

OTDK-dolgozat

2025

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

GAZDASÁGTUDOMÁNYI KAR



**A BUDAPESTI KÖZÖSSÉGI KÖZLEKEDÉS FUZZY ÉRTÉKELÉSI
SKÁLA ALAPÚ SZOLGÁLTATÁSMINŐSÉGI ELEMZÉSE**

**FUZZY RATING SCALE-BASED ANALYSIS OF THE PERCEIVED
AND EXPECTED QUALITY OF PUBLIC TRANSPORT SERVICES
IN BUDAPEST, HUNGARY**

Láng Kristóf Ágoston

Konzulens: Jónás Tamás

Kézirat lezárásának dátuma: 2024. 12. 17.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. LITERATURE REVIEW	3
2.1. MEASURING SERVICE QUALITY	3
2.2. SERVICE QUALITY IN PUBLIC TRANSPORT.....	5
2.3. LIMITATIONS OF LIKERT SCALES.....	6
3. METHODOLOGY	9
3.1. FUZZY SET THEORY	9
3.2. FUZZIFICATION AND DEFUZZIFICATION.....	12
3.3. MODIFIED SERVQUAL MODEL	13
3.4. DERIVING THE WEIGHTED MODEL	16
3.5. POTENTIAL WEAKNESSES OF THE METHODOLOGY.....	17
4. PRESENTING THE RESULTS.....	19
4.1. GENERAL CHARACTERISTICS OF THE SURVEY	19
4.2. PRESENTING THE SAMPLE	19
4.3. GAP ANALYSIS	21
4.4. WEIGHTED GAP ANALYSIS	30
4.5. MEAN COMPARISONS	32
4.6. CORRELATIONS	33
5. SUMMARY, CONCLUSIONS AND FOLLOWING RESEARCH.....	35
REFERENCE LIST	37
APPENDICES	I
APPENDIX 1: IMPLEMENTED QUESTIONNAIRE (HUNGARIAN)	II
APPENDIX 2: ONE SAMPLE T TESTS	V
APPENDIX 3: NORMALITY AND INDEPENDENT SAMPLES T TESTS	VI
5.1.1. Based on gender	VI
5.1.2. Based on age groups	VII
5.1.3. Based on PT travel frequency groups.....	VIII
5.1.4. Based on groups by average time spent on PT.....	IX
5.1.5. Based on car ownership status	X
APPENDIX 4: CORRELATIONS	XI

LIST OF TABLES AND FIGURES

Table 1: Dimensions and attributes of service quality used for the construction of the questionnaire for this paper	17
Table 2: Fuzzy expectation, perception and fuzzy gap values	23
Table 3: Defuzzified expectation, perception and gap values with gap value ranking	24
Table 4: Normalised importance weight values and comparison of unweighted and weighted gaps	31
Figure 1: Service Quality Model	4
Figure 2: Linguistic variables for expectation and perception	7
Figure 3: Influence of individual factors on the total assessment of the service quality of public transport in Kaliningrad	15
Figure 4: Public transport usage frequency of respondents.....	20
Figure 5: Time spent on public transportation on an average day	21
Figure 6: Attribute gap values colored by dimension	25
Figure 7: Average expectations colored by dimension	26
Figure 8: Average perceptions colored by dimension	27

1. INTRODUCTION

Public transportation is a service that most people are familiar with, however the importance of a well functioning public transport system cannot be overstated. Urbanization is an ongoing trend in the modern world that poses problems related to sustainability, mobility, and overall quality of life, among others. Hungary is not an exception, and, especially in the Budapest Metropolitan Region, there are serious sustainability concerns arising from the high concentration of population. Research shows that the development and promotion of the urban public transport system are among the most important policies that should be focused on in order to reduce the ecological footprint of the region (Kovács et al., 2022). Analysing the satisfaction of public transport passengers is a key element of the promotion of this service (Hussain et al., 2018), and one that can significantly help with the decision making processes regarding further developments.

As seen in the synthesis of Currie and Wallis (2008), in urban areas the quality of public transport services, as well as improvement measures on the service, are regarded primarily based on technical aspects, while the quality perception of passengers is taken substantially less into consideration. However, the perceived quality of services is well known to be a key element of service quality both in a more general (Hu et al., 2009; Tarn, 1999), and a more specific context (Echaniz et al., 2018).

According to de Oña and de Oña (2015), when it comes to measuring service quality in the public transport industry, the significant majority of the literature available falls between the performance perception and expectation approach (Parasuraman et al., 1985) and the only performance perception approach (Cronin. Jr & Taylor, 1992), and the use of a more sophisticated, weighted version of both these models is also widely adopted. The former one, which also lead to the development of the widely adopted SERVQUAL scale (Parasuraman et al., 1988), derives service quality from the gap between the service perceived and expected by the consumer. In contrast, the latter approach (along with other studies such as that of Babakus and Boller (1992)) suggests that the expected service component of the above proposed model might not be necessary in all settings, as consumer expectations are sometimes imbedded in the perceived quality. In both of these approaches customer satisfaction surveys have a critical role, and they also show that Likert scales, 5-point Likert scales in particular, are the most widely adopted rating measures in these surveys.

The Likert scale in general is popular for good reason: its ease of use and general applicability make it a strong tool for the quantification of otherwise unquantified, subjective data such as opinions and perceptions. However, its usage for this purpose is a more and more debated issue, as the variables such a scale tries to quantify are fuzzy and ambiguous by nature, opposed to the crisp number that the Likert scale would suggest (Tóth et al., 2019; Lupo, 2013a). Fuzzy set theory (Zadeh, 1965) provides a more and more frequently used toolset in the general field of service quality measurement (Kuo et al., 2007; Árva et al., 2019; Benítez et al., 2007; Chou et al., 2011), capable of better quantifying both the uncertainty associated with the decision making while rating subjective variables on Likert scales and the flexibility of the variables themselves.

The objective of this paper is to use a carefully selected combination of the above mentioned tools and methodologies in order to assess the service quality of the public transport system in Budapest, and show how these methodologies might be applied in the industry and in the more specific setting as well. First, the relevant academic literature is reviewed, further discussing the state of the art of service quality measurement both generally, and in the specific case of public transport. Second, the elements of the proposed methodology are presented and discussed, addressing their potential weaknesses, and finally a detailed discussion of the results is put forward along with conclusions and recommendations based on the analysed data.

2. LITERATURE REVIEW

2.1. MEASURING SERVICE QUALITY

Parasuraman, Zeithaml and Berry (1985) emphasize three important characteristics of services in comparison with goods that must be considered to fully, or at least better understand service quality, which are intangibility, heterogeneity and inseparability. They argue that services are not testable, and therefore their quality is not verifiable in advance to sales (intangibility), that service performance varies both over time and between customers (heterogeneity), and that the production of a service is not separable from its consumption. As a result of these characteristics, consumers evaluate the quality of services along different dimensions than goods, which are also more difficult to observe. They develop a model describing four gaps between consumer expectations and management perceptions of these expectations (Gap 1), these management perceptions and the service quality specifications of the firm (Gap 2), service quality specifications and the actual delivery of service (Gap 3), and finally a gap between the delivery of the service and external communication about it towards consumers (Gap 4). They propose that these gaps affect the service quality from the standpoint of the consumer, and that there exists a fifth gap between expected and perceived service that on the one hand depends on the nature of the above mentioned gaps, and on the other hand directly influences the perceived quality of the service with its magnitude and direction. Gap 5 is measurable as the difference between stated perceived and stated expected service values, and it can be calculated for both the entirety of the service, and its elements one by one. The structure of this conceptual model can be seen in Figure 1. This approach led to the development of the SERVQUAL (Parasuraman et al., 1988), a scale for the measurement of service quality, which became the foundation of methodology in a significant portion of related literature utilised in a wide range of industries (Wang et al., 2015), and more specifically in the public transport sector as well (de Oña & de Oña, 2015).

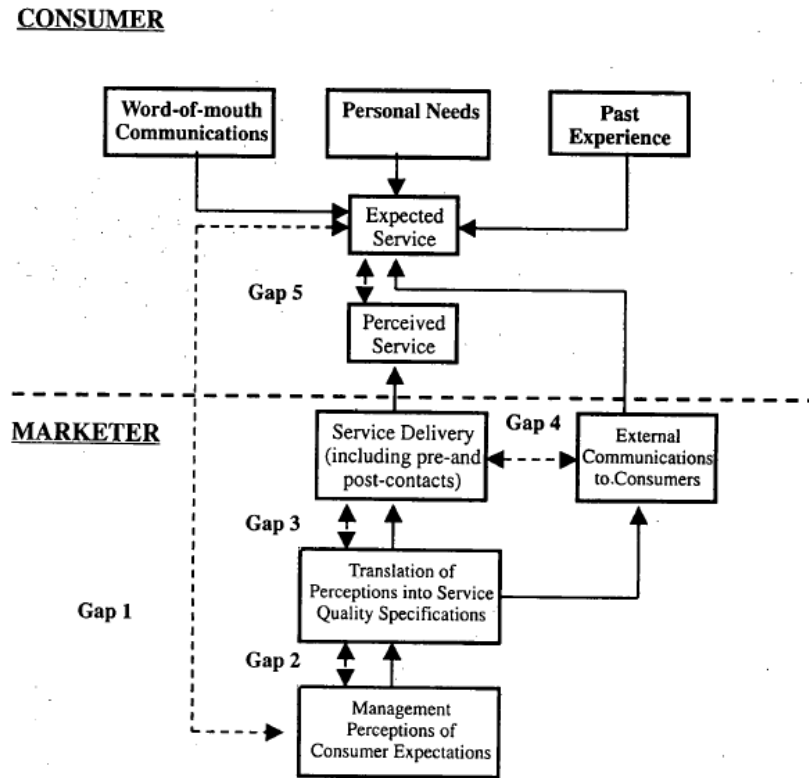


Figure 1: Service Quality Model

Source: Parasuraman et al. (1985, p. 4)

Explanation: The figure identifies the factors affecting service quality and shows the gaps between these factors as explained above

As with many other general models applied to a wide range of different problems and over large periods of time, in many articles some situation specific alterations have been implemented to the original SERVQUAL methodology. Parasuraman, Zeithaml and Berry (1988) put forward a five dimensional scale consisting of Tangibles, Reliability, Responsiveness, Assurance and Empathy, further breaking these each down into either four or five attributes (see Figure 3 for an example of breaking down service quality into its attributes). These proposed attributes of service quality are to be assessed by respondents according to both their expectations and their perceptions about the service in question on a seven point ordinal scale. This instrument (in line with the suggestion of the authors) is often adapted in recent literature to fit the needs of a specific research. Lupo (2013a) utilizes the AHP method (Saaty, 1977) for acquiring information about expectations while also applying fuzzy set theory through fuzzy-linguistic evaluation scales to the problem to better represent the inherent uncertainty characteristic of service performance analyses. He also

breaks down service quality into a different set of dimensions to develop a questionnaire that better suits the environment of higher education. Chou et al. (2011) demonstrates the effectiveness of employing fuzzy numbers as well, combined with a weighted SERVQUAL model obtaining importance weight values from respondents as an additional part of the questionnaire. They assess the usefulness of the weighted model by comparing its results with a model not taking the weight values into account, and conclude that there are significant differences between the results. Similarly to Lupo (2013a), they use fuzzy-linguistic scales instead of a traditional Likert scale, which is a common method to incorporate subjectivity and uncertainty. In terms of service quality dimensions they include some airline-specific elements in the general SERVQUAL model. Wahyudi (2017) investigates the role of using fuzzy numbers in the SERVQUAL method by comparing the Gap 5 results of a fuzzy-linguistic scale and a traditional Likert scale based, otherwise identical assessment method. He concludes that there are no statistically significant differences between the two methods, however he mentions several possible research-specific reasons for his results, and argues that conducting similar studies in different settings is necessary to better understand the role of fuzzy logic based assessment.

Implementations of a weighted SERVQUAL model are also decently common in the field. When developing such a model, the researcher assesses the gap 5 values after multiplying them with their respective importance weights (Cronin, Jr & Taylor, 1992). In the relevant literature, according to de Oña and de Oña (2015), these weights are either determined according to the stated importance of service quality attributes (Chou et al., 2011), that is, the subjective ratings of stakeholders on the importance of attributes, or derived using various statistical methods such as factor analysis, regression analysis or structural equation models. Jain and Gupta (2004) argue that while a weighted approach does not raise the validity or the explanatory power of the original SERVQUAL model (and is not superior in those regards to the weighted or unweighted SERVPERF models either), it does yield a significantly stronger diagnostic power, providing much deeper insight into areas in need of managerial intervention.

2.2. SERVICE QUALITY IN PUBLIC TRANSPORT

Looking at the public transport sector more closely, Ojo and Adom (2017) note that although some differing approaches are being applied to the field (Eboli & Mazzulla, 2007), SERVQUAL model based methods are between the most frequently applied in the industry.

Research-specific alterations however are still regularly incorporated regarding both the exact methodology of collecting and of analyzing the data.

Lupo (2013b) applies a very similar methodology as in his previously mentioned research, taking advantage of the AHP method, fuzzy set theory, and modifying the traditional service dimensions and attributes to better fit the setting. He derives these from the work of Eboli and Mazzulla (2007), who formulated a structural equation model to determine the overall structure of public transit. Moslem and Duleba (2018) used the AHP method as well, however they conducted their survey on transport experts, not the general passengers of public transport. Nevertheless, their results are of interest in this paper, as their research took place in Budapest, Hungary. They found that reliability is the most critical issue according to expert evaluators, followed by time availability, and that physical comfort and perspicuity are the least important attributes of service quality.

Mikhaylov et al. (2015) utilize the SERVQUAL instrument more traditionally, keeping the original five dimensions, but tailoring the attributes (and the number of attributes per dimension) to fit the public transport industry, with an emphasis on the reliability dimension. They assess expectations and perceptions on a seven point scale, and investigate the results not only in terms of the calculated Gap 5 values, but also comparing the percentage of negative (1 to 3), neutral (4), and positive (5–7) responses along their proposed attributes. Kokku et al. (2011) approach the problem similarly, but they include an additional quality dimension in the questionnaire, culture, with five different attributes.

