DOES WEIGHTED AVERAGE REALLY WORK?

Grzegorz Grela
Maria Curie-Skłodowska University, Poland
grzegorz.grela@umcs.lublin.pl

Abstract:
Since the eighties of the 20th century more and more computers were owned by entrepreneurs. Managers received very powerful tool in decision making process. It can be assumed that the biggest challenge for managers are multi-dimensional decisions, where each dimension may have different importance. One of the most common tasks, in this area, is to identify among many similar objects one, that best meets the requirements. This could be for example: selecting the project that best meets the established criteria, selecting the best candidates for a particular position or reward.

The purpose of this article is to attempt to answer the question whether the use of the weighted average results in significant differences compared with the arithmetic mean?The author used the Monte Carlo method and basic statistical analysis methods. Results of this study showed when weight in average has real impact on the final result.

Keywords:information technology, weighted average disadvantages, service quality.

1. INTRODUCTION

Since the eighties of the 20th century more and more computers were owned by entrepreneurs. Managers received very powerful tool in decision making process. It can be assumed that the biggest challenge for managers are multi-dimensional decisions, where each dimension may have different importance. One of the most common tasks, in this area, is to identify among many similar objects one, that best meets the requirements. This could be for example: selecting the project that best meets the established criteria, selecting the best candidates for a particular position or reward.

The computer is able to calculate very quickly, but it counts only what user gives it to count, and only in a way defined by the user. As M. Rosser put it: “perhaps the most important point which has to be made is that calculators and computers can only calculate what they are told to. They are machines that can perform arithmetic computations much faster than you can do by hand, and this speed does indeed make them very useful tools. However, if you feed in useless information you will get useless information back - hence the well-known phrase ‘rubbish in, rubbish out’” (Rosser, 2003, p.3). To make things worse it may happen that, despite the useful data in the input, through incorrect algorithm, one may also get ‘rubbish out’. When comparing many objects on computer the key is an aggregate function used in algorithm. Aggregate function is used to obtain single number from many variables. In practical terms the weighted average is often used.

This article is an attempt to answer the question of whether, and if so, in what circumstances the results obtained by using the weighted average differ from the results obtained using the arithmetic mean.

2. SELECTED USES OF WEIGHTED AVERAGE

The weighted average is very often used by both professional researchers and in amateur applications. The EBSKO database contain more than 21,000 peer reviewed scholarly articles including the phrase "weighted mean" and almost 70,000 peer reviewed scholarly articles with the phrase "weighted average". Upon entering the phrase "weighted mean" into Google, we get 1,090,000 results, whereas for "weighted average" we get 20,300,000. The use of weighted average has many advantages, such as: easy to count and understand, the result is the same scale as the component variables, offers the possibility of differentiating weights of the dimensions. Unfortunately, this function has some serious disadvantages, the main one is that, although the weights differ significantly, in certain circumstances, the result obtained by using the weighted average may slightly different from the result obtained using a simple average.

One example of the use of the weighted average in studies is the SERVQUAL method, used to assess the quality of services. This method was developed by A. Parasuraman, V. A. Zeithaml and L. Berry in the first half of the eighties of the last century. The authors of this method have developed a model of five gaps. Each of the gaps signifies the potential mismatch between the actual and the diagnosed state. The first gap exists between customer expectations regarding the service and perceptions of the management of these expectations. The second gap is the discrepancies between the management's perception of the customers' expectations regarding and the service's design specifications. The third gap is the discrepancy between the guidelines and the service actually delivered. The fourth gap results from the difference between the service actually provided and the service presented in the communication with the customer. The fifth gap between the perceived and the expected quality has been used to measure the quality. The model assumes the existence of 5 dimensions by which customers perceive the quality of services, these are (Parasuraman & Zeithaml & Berry, 1988, p. 23):

- tangibles: physical facilities, equipment, and appearance of personnel,
- reliability: ability to perform the promised service dependably and accurately,
- responsiveness: willingness to help customers and provide prompt service,
- assurance: knowledge and courtesy of employees and their ability to inspire trust and confidence,
- empathy: caring, individualized attention the firm provides its customers.

The survey questionnaire contains 22 questions. The Servqual mathematical model can be written according to the following formula, where I_i is the importance weight on dimension I, P_j is the respondents' score onperception question j, and E_j is the respondent's score on expectations question j.(Paul, 2003, p.93).

$$SQ = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n}[I_i(P_j - E_j)]$$

F. Buttle noticed that the "analysis of SERVQUAL data can take several forms: item-by-item analysis (e.g. P1 – E1, P2 – E2); dimension-by-dimension analysis (e.g. (P1 + P2 + P3 + P4/4) – (E1 + E2 + E3 + E4/4), where P1 to P4, and E1 to E4, represent the four perception and expectation statements relating to a single dimension); and computation of the single measure of service quality ((P1 + P2 + P3 … + P22/22) – (E1 + E2 + E3 + … + E22/22)), the so-called SERVQUAL gap." (Buttle, 1995, p. 10)

In view of the emerging empirical examples questioning the legitimacy the measurement of the expected value, J. Cronin and S. A. Taylor, on the basis of the modified SERVQUAL, have proposed the SERVPERF method (Cronin & Taylor, 1992, pp. 55-67). The modification consists of taking into account only the assessment of quality of the service provided. The revised questionnaire contains half of the questions (there are no questions about the desired service), and, according to the results of empirical studies, it explains more variance of quality of service (Sanjay & Gupta, 2004, p. 28).

