro-Wilk W	0.9559	**0.0129**

P-value obtained from the test confirms that it is unable to reject the assumption that F_0 Range Difference comes from a normal distribution. However, the researcher should now check if there are no outliers which may disturb the validity of results. Thus, the next step should be taken: visual methods.

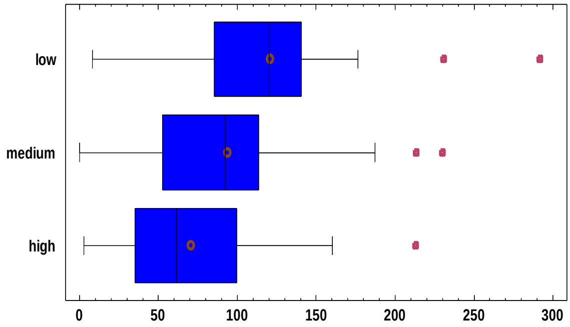

Figure 3. F_0 Range Difference, level low, medium and high.

The box-and-whisker plot shows that there are some outliers (dots outside the whiskers) which should be carefully analysed to ascertain the validity of the given data. The research was based on the recordings of students' utterances, thus the outliers we can easily notice on the box-and-whisker plot were the result of excluded from the research to ascertain its validity. After their removal, the procedure of testing the normal distribution of data set must be repeated in order to see if the outliers impacted the distribution of the studied group.

Before one decides whether parametric or non-parametric test should be applied, one more requirement must be met, which is the homogeneity of variances. In this case, Bartlett's test is used to verify if the variances are equal for all samples, however, data must come from a normal distribution. Also, there are two hypotheses for the test: the null one which indicates that the variances are equal for all studied samples, and the alternate hypothesis which implies that the variances are not equal for one or more pair.

Table 3. Bartlett's test results for F_0 Range Difference.

		Test	p-value
PRAAT	**F_0 D**	1.0213	0.4416

On the basis of the table above, the p-value is greater than (or equal to) 0.05, which means there is no significant difference between the standard deviations at the 95% confidence level. According to that, the parametric test – ANOVA – can be applied. This is step number five.

ANOVA enables the researcher to verify the null and alternative hypotheses. The null hypothesis implies that the arithmetic means obtained from the research (in this case low, medium and high groups) do not differ significantly, whereas alternative hypothesis indicates that at least one arithmetic mean of the given variables is significantly different from the others. The table below presents the results of the ANOVA test.

Table 4. ANOVA table for F_0 Difference.

Source	***Sum of Squares***	***Df***	***Mean Square***	***F-Ratio***	***P-Value***
Between groups	27905.9	2	13952.9	9.08	0.0003
Within groups	121337	79	1535.91		
Total (Corr.)	149243	81			

As seen above, the ANOVA test enumerates two components of F_0 variance: between and within groups. The p-value is less than 0.05, which means that the mean of F_0 Difference of low, medium and high groups are significantly different (95% confidence level). In order to verify which group means differ, one must follow the next step and apply Multiple Range Tests[2].

Table 5. Multiple Range Test for F_0 Range Difference (Tukey HSD).

Contrast	*Significance*	*Difference*	*+/- Limits*
high - low	*	-46.3315	26.042
high - medium		-18.4384	24.5998
low - medium	*	27.8932	25.6382

[2] It is important to note that if the p-value was greater than 0.05, there would be no statistically significant difference between the given groups. In such a case, the Multiple Regression Tests do not have to be applied.

On the basis of the above, the statistical difference between the means of the given groups appears among high-low and low-medium levels. This leads to the conclusion that the least significant difference between the model value and F_0D is between medium and high groups, which suggests positive interdependence among the level of musical intelligence and F_0 range.

Statistical analysis for speech rate difference

Due to the fact that the box-and-whisker plot revealed outliers that should have been removed from the dataset, the next parameter, speech rate difference (SRD), measures the results obtained by 89 participants. The basic statistical measures are presented in the table below.

Table 6. Speech rate difference – selected statistical results (89 participants).