2.3. LIMITATIONS OF LIKERT SCALES

Inspecting the summary of previous research with the help of de Oña and de Oña (2015), we can see that the most frequently used scale in the literature of public transport service quality evaluation is by far the Likert scale, with an emphasis on 5-point Likert scales. There are unquestionable advantages of these scales (ease of use on both the researcher and respondent side, general applicability, etc.), however, it is of vital importance to address the limitations of the Likert scale in this field. Lupo (2013a) notes that the preferences and judgments expressed by people are often characterised by subjectivity, inconsistency and vagueness, and that many popular methods to analyze service quality are unable to deal with the uncertainty arising from these characteristics. Li (2013) also states that the ordinal nature of the Likert scale leads to information distortion and loss, while Chou et al. (2011, p. 3) label Likert scales „inadequate and too simple” to handle the complexity of human decision making. They address the fact that in terms of interpreting the results, means and standard

deviations are less appropriate figures than frequencies or percentages are as well. Dombi and Jónás (2021) emphasize that due to the ordinal nature of data derived from Likert scales only statistical tools applicable to ordinal data should be used in the statistical analysis of the results. Árva et al. (2019) also note that the common practice of assuming adjacent values on a Likert scale to be of equal distances from each other (like on an interval scale) can lead to incorrect interpretations of their carried information.

All of the articles above (and many more) suggest the implementation of fuzzy set theory as a toolset that provides the necessary means to account for the inherent uncertainty in human decision making. Fuzzy set theory can be implemented into the SERVQUAL model on different levels, examples of which we have covered above. The use of a fuzzy-linguistic scale is a decently common approach in the literature, as it is a method that combines the ease of use of the Likert scale with the better information carrying characteristic of fuzzy numbers. When these scales are used, respondents are asked to indicate their opinions on a Likert scale, where each of the potential answers is explained by a linguistic expression. These expressions are attached to fuzzy numbers predetermined by the researcher(s), and the answers given by respondents are translated into fuzzy numbers based on these predetermined values. An example of such a fuzzy-linguistic evaluation scale with the corresponding fuzzy values can be seen on figure 2.

Table 2
Linguistic variables for expectation and perception.

Very Poor	(0.0,1.0,2.0)
Poor	(1.0,2.0,3.0)
Fair	(2.0,3.0,4.0)
Good	(3.0,4.0,5.0)
Very Good	(4.5,5.0,5.0)

Figure 2: Linguistic variables for expectation and perception.

Source: Chou et al. (2011, p. 5)

Explanation: An example of a fuzzy-linguistic conversion table

Although it is shown that these scales better represent human thinking than more traditional, crisp value based methods (Liou & Chen, 2006), they have their own disadvantages, as consumer perceptions of what these linguistic values represent might differ significantly from the predetermined fuzzy equivalents, and these perceptions may vary depending on the attribute of service quality being evaluated as well (Árva et al. 2019).

Chou et al. (2011, p. 10) also note that, when using a fuzzy linguistic scale, „it is difficult to justify appropriately and accurately the membership function of passenger’s perception and expectation”.

One proposed solution to these problems (Tóth et al., 2019) is the adoption of a 10 point fuzzy Likert scale, where instead of evaluating service performance with the crisp values on the scale, respondents are given the freedom of expressing their uncertainty and the heterogeneity of the service via selecting three values on the scale. These three values represent (in order) the level below which performance did not fall, the usual level of performance, and the level over which performance did not rise. With the help of these values a triangular fuzzy number (or more generally, a flexible fuzzy number (Dombi & Jónás, 2021)) is formulated, which overcomes not only the limitations of traditional Likert scales (by converting evaluations into a flexible environment), but also allows respondents to directly incorporate the level of variability in the service into their evaluation. Their evaluations can then be aggregated and analyzed to form interpretations of service quality.

3. METHODOLOGY

The proposed methodology of this paper is a combination of different tools and techniques mentioned above. This paper uses a SERVQUAL based approach that is, compared to the original tool, modified in three ways. Firstly, a fuzzy Likert scale is introduced in the assessment of perceived and expected service performance to address the weaknesses of both traditional Likert scales and fuzzy-linguistic scales, to better quantify the subjective data. Secondly, the general dimensions and attributes of service quality in the SERVQUAL model are modified to better fit the research setting. Thirdly, a weighted model is developed and its results are compared with those of the unweighted model, assessing the additional insights provided by incorporating weights.

In the following subchapters, a brief overview of the aspects and properties of fuzzy set theory and fuzzy numbers that is utilized in this research is included, along with a short explanation of the technique used to defuzzificate the results in order to be able to perform statistical analyses on them easier. The proposed modifications to the SERVQUAL model are also presented and justified, giving some context to the decision-making behind the modifications. Then, the necessary background for deriving weight values is discussed, and lastly, the potential weaknesses of the proposed methodology are addressed.

3.1. FUZZY SET THEORY

The fuzzy set theory was first introduced by Zadeh (1965). A fuzzy set is a subset of some classical set X (the universe) without a sharp membership boundary. Fuzzy set A is completely characterized by its membership function, $\mu_A: X \rightarrow [0; 1]$. This membership function assigns to each element of the universe a real number on the $[0, 1]$ interval, determining its degree of belonging to set A . The closer $\mu_A(x)$ is to 1, the more x belongs to A , with extreme values $\mu_A(x) = 0$ meaning x is not a member of A , and $\mu_A(x) = 1$ meaning x is a member of A to the fullest possible extent. A fuzzy set allows us to express values and intervals with no sharp boundaries. (Dubois & Prade, 1980).

The fuzzy subset of the real line A ($\mu_A: \mathbb{R} \rightarrow [0, 1]$) is a fuzzy number if its membership function fits the following criteria:

1. μ_A is normal, meaning that there exists such an x_0 where $\mu_A(x_0) = 1$
2. μ_A is fuzzy convex, meaning $\mu_A(tx + (1 - t)y) \geq \min\{\mu_A(x), \mu_A(y)\}$ for every $t \in [0, 1]$, where $x, y \in \mathbb{R}$

3. μ_A is upper semicontinuous on $\mathbb{R}$, meaning for every $\varepsilon > 0$ exists a $\delta > 0$ such that $\mu_A(x) - \mu_A(x_0) < \varepsilon, |x - x_0| < \delta$
4. μ_A is compactly supported, meaning that the set $\{x \in \mathbb{R}; \mu_A(x) > 0\}$ is bounded (Bede, 2013)

Triangular fuzzy numbers (TFN) are a special subset of fuzzy numbers. A triangular fuzzy number denoted as $T(l, m, r)$ with parameters $l, m, r \in \mathbb{R}, l \leq m \leq r$ has the following membership function:

$$\mu_T(x, l, m, r) = \begin{cases} \frac{x-l}{m-l}, & \text{if } l < x \leq m \\ \frac{x-r}{m-r}, & \text{if } m < x < r \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

(Benítez et al., 2007)

Triangular fuzzy number $T(l, m, r)$ can be regarded as one representation of the linguistic term “x is approximately equal to b”. TFN-s are of great use in this research, as data regarding the expectations and perceptions of passengers can be gathered in the form of TFN-s, or more specifically in the form of TFN parameters l , m , and r .

An important note here is that using these inputs there are many ways to construct fuzzy numbers, that is, many different functions could be used between the l , m and r values, leading to nontriangular fuzzy number representations. Flexible fuzzy numbers (Dombi & Jónás, 2021), for example, utilize the properties of the kappa function with an additional λ parameter as a membership function to represent different levels of uncertainty around m :

$$\mu(x, l, m, r, \lambda) = \begin{cases} \frac{1}{1 + \left(\frac{m-x}{x-l}\right)^\lambda}, & \text{if } l < x \leq m \\ \frac{1}{1 + \left(\frac{r-x}{x-m}\right)^{-\lambda}}, & \text{if } m < x < r \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The λ parameter allows for the creation of bell-shaped membership functions when $\lambda > 1$ that emphasize (give higher degree of membership) to values around m , and „reverse” bell-shaped membership functions when $1 > \lambda > 0$ that emphasize extreme values too. In this research setting this would mean either the emphasis or the diminishment of minimal and maximal expectations and perceptions indicated by respondents. None of these is a particular goal of this paper, however, therefore TFN-s are used, which also correspond to the usage of flexible fuzzy numbers with $\lambda = 1$.

As our data is to take the form of TFN-s, we must examine their properties regarding relevant operations, so that we can later execute calculations on them. The most relevant operation in this setting is taking weighted average of multiple TFN-s, which is necessary in order to aggregate them. The definition of the α -cut of a fuzzy set is a prerequisite to defining this operation.

The α -cut of a TFN denoted $T(l, m, r)$ is defined as the following interval (where $\alpha \in (0, 1)$):

$$T_\alpha = [(m - l)\alpha + l, (m - r)\alpha + r] \quad (3)$$

(Lupo, 2013b). This can be worded as the widest possible subset of the universe X where each element has a higher or equal level of membership in T than α . Using this definition, the proposition can be put forward that with TFN-s $T(a_1, b_1, c_1), T(a_2, b_2, c_2), \dots, T(a_n, b_n, c_n)$ where $a_i < b_i < c_i$ and $i = 1, 2, \dots, n$, $n \geq 1$ and with weights $w_1, w_2, \dots, w_n \in [0, 1]$, $\sum_{i=1}^n w_i = 1$, a triangular fuzzy number $A(a, b, c)$ can be constructed, the three parameters being

$$a = \sum_{i=1}^n w_i a_i, \quad b = \sum_{i=1}^n w_i b_i, \quad c = \sum_{i=1}^n w_i c_i$$

Here it is true that for every A_α α -cut of this TFN

$$A_\alpha = \sum_{i=1}^n w_i A_{\alpha,i} \quad (4)$$

where $A_{\alpha,i}$ denotes the α -cut of the TFN $T(a_i, b_i, c_i)$, where $i = 1, 2, \dots, n$ and $\alpha \in (0, 1)$. This is proven by Dombi and Jónás (2021) to the broader set of flexible fuzzy numbers, and they also show that TFN-s are a subset of flexible fuzzy numbers. Capitalizing on this proposition, we can interpret the weighted average of triangular fuzzy numbers as a TFN constructed with parameters calculated as the weighted average of the respective parameters of our analyzed set.

In order to perform a fuzzy gap analysis, that is, an analysis of the fuzzy representations of the gap 5 values, a clear definition of the difference of two fuzzy numbers is needed. Klir and Yuan (1995) define the α -cut of fuzzy set $A * B$ (denoted as $(A * B)_\alpha$) as

$$(A * B)_\alpha = A_\alpha * B_\alpha \quad (5)$$

where $\alpha \in (0, 1]$ and $*$ represents any of the four basic arithmetic operations, with the additional remark that $0 \notin B_\alpha$ is required for every $\alpha \in (0, 1]$ when $(*)$ represents $(/)$. They also define addition and subtraction on intervals as follows:

$$[a, b] + [d, e] = [a + d, b + e] \quad (6)$$

$$[a, b] - [d, e] = [a - e, b - d] \quad (7)$$

They also show that the result of these operations (in equation (5)) is also a fuzzy number. Using these definitions, it can be shown that the sum of TFN-s $A(l_a, m_a, r_a)$ and $B(l_b, m_b, r_b)$ is also a TFN, with parameters

$$A + B = C(l_a + l_b, m_a + m_b, r_a + r_b) \quad (8)$$

and so is their difference, with parameters

$$A - B = D(l_a - r_b, m_a - m_b, r_a - l_b) \quad (9)$$

We can see that the α -cut of the sum (or difference) of two TFN-s is always a wider interval than the individual α -cuts of the two original TFN-s (for a given α), while the middle point m of the resulting TFN can always be derived as the sum (or difference) of the original middle points. We can interpret this as the uncertainty of the result being close to the newly computed middle point growing with each consecutive operation. When computing fuzzy gap values, we can observe and analyze this growth in uncertainty, while this would not be possible with a traditional, crisp gap value analysis.

3.2. FUZZIFICATION AND DEFUZZIFICATION

As seen time and time again in the literature, the use of a fuzzy-linguistic scale is a popular method to generate fuzzy numbers from the expectations and perceptions of respondents. However, the use of predetermined fuzzy numbers tied to each value of the Likert scale can be arbitrary, as the perception of the respondent of what a given linguistic value represents can significantly differ from the predetermined values both generally, and between the different investigated attributes (Árva et al., 2019), and arguing in favor of the exact used membership function is difficult, especially in the case of linear membership functions (Chou et al., 2011).

To overcome this issue, the methodology used in this paper provides respondents more freedom to express their expectations and perceptions, giving them a higher influence on the way their opinions are translated into fuzzy values. In line with the methodology proposed by Tóth et al. (2019), respondents were asked to answer each of the service quality perception related questions with three separate values on a 10 point discrete scale, ranging from the worst possible to the best possible performance. The three values they had to indicate were, in order, the lowest perceived level, the general experience, and finally the highest perceived level of performance. The same method was used when assessing expectations, with respondents asked to indicate the lowest acceptable, average expected and highest realistically expectable values on the same scale. These values served as the

above mentioned input parameters l , m , and r used to formulate triangular fuzzy numbers. The advantage of this approach is that respondents are able to express both their uncertainty and the variability in service quality expectations and performance in a more precise and quantifiable way (Tóth et al., 2019).

Although the collection of fuzzy data makes it possible to gain deeper insights into the uncertainty and variability associated with service quality, the direct statistical analysis of fuzzy numbers is complicated and outside of the scope of this research. Additionally, crisp gap 5 values can be compared, weighted and interpreted with the ultimate goal of decision making much easier than their fuzzy counterparts (although fuzzy gap values are also analyzed in this paper). This makes the process of defuzzification necessary once the fuzzy expectation and perception values have been aggregated and evaluated, so that the following statistical and gap analysis can be carried out.

There are several defuzzification techniques, that is, techniques to convert fuzzy numbers into crisp values in the literature. These range from more complex techniques, such as the graded mean integration representation method utilized by Chou et al. (2011) to less computationally challenging approaches such as that utilized by Benítez et al. (2007). Ultimately, when computing crisp values from triangular fuzzy numbers that can be defined with parameters l , m , and r as seen above, most of the possible defuzzification methods can be simplified into taking a weighted average of these parameters, sometimes giving parameter m a slightly higher weight ($\frac{4}{6}$ in the case of the first, and $\frac{1}{2}$ in the case of the second mentioned method). In this paper, both for simplicity and in order to best represent the levels of volatility in service quality indicated by respondents, a simple centroid method is used. This is based on the idea of visualizing fuzzy number A with its membership function as a plane figure, and using the x coordinate of its centroid point to represent it with a crisp number. For a TFN, it can be easily shown that we can calculate this crisp value as follows:

$$x_0(A) = \frac{1}{3} (l + m + r) \quad (10)$$

where x_0 represents the sought x coordinate (Wang et al., 2006).

3.3. MODIFIED SERVQUAL MODEL

As seen before, in the relevant literature most studies use a modified version of the SERVQUAL approach. This is because the original model is very general, developed for a wide range of industries, and an adaptation for a specific research setting is not only

possible, but suggested by the original authors as well (Parasuraman et al., 1988). Many of these adaptations (including this paper) keep the five dimensions of the original scale, and tailor the specific attributes (along with the number of attributes per dimension) to the assessed sector, together with the occasional addition of one or two sector-specific dimensions (Chou et al., 2011; Kokku et al., 2011; Mikhaylov et al., 2015). However it is worth noting that a complete restructuring of the dimensions and attributes can be a beneficial approach as well, as seen in the research of Lupo (2013b).

When designing the questionnaire for this paper, a number of factors were taken into consideration. The original SERVQUAL scale was used as a framework with five dimensions as described by Parasuraman et al. (1988). The number of attributes belonging to each of these dimensions, however, were reduced significantly. In the case of some attributes under Responsiveness and Assurance this is a more frequently used practice (Mikhaylov et al., 2015; Kokku et al., 2011; Lupo, 2013b), as some of the employee behaviour related questions in the original instrument are significantly less applicable to public transport than in other industries.

Further attributes were removed due to concerns about the excessive length of the questionnaire: while the original instrument along with most of the reviewed literature investigates 22 attributes, adding up to 44 questions altogether, with the proposed methodology of this paper a similar questionnaire would require 132 values to be indicated by each respondent, not counting demographic related information. A survey of this length could lead to respondents losing interest and considering each question less, especially as the survey was conducted in a less committed environment, completely online. As the goal of the methodology is to give respondents the ability to express more complex opinions, their level of interest and consideration are of vital importance with regards to the results of this research, rendering a shorter survey maintaining a certain level of attention necessary.

For these reasons, altogether 10 attributes were determined, two in each dimension of service quality. When choosing which questions to include several of the above cited papers were considered together with some additional sources with a more similar setting (Mikhaylov et al., 2015; Hussain et al., 2018; Lupo, 2013b; Kokku et al., 2011; Imam, 2014; Borbás, 2022, Khademi–Vidra et al., 2024), selecting the more frequently appearing attributes investigated by these papers, and also those that are most applicable to the setting of Budapest. All the questions were slightly rephrased and translated, as the survey was conducted in Hungarian. The attributes used by Mikhaylov et al. (2015) can be seen on Figure 3 showing the results of their research, as an example source for the design of this

paper's questionnaire. Additionally, in Table 1 the structure of service quality used in this research is visible, broken down into investigated dimensions and service attributes. The last section of the survey, assessing demographic information, was also constructed based on the studies above, collecting the most important demographic characteristics for this industry. The entire survey is available in appendix 1.

Group title/statement	Number of respondents, in %			Expectations – perceptions gap	Rank
	Negative 1 to 3	Neutral 4	Positive 5 to 7		
Reliability					
Bus drivers wait until the bus is fully loaded	38	18	44	2.319	1
Buses leave the bus stop on schedule	35	17	48	2.135	2
Buses arrive to the bus stop on schedule	34	18	48	2.062	3
Information on the route and schedule of the bus is available to the passengers	24	18	58	1.764	4
Personnel does not short change passengers	13	9	78	1.158	5
Bus drivers stop at all bus stops	17	7	77	0.826	6
Bus drivers always comply with itinerary	10	5	85	0.655	7
Responsiveness					
Driver/conductor is responsive and ready to prompt the desired stop	23	15	62	1.491	1
Driver / conductor independently solve emerging issues, and not try shift the responsibility to third parties	27	17	56	1.436	2
Assurance					
Bus drivers adhere to careful driving style	33	20	47	2.348	1
Personnel prevent boarding of drunken, dirty and socially dangerous passengers	36	16	48	2.135	2
Driver/conductor know the city and can prompt the desired stop	15	11	74	1.081	3
Empathy					
Driver/conductor assist disabled and passengers with prams	49	13	39	2.387	1
Conductor is loyal to passengers in the event of an unforeseen situation	37	19	45	1.647	2
Bus drivers always comply with a request to release a passenger in a traffic jam	28	17	56	1.187	3
Drivers wait for a passenger running to the bus	29	13	58	1.138	4
Bus routes are designed optimally, no additional bus stops needed	18	6	76	1.010	5
Tangibles					
On-board temperature is always optimal	49	18	33	2.995	1
Appearance of Kaliningrad buses is aesthetic	46	16	38	2.683	2
Interior of buses is clean	39	18	43	2.616	3
Driver/conductor are neat and wear uniforms	47	19	34	2.535	4
Information on the route and schedule of the bus is intelligible, unambiguous	21	14	65	1.512	5

Figure 3: Influence of individual factors on the total assessment of the service quality of public transport in Kaliningrad

Source: Mikhaylov et al. (2015, p. 4)

Explanation: The figure shows the 5 main dimensions and the 22 attributes utilized by Mikhaylov et al. (2015) to develop their questionnaire. This figure also serves as an example to demonstrate a questionnaire developed in the SERVQUAL framework.

3.4. DERIVING THE WEIGHTED MODEL

When conducting a weighted SERVQUAL research (which is a goal of this paper), the researcher must collect information on the level of importance of each service quality attribute. These importance levels are then (sometimes in a normalised form) applied as scalars to the gap values of the unweighted model, amplifying the gaps corresponding to more important attributes while diminishing those of less important attributes.

As explained above, papers conducting a weighted SERVQUAL analysis either collect stated importance values from participants of the questionnaire or derive these values by more complex statistical means. In this paper, because the weighted gap analysis was conducted after survey responses were already collected, the average expectation values indicated by respondents regarding the different service quality attributes were also used as weights, in a normalised form. This simplification can be justified by the proposition that, based on the exact wording of the questions, respondents can interpret the expectation scales in many ways, including as a means to express the importance of each attribute (Ojo & Adom, 2017). As the questions regarding the minimally and maxially expectable level of service attributes are more explicitly connected to the expectation level of respondents, only their average expectations (the m values for the expectation TFN-s) are used to formulate the weight values. It is worth noting that while the rankings of respondents' expectations toward attributes and their stated importances are often in close relation to one another (Chou et al., 2011), the proposition that respondents rank attributes based on expectations and importances the same way is not well founded in the general literature, and the method of using expectation values as weights is unideal and primarily deployed due to resource constraints.

Table 1: Dimensions and attributes of service quality used for the construction of the questionnaire for this paper

Dimension	D1: Tangibles	D2: Reliability	D3: Responsi- veness	D4: Assurance	D5: Empathy
Attributes	A₁₁: Modern, up-to-date vehicles	A₂₁: Accurate and accessible information on schedules	A₃₁: Employees willing and ready to help	A₄₁: Ability to provide safety on board	A₅₁: Optimal design of public transport network
	A₁₂: Clean and comfortable travel conditions	A₂₂: Adherence to schedules	A₃₂: Fast and appropriate reaction to unforeseen complications	A₄₂: Polite and trustworthy employees	A₅₂: Capability of serving the specific needs of customers

Source: Own research

3.5. POTENTIAL WEAKNESSES OF THE METHODOLOGY

Although the proposed methodology addresses a number of previously mentioned frequent shortcomings of service quality assessment techniques, it is not without its own weaknesses that are worth mentioning.

The fuzzification of respondents' opinions through asking three separate input values for each service attribute might yield a more accurate representation of their feelings than fuzzy-linguistic scales, but the length and complexity of the questionnaire are significantly increased as a result. The increased complexity could lead to a lower level of valid responses and a lower level of responses that truly represent the exact opinions of respondents as well, while the increased length necessitates a reduction in the number of investigated attributes. This is of such a level that the validity of the results when talking about overall service quality needs to be questioned: with such few attributes the quality of the overall service might not be coverable. Additionally, implementing the AHP method to measure expectations or applying a weighted SERVQUAL approach collecting information on stated importance might lead to more relevant results, however these would complicate the completion of the survey even further. Applying factor analysis or utilizing structural equation models to determine weight values would also elevate the trustworthiness of the model, but these require statistical tools outside the scope of this research.

The implementation of a „calibration phase” like we can see in the research of Árvai et al. (2019) where a subset of the respondents is questioned separately in order to determine the fuzzy numbers that the linguistic values on a fuzzy-linguistic scale should represent could significantly shorten the questionnaire. However such a research requires more preparation, and significantly more investment on the part of the respondents calibrating the scale as well.

As the result of these limitations, this paper should be treated more as an example of how this specific methodology might be applied in the field and less as a representative analysis of service quality. This is also due to the method of data collection and the resulting disproportionate representations of socio-demographic characteristics, which are discussed in the following chapter.

4. PRESENTING THE RESULTS

4.1. GENERAL CHARACTERISTICS OF THE SURVEY

The survey was conducted between May 13th and May 30th in 2024, during which period a total of 32 respondents filled it out correctly, with an additional 6 incorrectly completed surveys. These 6 were filled out in a way that the three values given for some of the attributes were not always monotonically increasing, meaning that some average expectations or perceptions were lower than their minimal counterpart, or the highest expectable or perceived performance was indicated as lower than average. In accordance with equation (1) and the definition of triangular fuzzy numbers, equality between m and any (or both) of the two other values was allowed, however the above explained behavior led to the answer being treated as invalid. In some cases, there were only a few values not in line with the rules of filling out the questionnaire suggesting that these errors were due to mistakes on the respondent's part, while in other cases the excessive number of errors suggested that these respondents either did not understand the rules or were not invested enough to follow them. The following analyses were conducted only on the valid, correctly filled out surveys.

The survey was conducted entirely online through social media posts between friends, acquaintances, and those interested in Hungarian public transportation. A random sampling method would have been more suitable in order for the results to be representative of the population, but this was not possible due to resource limitations. As a result, conclusions to the whole population cannot be made on the basis of this sample, and as discussed before, this research serves more as an example of the application of the methodology.

The survey was conducted using Google Forms, as a more complex tool was not needed to construct the questionnaire. Introducing a functionality that ensures that respondents are unable to give invalid answers might have boosted the number of valid responses, however such a functionality would have hidden the problems most likely arising from the complexity and length of the questionnaire, or the ambiguous wording of the instructions. Analysis on the data was carried out using Microsoft Excel and SPSS Statistics, which together were suitable to handle and provide insights into the data, as well as create the necessary graphs and tables for the analysis.

4.2. PRESENTING THE SAMPLE

As the first step of the analysis, the sample composition of the research is presented based on more general socio-demographic factors, as well as more specific characteristics relevant

regarding opinions on public transport. Only individuals who travel by public transport in the capital on some level of regularity were asked to fill out the questionnaire.

Out of the 32 valid responses 20 (62.5%) were given by female respondents, while only the remaining 12 individuals were male (37.5%). When dividing the sample by age groups, we can see that a majority of the respondents fell in the 18–29 years age range, while the age groups of 30–39 and over 69 years were underrepresented. There were no respondents under 18 or in the 60–69 age range.

As seen on figure 4, a majority of respondents (59.38%) travel almost every day using public transportation, and altogether more than 80% of respondents travel at least multiple times a week, meaning that their opinions on public transportation services are relevant, as they are generally aware of the quality of the services provided.

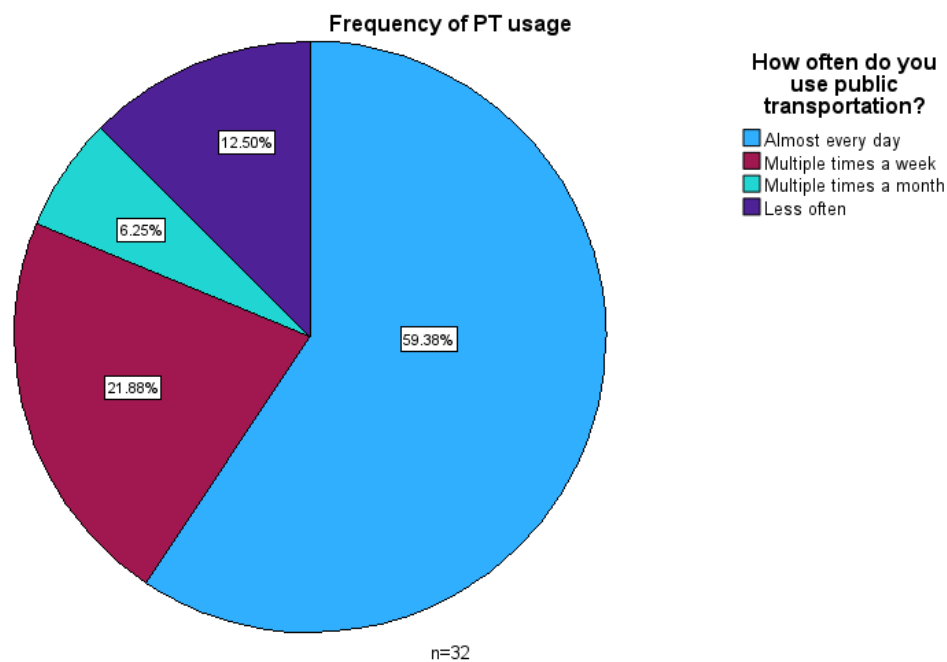


Figure 4: Public transport usage frequency of respondents

Source: Own research

Around one third (37.5%) of respondents travel between half and one hour on an average day, while a similar portion of them (31.25%) travel between one and 1,5 hours daily. Altogether these two groups make up almost two thirds of the sample, while individuals with average travelling times outside this range constitute to a minority. Proportions of daily travel time groups in the sample can be seen on figure 5. Exactly half of the respondents reported owning a car.