	Wing Level			**Total**
Statistic	**low**	**medium**	**High**	
Count	25	34	30	89
Average	0.3425	0.349	0.2567	0.3161
Median	0.289	0.2325	0.241	0.244
Standardised Skewness	1.1446	**5.7779**	**2.8524**	**8.1781**
Standardised Kurtosis	-0.5034	**7.8448**	**2.369**	**12.78**

Again, the greater the difference between Praat results and Praat model value, the less impact musical intelligence level has on L2 pronunciation skills. Standardised kurtosis and skewness indicate that the dataset is not normally distributed, however, similarly to the previous value, it should be tested by applying Shapiro-Wilk test.

Table 7. Test of normality for speech rate difference.

Test	***Statistic***	***P-Value***
Shapiro-Wilk W	0.8265	**≈0***

***Almost equal to 0**

As seen above, the p-value is less than the significance level of 0.05, which confirms that there are significant departures from normal distribution and results in rejection of null hypothesis. In order to see if there are any outliers which should be removed from the research, the box-and-whisker plot is presented.

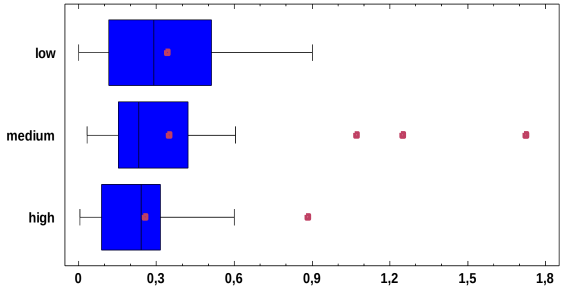

Figure 4. Speech rate difference (89 results).

The analysis of the above-presented box-and-whisker plot confirms that there are another outliers that should be excluded from the research in order to preserve the validity of the study. Following that, the Shapiro-Wilk test should be applied once again in order to validate the normal distribution of dataset.

Table 8. Shapiro-Wilk test for speech rate difference after outliers exclusion.

Variable	Statistic	P-value
SR D	0.9184	≈0

As expected, the results of Shapiro-Wilk test revealed that the above-mentioned variable is not normally distributed. On the basis of that, non-parametric test (in this case Kruskal-Wallis test) can be used for the further steps of statistical analysis.

Table 9. Kruskal-Wallis test results for speech rate difference (SRD).

Wing Level	***Sample Size***	***Average Rank***
Low	24	45.8333
Medium	30	40.9167
High	28	38.4107

Test statistic = 1.28382 **P-Value** = 0.526287

The null hypothesis which is verified with the use of Kruskal-Wallis test indicates that the medians of speech rate difference value within all three groups are the same. Since the p-value for SRD is greater than 0.05, it has been shown that the parameters are not significant. As a result, it is possible to state that the level of musical intelligence does not have an impact on participants' speech rate.

Discussion and Conclusion

The statistical analysis with the use of proper statistical methods and tools can give meaning to the meaningless numbers, bring a new life to the lifeless data, and finally lead to the meaningful research findings and their outstanding interpretations. Such a picture not only can be drawn in the scientific studies but also in the field of linguistics, which used to favour qualitative analyses in the last few decades.

In line with the above-mentioned data analysis, as well as literature review, it is without doubt that every researcher, including those who work in the field of linguistics should be familiar with at least some of the basic concepts of statistical methods, which by no means enable users to conduct well-designed and professional research study. The use of relevant statistical methods helps to obtain valid and reliable results of research and subsequently interpret them appropriately. Conversely, incorrect implementation of statistical techniques caused by insufficient knowledge and inexperience may lead to serious errors, erroneous conclusions which as a consequence may undermine the significance of the research study. Hence, it is of crucial importance to know the basic steps that should be taken in order to design and conduct high quality research. Hence, the major aim of this paper was to present five steps that should be taken in the process of gathering and interpreting statistical data. Based on the data drawn from the quantitative research on musical intelligence and its impact on L2 pronunciation skills, it is argued that an appropriate knowledge about particular statistical methods will result in improving research designs in any field, including linguistics.