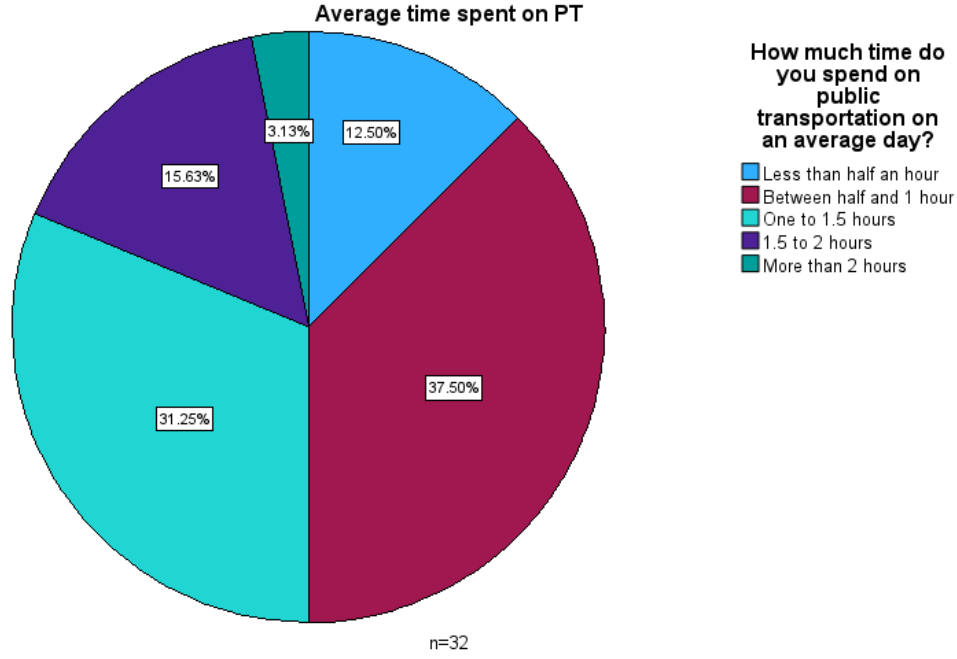


Figure 5: Time spent on public transportation on an average day

Source: Own research

4.3. GAP ANALYSIS

After reviewing the demographics, the expectation–perception gaps can be investigated deeper based on the results presented in tables 2 and 3. Table 2 shows each of the service quality dimensions and attributes assessed (further explained in Table 1 and in appendix 1 showing the full questionnaire), along with the fuzzy expectation and perception values associated with them. The fuzzy values for the attributes were calculated the same way a TFN was calculated for equation (4), with weights of $\frac{1}{n}$ for each parameter of each TFN, n being the sample size of 32. Therefore, the values within the brackets represent the arithmetic mean of l , m , and r values given by respondents for each of the attributes both in terms of expectations and perceptions, in this order.

Alongside, fuzzy representations of gap 5 values are shown, calculated in accordance with equation (9), showing the differences between their corresponding expectation and perception values. It is worth noting that in the relevant literature there are examples of gap 5 values being calculated both as expectations deducted from perceptions and vice versa; this paper uses the latter approach, meaning that a positive gap value shows expectations outweighing perceptions, so the higher a positive gap value is, the more the service is lacking in the given attribute of service quality.

As for the interpretation of the fuzzy gap values, we can derive from the calculation method that its first parameter (l) represents the amount by which maximal performance perception outweighs minimal expectations towards this attribute in the case of a negative value (which all l parameters of gap values are), its m parameter tells us how the average expectations outweigh the average perceptions, while its r parameter shows how the maximal expectable level of performance outweighs the minimal perceived level. We can interpret these three numbers as representing gap values for „best-case”, average and „worst-case” scenarios accordingly. For example, when a passenger possesses a minimal level (according to our respondents) of expectations towards an attribute, and meets with the highest level of performance that our service can provide in that regard (again, according to respondents), then the gap this passenger perceives in this ideal scenario between their expectations and perceptions is going to be the first parameter of this fuzzy representation. Along the analysis the „spread” of these fuzzy numbers is also mentioned several times, which here just means the difference between the l and r parameters.

In tables 2 and 3 (and later in table 4 as well) the highlighted values for the dimensions are calculated as an average of the values for the two attributes in the corresponding dimension in the case of the fuzzy numbers, pure numbers and Gap 5 values as well, representing respondents’ average opinions regarding the entirety of the dimension. This method of averaging is in line with the practice suggested by the authors of the original SERVQUAL instrument.

In table 3, the defuzzified expectation, perception and gap 5 values are presented, calculated in accordance with equation (10), based on their corresponding fuzzy number equivalents. Gap 5 values show the differences between the pure number representations of expectations and perceptions. The last „Rank” column indicates the order of gap values, rank 1 representing the highest observed gap. The column displays two separate rankings, one for the service quality attributes (ranks 1 through 10) and one for the dimensions (ranks (1) through (5)). These rankings are especially useful later, when the comparison with the weighted model is made.

To aid the analysis, one-sample T-tests were conducted and confidence intervals with 5% level of significance were set up for all the gap values presented in Table 3. It is important to emphasize that the applicability of confidence intervals lies on crucial assumptions, some of which cannot be made in this setting. Although the relatively large sample size (over 30) means that the sampling distribution of the sample mean can be regarded as approximately normal, the nonrandom sampling method means that the confidence intervals

should not be generalized to the whole population, as they could be biased. This precaution is extendable to the following statistical analyses as well, namely the independent samples T tests and correlations, that will be discussed in later chapters. Calculations for these confidence intervals can be seen in appendix 2. The p values visible are always regarding the null hypothesis that the mean of the observed gap values (sample mean) is equal to 0. The overwhelming majority of these p values (with the exception of one of the attributes shown as Q10gap, which indicates the gap value for A52) is lower than a 5% level of significance, suggesting the rejection of the null hypothesis and the existence of a positive gap value.

Table 2: Fuzzy expectation, perception and fuzzy gap values

Dimension	Attribute	Fuzzy Expectation	Fuzzy Perception	Fuzzy Gap 5
D1: Tangibles		(5.23,7.08,8.94)	(3.89,5.91,7.55)	(-2.31,1.17,5.05)
	A ₁₁	(4.50,6.66,8.78)	(4.09,6.16,7.63)	(-3.13,0.50,4.69)
	A ₁₂	(5.97,7.50,9.09)	(3.69,5.66,7.47)	(-1.50,1.84,5.41)
D2: Reliability		(6.66,7.97,9.17)	(4.77,6.91,8.53)	(-1.88,1.06,4.41)
	A ₂₁	(7.03,8.22,9.28)	(5.06,7.16,8.59)	(-1.56,1.06,4.22)
	A ₂₂	(6.28,7.72,9.06)	(4.47,6.66,8.47)	(-2.19,1.06,4.59)
D3: Responsiveness		(5.67,7.25,8.77)	(3.78,5.73,7.36)	(-1.69,1.52,4.98)
	A ₃₁	(6.06,7.53,8.94)	(3.78,5.91,7.75)	(-1.69,1.63,5.16)
	A ₃₂	(5.28,6.97,8.59)	(3.78,5.56,6.97)	(-1.69,1.41,4.81)
D4: Assurance		(6.33,7.75,9.06)	(3.64,5.80,7.58)	(-1.25,1.95,5.42)
	A ₄₁	(6.72,8.06,9.09)	(3.56,5.59,7.31)	(-0.59,2.47,5.53)
	A ₄₂	(5.94,7.44,9.03)	(3.72,6.00,7.84)	(-1.91,1.44,5.31)
D5: Empathy		(5.20,6.61,8.06)	(4.42,5.94,7.30)	(-2.09,0.67,3.64)
	A ₅₁	(6.00,7.31,8.69)	(4.97,6.63,8.25)	(-2.25,0.69,3.72)
	A ₅₂	(4.41,5.91,7.44)	(3.88,5.25,6.34)	(-1.94,0.66,3.56)

Source: Own research

Explanation: Aggregated fuzzy expectation and perception values for the investigated service quality dimensions and attributes along with fuzzy gap values calculated according to equation (9). Further explanation on the individual attributes can be seen in Table 1 or in the full questionnaire in appendix 1.

Table 3: Defuzzified expectation, perception and gap values with gap value ranking

Dimension	Attribute		Expectation	Perception	Gap 5	Rank
D1: Tangibles			7.08	5.78	1.30	(3)
	A ₁₁		6.65	5.96	0.69	10
	A ₁₂		7.52	5.60	1.92	2
D2: Reliability			7.93	6.73	1.20	(4)
	A ₂₁		8.18	6.94	1.24	6
	A ₂₂		7.69	6.53	1.16	7
D3: Responsiveness			7.23	5.63	1.60	(2)
	A ₃₁		7.51	5.81	1.70	3
	A ₃₂		6.95	5.44	1.51	5
D4: Assurance			7.71	5.67	2.04	(1)
	A ₄₁		7.96	5.49	2.47	1
	A ₄₂		7.47	5.85	1.61	4
D5: Empathy			6.63	5.89	0.74	(5)
	A ₅₁		7.33	6.61	0.72	9
	A ₅₂		5.92	5.16	0.76	8

Source: Own research

Explanation: Defuzzified values of aggregated expectation, perception and gap 5 fuzzy numbers. The „Rank” column indicates the order of gap values, from highest to lowest, separately for the dimensions (in parantheses) and attributes.

In terms of the gap values themselves, we can observe the relevant differences on figure 6, depicting the expectations–perceptions gap for each of the attributes, while the expectations and perceptions themselves can be seen on figures 7 and 8 respectively, in descending order.

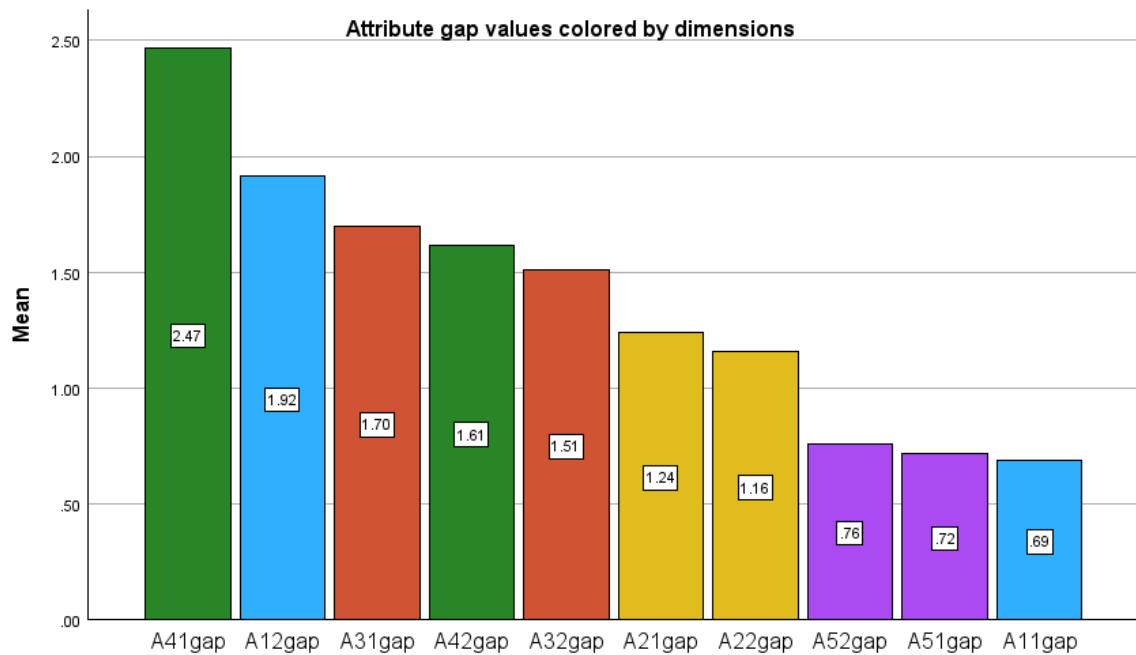


Figure 6: Attribute gap values colored by dimension

Source: Own research

Explanation: Gap 5 values of investigated attributes in descending order. Attributes are colored according to the dimension they belong to.

A₄₁ shows by far the biggest gap, suggesting that the ability of the public transport service providers to provide safety on board of their vehicles is the most lacking attribute between the ones investigated. This is both due to the high level of expectations in this regard (2nd highest overall) and to the low level of perceived safety, which is the third lowest out of all the attributes. Additionally, exploring the fuzzy values we can observe that both the expectations and perceptions regarding this attribute show a higher likelihood of being lower than the average values marked by respondents, as the m parameters of the fuzzy numbers are more distant from the l than the r parameters in both cases. The fuzzy gap value indicates that maximal perceived perceptions just barely outweigh the minimal expectations towards safety, but both the average gap and the „worst-case scenario” gap, that is, the gap between the highest expectable and lowest perceived level of performance, are the highest of all the attributes. Considering a confidence interval at the 5% level of significance, we can observe that A₄₁ is one of the only two attributes (along with A₁₂) for which the gap value is significantly higher than one. Based on all these results, safety seems to be a crucial attribute in need of managerial attention.

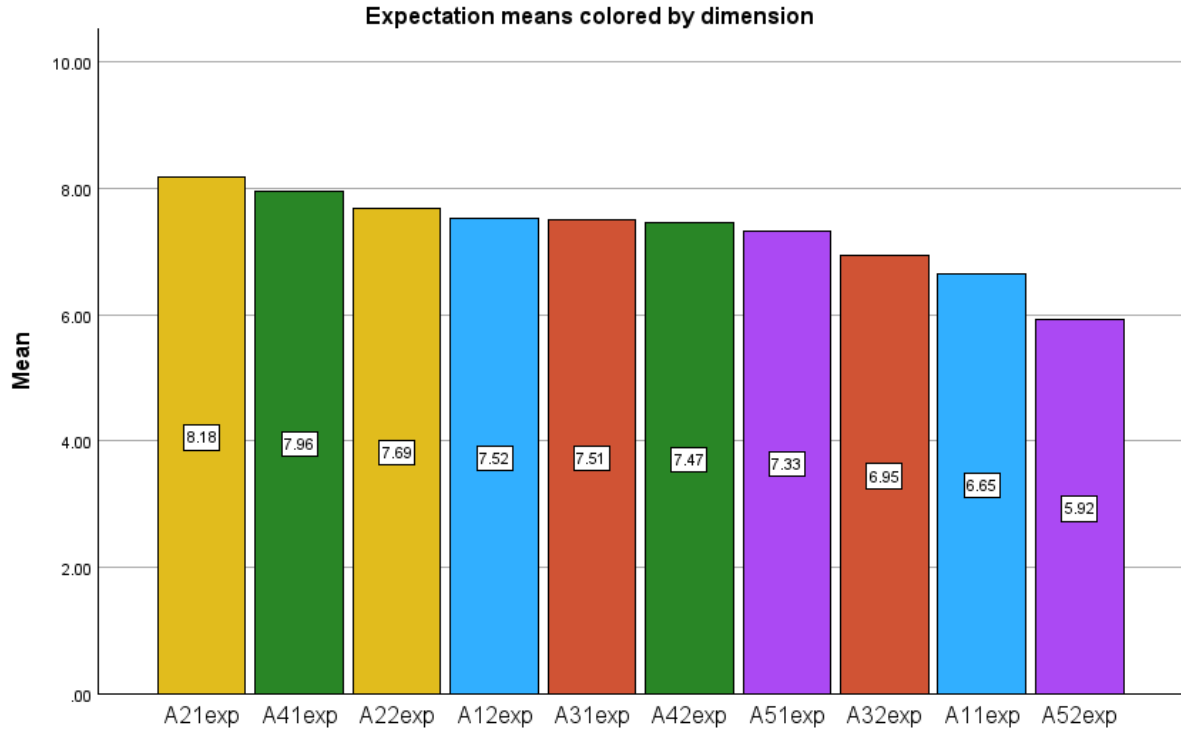


Figure 7: Average expectations colored by dimension

Source: Own research

Explanation: Average expectation values for attributes in decreasing order, colored by dimension.

The following three highest gap values correspond to attributes (A_{12} , A_{31} and A_{42}) all in the middle field both in terms of expectations and perceptions, being the 4th, 5th and 6th highest in the former, and 4th, 5th and 6th lowest in the latter respectively, in terms of defuzzified values. However, when we look at the fuzzy equivalents of these numbers, we can observe that although the numbers representing expectations are fairly symmetrical (meaning that their three parameters are of roughly equal distance to each other), the perception scores show a higher variability from the average perceived value in the negative direction, similarly to the previously discussed attribute A_{41} . Fuzzy gap values show that the „worst-case scenario” perceived gap levels for these attributes are still over 5, which, on a 10 point scale, is quite considerable, while „best-case scenario” gap values are still below 2 (in absolute value), showing that the highest perceived performance is still not much better than the lowest expectable level. The spread of these fuzzy gaps, that is, the distance between worst and best case scenarios is also relatively high, at 3th, 4th and 2nd highest respectively, suggesting a high volatility of user experiences. These attributes correspond to the

questions regarding clean and comfortable travel conditions, as well as the helpfulness, politeness and trustworthiness of employees.

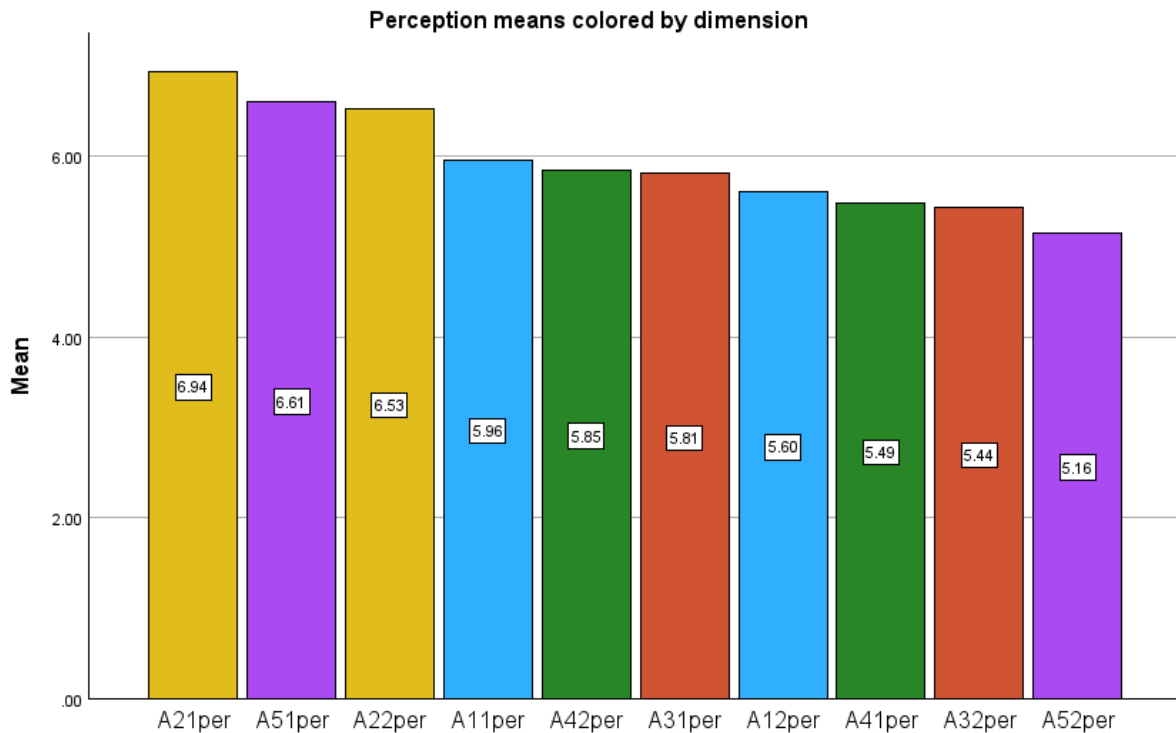


Figure 8: Average perceptions colored by dimension

Source: Own research

Explanation: Average perception values for attributes in decreasing order, colored by dimension.

Attribute A_{32} should be considered between these as well, as it shows a higher-than-average gap and displays a similar pattern regarding its fuzzy perceptions. It is interesting to note however, that this attribute, measuring opinions on fast and appropriate reactions of the service provider to unforeseen complications, is the third lowest both in terms of expectations and perceptions when assessing the crisp values. This indicates that although passengers already expect a relatively low level of performance in this regard, their expectations are still not met. Fuzzy gap values show that when the lowest expectations meet the highest performance, similarly to the previous three, an absolute gap value of 2 is still not reached, the „worst-case” gap however is slightly improved compared to those three, as it shows a gap of just over 4.8. As all these four gap values observed are still relatively high, the service provider needs to look into these factors, as they influence the consumer perception of the overall service quality significantly.

Dimensions D_2 and D_5 (reliability and empathy) colored yellow and purple respectively, show lower gap levels across their attributes. In the case of D_2 , we can observe that while the expectations regarding these attributes are generally high (1st and 3rd highest), the perceptions of the level of perceived performance in these areas (namely information availability and adherence to schedules) follows closely with similarly high rates. Fuzzy representations of the scores show the previously observed pattern of a higher likelihood of the service performance differing from the middle value in the negative direction. Fuzzy gap analysis reveals that while all fuzzy gap values are slightly skewed to the right (meaning that the average experience is more likely to deviate towards the worst-case scenario), A_{21} displays this on the highest level. Based on all this information we can propose that the adherence to—and availability and accuracy of schedules is not an outstandingly acute problem that needs immediate managerial action, but the higher tendency of worse-than-average user experiences should be looked into in the future.

In the case of D_5 attributes, it is clearly visible that the expectations of customers are generally on the lower side regarding empathy. In the case of A_{51} this is paired with the second highest perceived performance level between the attributes, which together lead to a low gap value. On the contrary, A_{52} shows the lowest level of perceived performance, suggesting that the only reason for the low gap value is the unusually low customer expectations towards this attribute, the capability of the service provider to cater to specific customer needs. This is understandable, as the role of a public transport service provider is generally less to serve these specific customer needs, and more to provide value to the customer base as a whole. It is also worth noting that the 95% confidence interval for the gap values shows that this gap is the only one not statistically different from 0 (even though it is only the third lowest), so we cannot state that there is a significant dissatisfaction regarding this attribute. The fuzzy numbers for these attributes show relative symmetry both in terms of expectations and perceptions. When analyzing their fuzzy gap values, we can observe the first and third lowest spreads between best-case and worst-case gaps suggesting a relatively low volatility in user experiences, together with the lowest gap values for worst-case scenarios and best-case gaps of around -2, that can be considered better than average.

A_{11} , the attribute assessing the need for—and presence of modern and up-to-date vehicles shows the lowest gap level. We can see that the expectations regarding this attribute are relatively low, while the perceived performance of the service provider in this regard is relatively high, which leads to an overall low gap. Although this low level suggests that this attribute is not the most relevant problem regarding overall service quality, it is worth noting

that the fuzzy perception of this attribute shows a significantly higher difference between the lowest and the average perceived performance, than between the average and the highest level. This indicates that although the average perception seems to be close to meeting customer expectations, most customers have experienced unusually low levels of performance in this area at some point. This idea is further backed by the fact that the fuzzy gap value is skewed to the right, and displays the highest spread among fuzzy attribute gaps.

Taking a closer look at the dimensional gaps, D_4 , assurance, seems to exhibit the highest discrepancies between customer expectations and perceptions with an average gap value of 2.04 between its attributes. This is both due to the 2nd highest level of expectations between the dimensions and the 2nd lowest level of perceived performance. This is the only dimension where the gap value is statistically over 1 using a 95% confidence interval, which, on a 10-point scale can be considered quite significant. A high gap value for assurance indicates that a feeling of safety and trust is not sufficiently evoked in customers by neither the environment nor the employees of the service provider. The fuzzy gap representation indicates that this is the worst performing dimension in terms of both l and r values, also showing the second highest spread between dimensions. Additionally, the fuzzy perception representations show that the perceived performance instability is the highest in this dimension (the difference between l and r values), indicating a high volatility in safety and trustworthiness perception, which is suboptimal, as these characteristics are generally preferred to be stable.

The dimension of responsiveness (D_3) is a somewhat close second with its dimensional gap value of 1.60. Although the expectations for this dimension are not outstandingly large (the crisp value score is just below the grand average for all of the expectations), it shows the lowest perception level. As this dimension reflects the ability of employees and the service provider to react to and assist with unforeseen situations, this low of a performance perception level is alarming. The lower side of its 95% confidence interval is just below 1, showing that the gap is of a significant size. The fuzzy gap displays a relatively high spread indicating volatility.

Tangibles and reliability show similar gaps but for differing reasons: while expectations towards—and perceptions of reliability are both of the highest level between the dimensions, tangibles are average in terms of both of those values. The existence of a gap in reliability is therefore mostly due to the high level of expectations, and resources might be better focused elsewhere as the performance in this field is already decent compared to other dimensions. Analyzing the fuzzy gap supports this idea, as it presents around average l and

r values, with a relatively low spread. As already discussed, however, this dimension displays the highest level of skewness, raising some potential concerns. On the other hand, taking a look at the fuzzy gap of the tangibles dimension reveals that it displays the best of the best case scenarios and the second worst of the worst case scenarios at the same time, showing the highest spread between dimensions. This indicates an unusually high volatility regarding the cleanliness, comfortableness and modernity of vehicles.

The empathy dimension shows the lowest gap value, considerably lower than all the others. This is a common phenomenon in the industry specific literature: empathy is often found to be producing relatively lower gap scores. The average expectations towards this dimension are the lowest, while the perceived performance is relatively high, resulting in a low gap value. The gap however is still statistically higher than 0 with a 95% level of confidence, suggesting that the gap still exists and is statistically significant. Its fuzzy gap shows the lowest spread and skewness along dimensions, together with the least negative worst case scenarios.

4.4. WEIGHTED GAP ANALYSIS

Following the traditional and fuzzy gap analysis, the weighted SERVQUAL model can be developed, and, perhaps most importantly, its results can be compared with those of the original model. The relevant figures for this process are available in table 4. As discussed before, importance weights are derived from the average expectations of respondents. These values were normalised (divided by the sum of all average expectations) and then multiplied by 10, resulting in importance weights around 1, that are easy to interpret and work with. As we are only considering the ranking of weighted values, the multiplication of weights with positive constants does not alter the results (even normalisation is not required, but it aids the interpretation). Gap 5 values and their rankings are identical to those already seen in table 3, while a similar ranking of the newly computed weighted gaps is added. The weighted gaps themselves are the products of corresponding gap 5 and importance values.

As it is most appropriate to consider rankings when conducting a gap analysis (as ordinal data is used), a particularly major point of interest regarding the weighted model is if and how it changes the ordering of attribute gap values, along with the order of importance between these values.

The two elements of reliability (adherence to—and availability of accurate schedules) are the first and third most important attributes, leading to the dimension of reliability showing a higher weighted gap than tangibles, contrary to the unweighted model, however, the

attributes themselves do not move up in ranking. The second most important dimension is assurance, that consists of the second (safety) and sixth (polite staff) most important attributes. This dimension still shows the highest weighted gap level, and its attributes did not change ranking either. The only other notable change is regarding A52, the service provider's ability to serve specific needs, which is weighed as by far the least important attribute, falling two places in gap value ranking, while the former last two attributes move up one place each.

Table 4: Normalised importance weight values and comparison of unweighted and weighted gaps

Dimension	Attribute	Importance weight	Gap 5	Rank	Weighted Gap 5	Weighted Rank
D1: Tangibles		0.97	1.30	(3)	1.26	(4)
	A ₁₁	0.91	0.69	10	0.62	9
	A ₁₂	1.02	1.92	2	1.96	2
D2: Reliability		1.09	1.20	(4)	1.30	(3)
	A ₂₁	1.12	1.24	6	1.39	6
	A ₂₂	1.05	1.16	7	1.22	7
D3: Responsiveness		0.99	1.60	(2)	1.59	(2)
	A ₃₁	1.03	1.70	3	1.74	3
	A ₃₂	0.95	1.51	5	1.44	5
D4: Assurance		1.06	2.04	(1)	2.16	(1)
	A ₄₁	1.10	2.47	1	2.71	1
	A ₄₂	1.01	1.61	4	1.64	4
D5: Empathy		0.90	0.74	(5)	0.67	(5)
	A ₅₁	1.00	0.72	9	0.72	8
	A ₅₂	0.81	0.76	8	0.61	10

Source: Own research

Explanation: Normalised importance weights derived from average expectations together with gap values of the unweighted (already seen in Table 3) and weighted models, and the dimensional and attribute level rankings for both sets of gap values.