References

Aiken, Leona S., Stephen G. West, Roger E. Millsap (2008) "Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America." *American Psychologist*, 63, 32-50. doi:10.037/0003-066X.63.1.32.

Baayen, Harald R. (2008) *Analyzing Linguistics Data. A practical introduction to statistics*. Cambridge University Press.

Cantos Gómez, P. (2013) *Statistical Methods in Language and Linguistic Research*. Equinox, Sheffield.

Chambers, John M., William S. Cleveland, Beat Kleiner, Paul A. Tukey (1983) *Graphical methods for data analysis*. Wadsworth & Brooks/Cole, Belmont, CA.

Eddington, David (2015) *Statistics for linguists: A Step-by-Step Guide for Novices*. Cambridge Scholars Publishing.

Ghasemi, Asghar, Saleh Zahediasl (2012) "Normality tests for statistical analysis: a guide for non-statisticians." *International Journal of Endocrinol Metab*. 10(2), 486-489. doi: 10.5812/ijem.3505

Gries, Stefan (2013) *Statistics for Linguistics with R. A Practical Introduction*. De Gruyter Mouton.

Kolmogorov, Andriej (1933). "Sulla determinazione empirica di una legge di distribuzione." *G. Istituto Italiano degli Attuari* 4, 83–91.

Kusz, Ewa (2019) *Musical intelligence and its impact on English pronunciation skills in the process of second language acquisition*. Frankfurt am Mann, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang.

Larson-Hall, Jenifer (2010) *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York: Routledge.

Lazaraton, Anne (2000). "Current trends in research methodology and statistics in applied linguistics." TESOL *Quarterly,* 34(1), 175–181. doi: 10.2307/3588103

Lazaraton, Anne (2005) "Quantitative research methods." [In:] Eli Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 209-224). Mahwah, NJ: Lawrence Erlbaum 2005.

Loewen, Shawn, Susan Gass (2009) "The use of statistics in L2 acquisition research". *Language Teaching,* 42(2), 181-196. doi:10.1017/S0261444808005624

Loewen, Shawn, Elizabeth Lavolette, Le Anne Spino (2014) "Statistical Literacy Among Applied Linguists and Second Language Acquisition Researchers." TESOL *Quarterly* 48.2 (June 2014), 360–388.

McKnight, Patrick E., Julius Najab (2010) "Kruskal-Wallis Test." *Corsini Encyclopedia of Psychology*, 1. https://doi.org/10.1002/9780470479216.corpsy0524

McHugh, Mary L. (2011) "Multiple comparison analysis testing in ANOVA." *Biochemia Medica,* 21(3), 203-209.

Mendes, Mehmet, and Akin Pala (2003) "Type I Error Rate and Power of Three Normality Tests." *Pakistan Journal of Information and Technology* 2(2), 135-139.

Olleveant, Nicola A. (1999) "Tukey Multiple Comparison test" Blackwell Science Ltd. *Journal of Clinical Nursing*, 8, 299-304.

Plonsky, Luke, Susan Gass (2011) "Quantitative research methods, study quality, and outcomes: The case of interaction research." *Language Learning*, 61, 325-366. doi: 10.1111/j.1467-9922.2011.00640.x

Rasinger, Sebastian M. (2008) *Quantitative Research in Linguistics: An Introduction. Research Methods in Linguistics*. Bloomsbury.

Shapiro, Samuel S., and Martin, B. Wilk (1965) "An analysis of variance test for normality (complete samples)." *Biometrika* 52(3/4), 591-611.

Smirnov, Nikolai (1948) "Table for estimating the goodness of fit of empirical distributions." *Annals of Mathematical Statistics* 19(2): 279–281. doi: 10.1214/aoms/1177730256.

Snedecor, George W., William G. Cochran (1989) *Statistical Methods, Eighth Edition*. Ames, IA: Iowa State University Press.

Thode, Henry J. (2002) *Testing for normality*. New York: Marcel Dekker.

Woods, Anthony, Paul Fletcher, Arthur Hughes (1986) *Statistics in language studies***.** Cambridge University Press.