The main managerial implication deduceable from the weighted model is that although reliability does not display particularly high gap values, its importance level to passengers is of such a degree that taking action to improve this dimension has to be considered. This finding is especially interesting as the conclusion regarding reliability based on the unweighted analysis was completely different, belittling the necessity of intervention. However, the most important areas of interest are still assurance and responsiveness, or, in terms of service quality attributes, safety, cleanliness and comfort, and the willingness of employees to help in cases of need.

4.5. MEAN COMPARISONS

Based on the demographic data collected and in pursuit of discovering potential connections between demographic data and measured expectations and perceptions, a series of independent samples T tests were conducted. The difference between the means of expectation values towards a dimension, as well as perception values with regards to a dimension were compared for sample groups derived from demographic characteristics. These groups were derived as follows: the prominent age group of 18–29 years old was compared to the group of all other respondents; passengers reporting daily use of PT services were compared to those travelling less often; and finally those travelling approximately less than an hour daily were compared to passengers using PT services for longer commutes. Additionally, mean values were also compared based on gender and car ownership status.

The independent samples T test using the above explained categories was chosen due to low sample sizes. Otherwise, an ANOVA analysis eliminating the need to make further demographic classifications might have been more appropriate. The nonrandom sampling method is a problem in both cases (better explained above) making the results non-generalizable.

The normality assumption was checked in all cases via the Shapiro–Wilk test and by examining Q-Q plots. The null hypothesis of a normally distributed underlying population had to be rejected in several cases; consecutively, the independent samples T tests were only conducted (and their results were only analyzed) in the cases where normality could not be rejected. The equal variances assumption was tested using Levene’s test in all cases, resulting in p values all higher than the 5% level of significance, suggesting that the null hypothesis of equal variances should not be rejected. As we have no prior hypothesis about the direction of the relationships, two-tailed p values were considered in each case.

All Shapiro–Wilk normality tests, along with independent samples T test tables where the normality assumption was not rejectable can be found in appendix 3.

When assessing the expectations and perceptions for service quality dimensions based on gender, our sample suggests statistically significant differences between expectations towards responsiveness and assurance (D3 and D4) at the 5% significance level. Based on the mean differences it appears that female respondents tend to indicate higher levels of expectations for these dimensions than males. As these are the dimensions showing the biggest gap levels, and females are overrepresented in the sample (62.5%) compared to the population, we can assume that at least some of the calculated gap level is due to sampling error. However, as sampling error affects this very statistic as well, we cannot make definite conclusions.

For age and travel frequency groups we cannot note any significant mean differences at the 5%, but not even at a potential 10% significance level. This is true to classification based on car ownership status too, which is surprising, as this demographic is often a good measure for segmentation in the literature, but the lack of significant mean differences could be due to low sample sizes and the unrepresentative sample. It is notable however that the lowest p value can be observed for the expectations of assurance (D4) in the case of car ownership (0.110), while the second highest p value is visible for the expectations on responsiveness (D3, 0.234), which, when assessing based on travel frequency groups, is actually the highest (0.125). Although these p values are too high to make any conclusions, it is interesting to see that the expectations towards these dimensions (which are also the most lacking dimensions of service quality based on their gap values) seem to be closer related to assessed demographic features than other expectations or perceptions. This idea can be further supported by investigating the mean difference between the two groups derived from the amount of average time spent travelling on PT (these were those travelling less than or more than one hour), as the p value calculated for difference between D3 expectations is 0.124, the lowest one observed.

4.6. CORRELATIONS

With the goal of discovering stochastic relationships between expectations and perceptions for the different service quality dimensions, a series of Pearson correlation coefficients were calculated. The initial hypotheses set up regarding these were that a higher expectation level in one of the dimensions is associated with a higher expectation level across all the other

dimensions, and similarly, a higher perception level in one of the dimensions is associated with higher perceptions of all the other dimensions.

Correlations between dimensional expectations were found to be generally strong (all correlation coefficients are above 0.65) and all significant even at the 1% significance level. This suggests that there is a strong positive relationship between customer expectations for the different dimensions, meaning that a higher expectation towards one of them is generally associated with higher expectations towards all the others.

Similar results can be seen addressing customer perceptions at a 5% significance level. Although these relationships are more in the moderate range, the results still suggest a moderate positive relationship between the perception of service quality of any two dimensions. A table containing all the calculated coefficients is visible in appendix 4.

5. SUMMARY, CONCLUSIONS AND FOLLOWING RESEARCH

The goal of this paper was to assess the quality of public transport services in Budapest from the point of view of customers, with a novel approach, using a methodology capable of quantifying otherwise unmeasurable characteristics. These include volatility regarding the service itself and uncertainties within the process of gathering evaluations from customers, along with the inherent variable nature of customers' opinions on the quality of service. For this purpose, a detailed review of the relevant literature is presented, and a methodology combining solutions to different shortcomings of previously utilized methods is proposed. A questionnaire in the SERVQUAL framework is developed and is highly adapted both to better fit the specific setting and to introduce fuzzy logic, which provides solutions that help in reaching the initial goals. Finally, the results of this questionnaire are presented in detail, including a crisp value and fuzzy gap analysis, the analysis and implications of a weighted model, mean comparisons between identified demographic groups and a brief mention of discovered correlations in the data.

As it is emphasized multiple times in this paper, a nonrandom sampling method was used to collect data, and therefore the sample cannot be treated as representative to the whole population of public transport users. For this reason, the following conclusions are to be interpreted for this sample only.

With the help of the gap analysis the appropriate gap 5 (expectation–perception gap) values were calculated for the five identified dimensions of service quality. All these gap values were shown to be positive and significantly different from zero, showing that the perceived service quality in Budapest is significantly lower than the expectations regarding it. Assurance and responsiveness were shown to exhibit the highest gap values, suggesting that the ability of the service provider to invoke a sense of safety and security, along with its ability to react to and help with unforeseen complications are lacking. Turning a sharp focus towards the improvement of these areas could help the service provider generate a higher level of perceived service quality. Reliability, the dimension containing the adherence to—and the availability of accurate schedules was shown to have a high importance weight for passengers, even though its displayed gap values are below average. It was also shown that women are likely to rate their expectations towards these dimensions higher than men, and that a higher expectation or perception for one of the dimensions is associated with an increase in the same statistic for all other dimensions as well.

Potential critics for both the methodology and the statistical tools utilized were discussed along the main text. The use of a more direct input system for customers to indicate their feelings results in a necessary reduction in the number of investigated service attributes. It is also visible that the relatively complicated wording and therefore high attention demand of the survey lead to a significant level of invalidly submitted answers. Setting up an appropriate fuzzy-linguistic scale with the help of a potential preliminary research assessing customers' perceptions of the linguistic scale itself could be a more appropriate and generally more useful methodology, allowing for a full-length SERVQUAL questionnaire and simplifying the completion of the survey for most respondents. Gathering data on stated levels of importance would lead to a more accurate weighted model, with the (recurring) drawback of lengthening the questionnaire. The use of a better sampling method is also required in order to make conclusions regarding the whole population. This, combined with a higher sample size would lead to a significantly more meaningful statistical analysis. For the creation of this paper however these options were not available due to resource limitations.

A future study conducted with the above explained methodology and the easily observable limitations of this research in mind could unveil a more accurate assessment of service quality, and ultimately provide a better understanding of the areas in need of improvement.

REFERENCE LIST

- Árva, G., Tóth, Z. E., Jónás, T., & Surman, V. (2019). *Will the Likert scale pass the final exam? A novel, fuzzy-number-based evaluation of supervisors' performance*. In *Proceedings of the 10th International Odyssey Conference on Economics and Business* (pp. 300–316). Opatija, Croatia.
- Babakus, E., & Boller, G. W. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24(3), 253–268. [https://doi.org/10.1016/0148-2963\(92\)90022-4](https://doi.org/10.1016/0148-2963(92)90022-4)
- Bede, B. (2013). Fuzzy Numbers. In B. Bede (Ed.), *Mathematics of Fuzzy Sets and Fuzzy Logic* (pp. 51–64). Springer. https://doi.org/10.1007/978-3-642-35221-8_4
- Benítez, J. M., Martín, J. C., & Román, C. (2007). Using fuzzy number for measuring quality of service in the hotel industry. *Tourism Management*, 28(2), 544–555. <https://doi.org/10.1016/j.tourman.2006.04.018>
- Borbás, P. (2022). The Importance of the Quality of Urban Public Transport and a Methodological Example for Weighting the Typical Quality Criteria. *Regionalnaya Ekonomika. Yug Rossii*, 2, 89–96. <https://doi.org/10.15688/re.volsu.2022.2.9>
- Chou, C.-C., Liu, L.-J., Huang, S.-F., Yih, J.-M., & Han, T.-C. (2011). An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. *Applied Soft Computing*, 11(2), 2117–2128. <https://doi.org/10.1016/j.asoc.2010.07.010>
- Cronin, Jr, J., & Taylor, S. (1992). Measuring Service Quality—A Reexamination And Extension. *The Journal of Marketing*, 56, 55–68. <https://doi.org/10.2307/1252296>
- Currie, G., & Wallis, I. (2008). Effective ways to grow urban bus markets – a synthesis of evidence. *Journal of Transport Geography*, 16(6), 419–429. <https://doi.org/10.1016/j.jtrangeo.2008.04.007>
- de Oña, J., & de Oña, R. (2015). Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches. *Transportation Science*, 49(3), 605–622. <https://doi.org/10.1287/trsc.2014.0544>
- Dombi, J., & Jónás, T. (2021). *Advances in the Theory of Probabilistic and Fuzzy Data Scientific Methods with Applications* (Vol. 814). Springer International Publishing. <https://doi.org/10.1007/978-3-030-51949-0>
- Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems: Theory and applications*. Academic Press.
- Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation*, 10(3), 21–34. <https://doi.org/10.5038/2375-0901.10.3.2>
- Echaniz, E., dell’Olio, L., & Ibeas, Á. (2018). Modelling perceived quality for urban public transport systems using weighted variables and random parameters. *Transport Policy*, 67, 31–39. <https://doi.org/10.1016/j.tranpol.2017.05.006>
- Hu, H.-H. (Sunny), Kandampully, J., & Juwaheer, T. D. (2009). Relationships and impacts of service quality, perceived value, customer satisfaction, and image: An empirical study. *The Service Industries Journal*, 29(2), 111–125. <https://doi.org/10.1080/02642060802292932>
- Hussain, B., Zefreh, M. M., & Török, Á. (2018). Designing the appropriate data collection method for public transport passenger satisfaction analysis. *International Journal for Traffic and Transport Engineering*, 8(2), 177–183. [https://doi.org/10.7708/ijtte.2018.8\(2\).03](https://doi.org/10.7708/ijtte.2018.8(2).03)

- Imam, R. (2014). Measuring Public Transport Satisfaction from User Surveys. *International Journal of Business and Management*, 9(6), p106. <https://doi.org/10.5539/ijbm.v9n6p106>
- Jain, S. K., & Gupta, G. (2004). Measuring Service Quality: Servqual vs. Servperf Scales. *Vikalpa*, 29(2), 25–38. <https://doi.org/10.1177/0256090920040203>
- Khademi-Vidra, A., Nemecz, G., & Mária Bakos, I. (2024). Satisfaction measurement in the sustainable public transport of Budapest. *Transportation Research Interdisciplinary Perspectives*, 23, 100989. <https://doi.org/10.1016/j.trip.2023.100989>
- Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. Prentice Hall PTR.
- Kokku, R., Almotawa, A., & Vijai, P. (2011). Measuring Commuters' Perception on Service Quality Using SERVQUAL in Public Transportation. *International Journal of Marketing Studies*, 3. <https://doi.org/10.5539/ijms.v3n1p21>
- Kovács, Z., Farkas, J. Z., Szigeti, C., & Harangozó, G. (2022). Assessing the sustainability of urbanization at the sub-national level: The Ecological Footprint and Biocapacity accounts of the Budapest Metropolitan Region, Hungary. *Sustainable Cities and Society*, 84, 104022. <https://doi.org/10.1016/j.scs.2022.104022>
- Kuo, M.-S., Wu, J.-W., & Pei, L. (2007). A soft computing method for selecting evaluation criteria of service quality. *Applied Mathematics and Computation*, 189(1), 241–254. <https://doi.org/10.1016/j.amc.2006.11.084>
- Li, Q. (2013). A novel Likert scale based on fuzzy sets theory. *Expert Systems with Applications*, 40(5), 1609–1618. <https://doi.org/10.1016/j.eswa.2012.09.015>
- Liou, T., & Chen, C. (2006). Subjective appraisal of service quality using fuzzy linguistic assessment. *International Journal of Quality & Reliability Management*, 23(8), 928–943. <https://doi.org/10.1108/02656710610688149>
- Lupo, T. (2013a). A fuzzy ServQual based method for reliable measurements of education quality in Italian higher education area. *Expert Systems with Applications*, 40(17), 7096–7110. <https://doi.org/10.1016/j.eswa.2013.06.045>
- Lupo, T. (2013b). Handling stakeholder uncertain judgments in strategic transport service analyses. *Transport Policy*, 29, 54–63. <https://doi.org/10.1016/j.tranpol.2013.04.002>
- Mikhaylov, A., Gumenyuk, I., & Mikhaylova, A. (2015). The SERVQUAL model in measuring service quality of public transportation: Evidence from Russia. *Quality - Access to Success*, 16, 78–83.
- Moslem, S., & Duleba, S. (2018). *Analytic Hierarchy Process Evaluation of Public Transport System in Budapest, Hungary*. In *Proceedings of the Fourth International Conference on Traffic and Transport Engineering* (pp. 660–665). Belgrade, Serbia.
- Ojo, T., & Adom, A. (2017). *Managing Service Quality in Public Transport through Servqual Model: A Literature Review*.
- Parasuraman, A. P., Zeithaml, V., & Berry, L. (1985). A Conceptual Model of Service Quality and its Implication for Future Research (SERVQUAL). *The Journal of Marketing*, 49, 41–50. <https://doi.org/10.2307/1251430>
- Parasuraman, A. P., Zeithaml, V., & Berry, L. (1988). SERVQUAL: A multiple- Item Scale for measuring consumer perceptions of service quality. *Journal of Retailing*.
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234–281. [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5)

- Tarn, J. L. M. (1999). The Effects of Service Quality, Perceived Value and Customer Satisfaction on Behavioral Intentions. *Journal of Hospitality & Leisure Marketing*, 6(4), 31–43. https://doi.org/10.1300/J150v06n04_04
- Tóth, Z. E., Jónás, T., & Dénes, R. V. (2019). Applying flexible fuzzy numbers for evaluating service features in healthcare – patients and employees in the focus. *Total Quality Management & Business Excellence*, 30(sup1), S240–S254. <https://doi.org/10.1080/14783363.2019.1665863>
- Wahyudi, R. D. (2017). Investigating the role of Fuzzy as confirmatory tool for service quality assessment (Case study: Comparison of Fuzzy SERVQUAL and SERVQUAL in hotel service evaluation). *IOP Conference Series: Materials Science and Engineering*, 273, 012032. <https://doi.org/10.1088/1757-899X/245/1/012032>
- Wang, Y. L., Luor, T., Luarn, P., & Lu, H. (2015). Contribution and Trend to Quality Research—A literature review of SERVQUAL model from 1998 to 2013. *Informatica Economica*, 19(1/2015), 34–45. <https://doi.org/10.12948/issn14531305/19.1.2015.03>
- Wang, Y.-M., Yang, J.-B., Xu, D.-L., & Chin, K.-S. (2006). On the centroids of fuzzy numbers. *Fuzzy Sets and Systems*, 157(7), 919–926. <https://doi.org/10.1016/j.fss.2005.11.006>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)

APPENDICES

APPENDIX 1: IMPLEMENTED QUESTIONNAIRE (HUNGARIAN)

A budapesti tömegközlekedési szolgáltatás minőségének felmérése

Kedves Kitöltő!

Láng Kristóf vagyok, az ELTE GTK pénzügy és számvitel szakos, harmadéves hallgatója. Szakdolgozatomban egy újszerű módszerrel vizsgálom a budapesti tömegközlekedés minőségét felhasználói visszajelzések alapján.

A kérdőív kitöltése anonim, körülbelül 10 percet vesz igénybe. Kitöltését előre is köszönöm, ezzel nagyban segíti a munkámat.

ELSŐ SZAKASZ

Az első szakaszban a kérdések egy tömegközlekedési vállalattal kapcsolatos elvárásaira vonatkoznak. Az alapján válaszoljon, hogy Ön szerint egy tömegközlekedési szolgáltatónak milyen mértékben kellene az adott állításnak megfelelnie. Minden állításhoz három különböző 10 fokú skála tartozik, ahol az 1 a lehető legrosszabb, a 10 pedig a lehető legjobb teljesítményt jelöli az adott állítással kapcsolatban. Az első skálán jelölje a legkisebb olyan szintet, amit elfogadhatónak tart, a másodikon jelölje az általános elvárásainak megfelelő számot, míg az utolsón jelölje az Ön szerint maximális elvárható teljesítményt az állítással kapcsolatban. Kérem, hogy a minimális elvárásánál ne jelöljön meg alacsonyabb szintet általánosként, illetve a maximális elvárása se legyen alacsonyabb az általánosnál!

Egy tömegközlekedési szolgáltató...

...modern, korszerű járműveket kellene, hogy fenntartson.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...tisztaságot és kényelmet kellene, hogy biztosítson a járművein.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...pontos és elérhető információt kellene, hogy biztosítson a megállókkal és a menetrendekkel kapcsolatban.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...a menetrendeket pontosan be kellene, hogy tartsa.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...alkalmazottainak segítőkésznek kellene lennie.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...gyorsan és megfelelően kellene, hogy tudjon reagálni váratlan problémákra.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...képes kellene, hogy legyen biztonságos környezetet fenntartani a járatain.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...alkalmazottainak udvariasnak és megbízhatónak kellene lennie.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...járatai és megállói elhelyezkedését optimálisan, a felhasználók érdekeit figyelembe véve

kellene, hogy megtervezze.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

...a felhasználói különleges igényeit szem előtt kellene, hogy tartsa.

1 2 3 4 5 6 7 8 9 10

Legkisebb elfogadható szint

Általános elvárás

Legnagyobb elvárható szint

MÁSODIK SZAKASZ

A második szakaszban a kérdések a budapesti tömegközlekedéssel kapcsolatos érzéseire vonatkoznak.. Az alapján válaszoljon, hogy Ön milyen mértékben ért egyet az állítással a budapesti tömegközlekedésre vonatkozóan. Minden állításhoz három különböző 10 fokú skála tartozik, ahol az 1 a lehető legrosszabb, a 10 pedig a lehető legjobb teljesítményt jelöli az adott állítással kapcsolatban. Az első skálán jelölje a szintet, aminél nem észlelte még rosszabbnak, a másodikon jelölje az átlagos tapasztalatainak megfelelő számot, míg az utolsón jelölje az értéket, aminél jobbnak még nem tapasztalta a BKK teljesítményét az állítással kapcsolatban. Kérem, hogy a minimális értéknél ne jelöljön meg alacsonyabb szintet általánosként, illetve a maximális tapasztalata se legyen alacsonyabb az általánosnál!

A BKK (Budapesti Közlekedési Központ)...

...modern, korszerű járműveket tart fent.

1 2 3 4 5 6 7 8 9 10

Legalacsonyabb észlelt szint

Általános tapasztalat

Legnagyobb észlelt szint

...tisztaságot és kényelmet biztosít a járművein.

1 2 3 4 5 6 7 8 9 10

Legalacsonyabb észlelt szint

Általános tapasztalat

Legnagyobb észlelt szint

...pontos és elérhető információt biztosít a megállókkal és a menetrendekkel kapcsolatban.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...a menetrendeket pontosan betartja.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...alkalmazottai segítőkészek.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...gyorsan és megfelelően tud reagálni váratlan problémákra.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...képes biztonságos környezetet fenntartani a járatain.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...alkalmazottai udvariasak és megbízhatóak.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...járatái és megállóinak elhelyezkedése optimálisan, az Ön érdekeit figyelembe véve van megtervezve.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

...az Ön különleges igényeit szem előtt tudja tartani.

	1	2	3	4	5	6	7	8	9	10
Legalacsonyabb észlelt szint										
Általános tapasztalat										
Legnagyobb észlelt szint										

Demográfiai adatok

Neme

- Férfi
- Nő
- Egyéb / nem kíván válaszolni

Kora

- 18 év alatt
- 18-29
- 30-39

- 40-49
- 50-59
- 60-69
- 69 év felett

Milyen gyakran használ tömegközlekedést?

- Szinte minden nap
- Hetente többször
- Havonta többször
- Ennél ritkábban

Egy átlagos napon körülbelül mennyi időt tölt tömegközlekedésen?

- Kevesebb, mint fél órát
- Fél és egy óra között
- Egy és másfél óra között
- Másfél és két óra között
- Több, mint két órát

Van-e autója?

- Igen
- Nem

APPENDIX 2: ONE SAMPLE T TESTS

One-Sample Test

Test Value = 0

	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
Q1gap	2.584	31	.007	.015	.68750	.1449	1.2301
Q2gap	5.945	31	<.001	<.001	1.91667	1.2591	2.5742
Q3gap	4.483	31	<.001	<.001	1.23958	.6757	1.8035
Q4gap	3.552	31	<.001	.001	1.15625	.4924	1.8201
Q5gap	4.721	31	<.001	<.001	1.69792	.9644	2.4314
Q6gap	4.522	31	<.001	<.001	1.51042	.8292	2.1916
Q7gap	5.447	31	<.001	<.001	2.46875	1.5444	3.3931
Q8gap	4.415	31	<.001	<.001	1.61458	.8687	2.3605
Q9gap	2.655	31	.006	.012	.71875	.1667	1.2708
Q10gap	1.871	31	.035	.071	.76042	-.0684	1.5893

One-Sample Test

Test Value = 0

	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
D1Gap	5.523	31	<.001	<.001	1.30208	.8213	1.7829
D2Gap	4.740	31	<.001	<.001	1.19792	.6825	1.7133
D3Gap	5.331	31	<.001	<.001	1.60417	.9905	2.2179
D4Gap	5.474	31	<.001	<.001	2.04167	1.2810	2.8024
D5Gap	2.574	31	.008	.015	.73958	.1536	1.3256

APPENDIX 3: NORMALITY AND INDEPENDENT SAMPLES T TESTS

5.1.1. Based on gender

Tests of Normality							
Gender	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Sig.
	Statistic	df	Sig.	Statistic	df	Sig.	
D1exp	Male	.119	11	.200 [*]	.980	11	.964
	Female	.190	20	.057	.915	20	.079
D2exp	Male	.172	11	.200 [*]	.916	11	.289
	Female	.126	20	.200 [*]	.898	20	.038
D3exp	Male	.152	11	.200 [*]	.938	11	.501
	Female	.106	20	.200 [*]	.982	20	.956
D4exp	Male	.175	11	.200 [*]	.916	11	.288
	Female	.172	20	.124	.932	20	.172
D5exp	Male	.150	11	.200 [*]	.958	11	.746
	Female	.101	20	.200 [*]	.981	20	.947
D1per	Male	.241	11	.074	.929	11	.400
	Female	.118	20	.200 [*]	.969	20	.736
D2per	Male	.188	11	.200 [*]	.916	11	.290
	Female	.189	20	.060	.866	20	.010
D3per	Male	.105	11	.200 [*]	.973	11	.918
	Female	.153	20	.200 [*]	.958	20	.498
D4per	Male	.153	11	.200 [*]	.953	11	.686
	Female	.127	20	.200 [*]	.971	20	.777
D5per	Male	.104	11	.200 [*]	.963	11	.808
	Female	.110	20	.200 [*]	.974	20	.843

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Independent Samples Test											
Levene's Test for Equality of Variances				t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
D1exp	Equal variances assumed	.020	.887	-1.466	30	.077	.153	-.80000	.54583	-1.91473	.31473
	Equal variances not assumed			-1.392	19.740	.090	.179	-.80000	.57477	-1.99996	.39996
D3exp	Equal variances assumed	.440	.512	-3.344	30	.001	.002	-1.65556	.49510	-2.66668	-.64444
	Equal variances not assumed			-3.117	18.594	.003	.006	-1.65556	.53118	-2.76898	-.54213
D4exp	Equal variances assumed	2.428	.130	-2.557	30	.008	.016	-1.34167	.52475	-2.41335	-.26998
	Equal variances not assumed			-2.312	16.925	.017	.034	-1.34167	.58024	-2.56627	-.11706
D5exp	Equal variances assumed	1.904	.178	-1.591	30	.061	.122	-.91111	.57283	-2.08098	.25876
	Equal variances not assumed			-1.486	18.733	.077	.154	-.91111	.61314	-2.19567	.37344
D1per	Equal variances assumed	1.182	.286	.760	30	.227	.453	.30556	.40210	-.51565	1.12676
	Equal variances not assumed			.660	14.993	.260	.520	.30556	.46328	-.68193	1.29304
D3per	Equal variances assumed	.004	.948	.476	30	.319	.637	.24444	.51302	-.80329	1.29218
	Equal variances not assumed			.465	21.504	.323	.647	.24444	.52611	-.84809	1.33698
D4per	Equal variances assumed	.147	.704	.890	30	.190	.380	.52500	.58968	-.67929	1.72929
	Equal variances not assumed			.914	25.253	.185	.369	.52500	.57425	-.65710	1.70710
D5per	Equal variances assumed	1.160	.290	.080	30	.468	.937	.05000	.62442	-1.22524	1.32524
	Equal variances not assumed			.075	18.540	.471	.941	.05000	.67055	-1.35585	1.45585

5.1.2. Based on age groups

Tests of Normality							
AgeGroups	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Sig.
	Statistic	df	Sig.	Statistic	df	Sig.	
D1exp	1.00	.139	18	.200 [*]	.956	18	.520
	2.00	.185	13	.200 [*]	.929	13	.329
D2exp	1.00	.133	18	.200 [*]	.913	18	.096
	2.00	.212	13	.115	.882	13	.075
D3exp	1.00	.143	18	.200 [*]	.958	18	.570
	2.00	.157	13	.200 [*]	.948	13	.567
D4exp	1.00	.143	18	.200 [*]	.932	18	.211
	2.00	.240	13	.039	.757	13	.002
D5exp	1.00	.095	18	.200 [*]	.988	18	.996
	2.00	.114	13	.200 [*]	.953	13	.641
D1per	1.00	.170	18	.180	.913	18	.097
	2.00	.248	13	.028	.881	13	.073
D2per	1.00	.152	18	.200 [*]	.939	18	.273
	2.00	.128	13	.200 [*]	.947	13	.552
D3per	1.00	.119	18	.200 [*]	.963	18	.665
	2.00	.149	13	.200 [*]	.906	13	.164
D4per	1.00	.106	18	.200 [*]	.981	18	.958
	2.00	.195	13	.190	.944	13	.514
D5per	1.00	.113	18	.200 [*]	.953	18	.479
	2.00	.128	13	.200 [*]	.951	13	.620

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Independent Samples Test											
Levene's Test for Equality of Variances						t-test for Equality of Means					
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
D1 exp	Equal variances assumed	.044	.835	.058	30	.477	.954	.03239	.55693	-1.10502	1.16980
	Equal variances not assumed			.059	26.765	.477	.954	.03239	.55176	-1.10020	1.16498
D2 exp	Equal variances assumed	1.697	.203	-.471	30	.321	.641	-.28677	.60931	-1.53114	.95759
	Equal variances not assumed			-.503	29.957	.309	.619	-.28677	.57049	-1.45194	.87840
D3 exp	Equal variances assumed	2.717	.110	-.845	30	.202	.405	-.47773	.56510	-1.63182	.67636
	Equal variances not assumed			-.907	29.994	.186	.372	-.47773	.52698	-1.55397	.59851
D5 exp	Equal variances assumed	.902	.350	-.489	30	.314	.629	-.28610	.58565	-1.48216	.90997
	Equal variances not assumed			-.510	29.163	.307	.614	-.28610	.56122	-1.43364	.86144
D1 per	Equal variances assumed	.055	.816	-.656	30	.259	.517	-.26046	.39733	-1.07191	.55099
	Equal variances not assumed			-.634	22.896	.266	.532	-.26046	.41059	-1.11004	.58913
D2 per	Equal variances assumed	.969	.333	-1.503	30	.072	.143	-.79285	.52766	-1.87048	.28478
	Equal variances not assumed			-1.554	28.644	.066	.131	-.79285	.51011	-1.83670	.25101
D3 per	Equal variances assumed	.011	.918	-1.050	30	.151	.302	-.52362	.49853	-1.54175	.49451
	Equal variances not assumed			-1.056	26.425	.150	.301	-.52362	.49584	-1.54202	.49479
D4 per	Equal variances assumed	.006	.941	-1.384	30	.088	.177	-.79015	.57095	-1.95618	.37589
	Equal variances not assumed			-1.398	26.831	.087	.174	-.79015	.56521	-1.95021	.36991
D5 per	Equal variances assumed	.123	.728	-.598	30	.277	.555	-.36572	.61194	-1.61548	.88403
	Equal variances not assumed			-.588	24.483	.281	.562	-.36572	.62171	-1.64752	.91608

5.1.3. Based on PT travel frequency groups

Tests of Normality							
TravelFreqGroups	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Sig.
	Statistic	df	Sig.	Statistic	df	Sig.	
D1exp	1.00	.143	18	.200 [*]	.928	18	.181
	2.00	.152	13	.200 [*]	.962	13	.778
D2exp	1.00	.139	18	.200 [*]	.916	18	.110
	2.00	.161	13	.200 [*]	.911	13	.191
D3exp	1.00	.105	18	.200 [*]	.957	18	.554
	2.00	.186	13	.200 [*]	.950	13	.599
D4exp	1.00	.232	18	.011	.885	18	.032
	2.00	.264	13	.014	.869	13	.050
D5exp	1.00	.094	18	.200 [*]	.968	18	.758
	2.00	.091	13	.200 [*]	.968	13	.864
D1per	1.00	.241	18	.007	.893	18	.044
	2.00	.188	13	.200 [*]	.941	13	.464
D2per	1.00	.154	18	.200 [*]	.942	18	.312
	2.00	.173	13	.200 [*]	.932	13	.367
D3per	1.00	.122	18	.200 [*]	.975	18	.888
	2.00	.169	13	.200 [*]	.911	13	.192
D4per	1.00	.146	18	.200 [*]	.967	18	.738
	2.00	.138	13	.200 [*]	.963	13	.806
D5per	1.00	.106	18	.200 [*]	.975	18	.880
	2.00	.109	13	.200 [*]	.981	13	.984

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Independent Samples Test											
Levene's Test for Equality of Variances				t-test for Equality of Means							
		F	Sig.	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
										Lower	Upper
D1exp	Equal variances assumed	1.028	.319	-.723	30	.238	.475	-.39946	.55217	-1.52714	.72822
	Equal variances not assumed			-.706	23.642	.244	.487	-.39946	.56605	-1.56867	.76975
D2exp	Equal variances assumed	.343	.563	-.978	30	.168	.336	-.58907	.60202	-1.81856	.64042
	Equal variances not assumed			-1.014	28.761	.160	.319	-.58907	.58092	-1.77761	.59947
D3exp	Equal variances assumed	.077	.783	-1.577	30	.063	.125	-.86640	.54948	-1.98858	.25578
	Equal variances not assumed			-1.603	27.371	.060	.120	-.86640	.54045	-1.97462	.24182
D5exp	Equal variances assumed	.002	.961	-.267	30	.396	.792	-.15655	.58728	-1.35594	1.04285
	Equal variances not assumed			-.264	24.915	.397	.794	-.15655	.59390	-1.37992	1.06683
D2per	Equal variances assumed	.579	.453	-1.503	30	.072	.143	-.79285	.52766	-1.87048	.28478
	Equal variances not assumed			-1.545	28.215	.067	.134	-.79285	.51333	-1.84399	.25829
D3per	Equal variances assumed	.296	.590	-.394	30	.348	.696	-.19973	.50630	-1.23373	.83427
	Equal variances not assumed			-.385	23.580	.352	.704	-.19973	.51937	-1.27268	.87322
D4per	Equal variances assumed	1.849	.184	.456	30	.326	.651	.26788	.58686	-.93064	1.46641
	Equal variances not assumed			.432	20.882	.335	.670	.26788	.62012	-1.02216	1.55793
D5per	Equal variances assumed	.231	.634	.283	30	.389	.779	.17409	.61475	-1.08141	1.42958
	Equal variances not assumed			.290	27.878	.387	.774	.17409	.60078	-1.05679	1.40497

5.1.4. Based on groups by average time spent on PT

Tests of Normality							
TimeSpentGroups	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Sig.
	Statistic	df	Sig.	Statistic	df	Sig.	
D1exp	1.00	.131	16	.200 [*]	.945	16	.414
	2.00	.184	15	.184	.937	15	.347
D2exp	1.00	.190	16	.126	.824	16	.006
	2.00	.126	15	.200 [*]	.927	15	.242
D3exp	1.00	.200	16	.087	.891	16	.058
	2.00	.095	15	.200 [*]	.978	15	.950
D4exp	1.00	.265	16	.004	.780	16	.002
	2.00	.225	15	.040	.923	15	.216
D5exp	1.00	.159	16	.200 [*]	.964	16	.734
	2.00	.163	15	.200 [*]	.923	15	.210
D1per	1.00	.187	16	.139	.941	16	.365
	2.00	.194	15	.135	.919	15	.187
D2per	1.00	.147	16	.200 [*]	.938	16	.322
	2.00	.192	15	.141	.909	15	.130
D3per	1.00	.086	16	.200 [*]	.956	16	.597
	2.00	.159	15	.200 [*]	.973	15	.902
D4per	1.00	.137	16	.200 [*]	.964	16	.731
	2.00	.277	15	.003	.875	15	.040
D5per	1.00	.121	16	.200 [*]	.946	16	.423
	2.00	.150	15	.200 [*]	.951	15	.544

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Independent Samples Test											
Levene's Test for Equality of Variances				t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
D1exp	Equal variances assumed	.998	.326	1.024	29	.157	.314	.56667	.55343	-.56523	1.69856
	Equal variances not assumed			1.035	27.164	.155	.310	.56667	.54749	-.55638	1.68971
D3exp	Equal variances assumed	1.287	.266	1.442	29	.080	.160	.80069	.55512	-.33464	1.93603
	Equal variances not assumed			1.462	25.889	.078	.156	.80069	.54761	-.32517	1.92655
D5exp	Equal variances assumed	.288	.596	.959	29	.173	.346	.56319	.58735	-.63808	1.76447
	Equal variances not assumed			.960	28.985	.172	.345	.56319	.58649	-.63634	1.76273
D1per	Equal variances assumed	.554	.463	.732	29	.235	.470	.29444	.40242	-.52860	1.11749
	Equal variances not assumed			.741	26.633	.233	.465	.29444	.39761	-.52191	1.11080
D2per	Equal variances assumed	.016	.902	.169	29	.433	.867	.09375	.55392	-1.03914	1.22664
	Equal variances not assumed			.170	28.854	.433	.866	.09375	.55139	-1.03421	1.22171
D3per	Equal variances assumed	1.331	.258	.894	29	.189	.378	.45486	.50854	-.58522	1.49494
	Equal variances not assumed			.902	27.988	.187	.375	.45486	.50424	-.57804	1.48776
D5per	Equal variances assumed	.459	.504	-.385	29	.352	.703	-.23750	.61752	-1.50047	1.02547
	Equal variances not assumed			-.387	28.722	.351	.702	-.23750	.61414	-1.49409	1.01909

5.1.5. Based on car ownership status

Tests of Normality							
Do you own a car?		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
D1exp	Yes	.103	16	.200 [*]	.978	16	.941
	No	.143	15	.200 [*]	.922	15	.207
D2exp	Yes	.122	16	.200 [*]	.949	16	.474
	No	.182	15	.196	.885	15	.056
D3exp	Yes	.138	16	.200 [*]	.969	16	.826
	No	.126	15	.200 [*]	.946	15	.470
D4exp	Yes	.171	16	.200 [*]	.906	16	.099
	No	.187	15	.167	.903	15	.107
D5exp	Yes	.131	16	.200 [*]	.962	16	.707
	No	.146	15	.200 [*]	.968	15	.820
D1per	Yes	.232	16	.021	.883	16	.044
	No	.254	15	.010	.842	15	.013
D2per	Yes	.137	16	.200 [*]	.928	16	.229
	No	.158	15	.200 [*]	.881	15	.050
D3per	Yes	.139	16	.200 [*]	.963	16	.712
	No	.105	15	.200 [*]	.964	15	.757
D4per	Yes	.124	16	.200 [*]	.969	16	.824
	No	.130	15	.200 [*]	.985	15	.993
D5per	Yes	.130	16	.200 [*]	.973	16	.886
	No	.155	15	.200 [*]	.956	15	.630

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Independent Samples Test											
Levene's Test for Equality of Variances				t-Test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
D1exp	Equal variances assumed	.011	.916	.808	30	.213	.425	.43750	.54123	-.66783	1.54283
	Equal variances not assumed			.808	29.921	.213	.425	.43750	.54123	-.66795	1.54295
D2exp	Equal variances assumed	1.274	.268	.646	30	.262	.523	.38542	.59657	-.83294	1.60377
	Equal variances not assumed			.646	27.451	.262	.524	.38542	.59657	-.83770	1.60854
D3exp	Equal variances assumed	.693	.412	1.216	30	.117	.234	.66667	.54830	-.45312	1.78645
	Equal variances not assumed			1.216	28.853	.117	.234	.66667	.54830	-.45499	1.78832
D4exp	Equal variances assumed	1.304	.263	1.649	30	.055	.110	.88542	.53691	-.21110	1.98194
	Equal variances not assumed			1.649	28.725	.055	.110	.88542	.53691	-.21315	1.98398
D5exp	Equal variances assumed	.001	.976	.144	30	.443	.886	.08333	.57735	-1.09577	1.26244
	Equal variances not assumed			.144	29.992	.443	.886	.08333	.57735	-1.09579	1.26245
D3per	Equal variances assumed	.334	.567	.251	30	.402	.804	.12500	.49809	-.89223	1.14223
	Equal variances not assumed			.251	29.330	.402	.804	.12500	.49809	-.89320	1.14320
D4per	Equal variances assumed	.061	.807	-.857	30	.199	.398	-.48958	.57150	-1.65674	.67758
	Equal variances not assumed			-.857	29.695	.199	.398	-.48958	.57150	-1.65725	.67808
D5per	Equal variances assumed	.379	.543	-.872	30	.195	.390	-.52083	.59713	-1.74034	.69868
	Equal variances not assumed			-.872	28.913	.195	.390	-.52083	.59713	-1.74227	.70060

APPENDIX 4: CORRELATIONS

		Correlations									
		D1exp	D2exp	D3exp	D4exp	D5exp	D1per	D2per	D3per	D4per	D5per
D1exp	Pearson Correlation	1	.706**	.823**	.706**	.664**	.521**	.392*	.312	.317	.376*
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	.002	.027	.082	.078	.034
	N	32	32	32	32	32	32	32	32	32	32
D2exp	Pearson Correlation	.706**	1	.802**	.761**	.708**	.324	.597**	.374*	.345	.361*
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	.070	<.001	.035	.053	.043
	N	32	32	32	32	32	32	32	32	32	32
D3exp	Pearson Correlation	.823**	.802**	1	.859**	.735**	.432*	.517**	.339	.199	.365*
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	.013	.002	.058	.274	.040
	N	32	32	32	32	32	32	32	32	32	32
D4exp	Pearson Correlation	.706**	.761**	.859**	1	.757**	.399*	.471**	.352*	.114	.411*
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	.024	.007	.048	.534	.019
	N	32	32	32	32	32	32	32	32	32	32
D5exp	Pearson Correlation	.664**	.708**	.735**	.757**	1	.511**	.535**	.404*	.304	.513**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		.003	.002	.022	.091	.003
	N	32	32	32	32	32	32	32	32	32	32
D1per	Pearson Correlation	.521**	.324	.432*	.399*	.511**	1	.636**	.727**	.647**	.542**
	Sig. (2-tailed)	.002	.070	.013	.024	.003		<.001	<.001	<.001	.001
	N	32	32	32	32	32	32	32	32	32	32
D2per	Pearson Correlation	.392*	.597**	.517**	.471**	.535**	.636**	1	.651**	.491**	.515**
	Sig. (2-tailed)	.027	<.001	.002	.007	.002	<.001		<.001	.004	.003
	N	32	32	32	32	32	32	32	32	32	32
D3per	Pearson Correlation	.312	.374*	.339	.352*	.404*	.727**	.651**	1	.805**	.478**
	Sig. (2-tailed)	.082	.035	.058	.048	.022	<.001	<.001		<.001	.006
	N	32	32	32	32	32	32	32	32	32	32
D4per	Pearson Correlation	.317	.345	.199	.114	.304	.647**	.491**	.805**	1	.425*
	Sig. (2-tailed)	.078	.053	.274	.534	.091	<.001	.004	<.001		.015
	N	32	32	32	32	32	32	32	32	32	32
D5per	Pearson Correlation	.376*	.361*	.365*	.411*	.513**	.542**	.515**	.478**	.425*	1
	Sig. (2-tailed)	.034	.043	.040	.019	.003	.001	.003	.006	.015	
	N	32	32	32	32	32	32	32	32	32	32

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).