Articles using the SERVQUAL or SERVPERF methods are well suited to compare the effectiveness of the weighted average because researchers often publish the results both using the weighted average and the simple arithmetic mean. In addition, they include the opinion, that the measurement without weights is a better method for measuring the quality of service (Mazis, 1975, pp. 38-52, Cronin & Taylor, 1992, pp. 55-67, Teas, 1993, pp. 18–24, Blery& Gilbert, 2006 pp. 10–30). For example, D. P. Paul, in an article about the prosthetic dental research quality of services, came to the following conclusions: "SERVQUAL with the inclusion of importance weights was the most statistically significant model, but SERVPERF without importance weights accounted for the most variance. Prosthodontists may for the sake of brevity, decide to utilize SERVPERF withoutimportance weights to measure their perceived service quality. Considering thetrade-off between a marginal, albeit statistically significant, improvement instatistical significance, and the administration of a survey instrument one thirdthe length, SERVPERF without importance weights seems quite adequate forprosthetic dental practitioners' purposes" (Paul, 2003, p.89)."While the situation regarding the statisticalappropriateness of the decision as to whether or not to include importance weights inservice quality measurement is unclear, it may be that the inclusion of importanceweights add little, if anything, to practical perceived service quality modeling" (Paul, 2003, p.93). Another example are the results obtained by B. Corneliu on a sample of 250 respondents in the study of the quality of banking services, he stated that "we can notice that the weighted average score did not change in the situation of banking servicequality dimensions, resulting in the overall average positive score of 0.00751, the minimum andmaximum recorded size remains the responsiveness and tangibles."(Corneliu, 2012, p. 893) A. Anvary-Rostamy in an article about the perceived quality of banking services from the perspective of customers and employees of the bank in Iran, he has published a study using the SERVPERF and weighted SERVEPERF methods that differed in hundredths or thousandths decimal places (cf. Tab. 1). G. Santhiyavalli and B. Sandhya, in an article about the quality of services of selected banks in India, have obtained a very similar result for the result of the version with the weights and the version without weights - Unweighted SERVQUAL Score Bank_1=2.746 vs Total Weighted SERVQUAL Score Bank_1=2.731, Unweighted SERVQUAL Score Bank_2= 2.434 vs Total Weighted SERVQUAL Score Bank_2= 2.443.

The presented empirical examples raise the question whether, and if so, under what conditions the weighted average differs from the arithmetic mean?

Table 1: Results of SERVPERF andweighted SERVPERF Models

Number of Quality Dimensions	Bank Service Quality Average Score Using Simple SERVEPERF Model	Bank Service Quality Average Score Using Weighted SERVEPERF Model	
8	6.940913	6.94507	Employees
8	7.548330	7.45119	Customers

Source: Ali Asghar Anvary, 2009, p. 249.

3. Mathematical analysis of weighted average and the arithmetic mean formulas

Analyzing the equations of the arithmetic mean (a) and weighted arithmetic average (wa), one can easily observe that the two averages are equal in the two cases.

weighted arithmetic $(wa) = \frac{\sum_{i=1}^{i=n} W_i x_i}{\sum_{i=1}^{i=n} W_i}$

arithmetic mean (a) $= \frac{\sum_{i=1}^{i=n} x_i}{n}$

where: W_i - weight, x_i – feature, n – number of features

First, if the weights in the weighted average are the same, then the pattern after the simplifications takes the form of the arithmetic mean. Secondly, if the attributes' values are the same, then regardless of the weight values, the weighted average will give the same result as the arithmetic mean, and in this case it will be equal to the values of the attributes. The obvious proof can be found below.

Proof for equal weights ,$W_1=W_2=\ldots=W_n=a$

weighted arithmetic average $= \frac{\sum_{i=1}^{i=n} W_i x_i}{\sum_{i=1}^{i=n} W_i} = \frac{\sum_{i=1}^{i=n} a x_i}{\sum_{i=1}^{i=n} a} = \frac{a\sum_{i=1}^{i=n} x_i}{na} = \frac{\sum_{i=1}^{i=n} x_i}{n} =$ arithmetic mean

Proof for equal attributes' values $x_1=x_2=\ldots=x_n=a$

weighted arithmetic average $= \frac{\sum_{i=1}^{i=n} W_i x_i}{\sum_{i=1}^{i=n} W_i} = \frac{\sum_{i=1}^{i=n} W_i a}{\sum_{i=1}^{i=n} W_i} = \frac{a\sum_{i=1}^{i=n} W_i}{\sum_{i=1}^{i=n} W_i} = a = \frac{n x_i}{n} = \frac{\sum_{i=1}^{i=n} x_i}{n}$

= arithmetic mean

In case of equal weight values or equal values of attributes the matter is clear. However, the question arises whether, if not, all weights are not equal and the values of all attributes are not equal, then may one get little or no difference between the weighted average and the arithmetic mean. To try to answer that question the Monte Carlo method has been used. For randomly generated attributes and weights, the results obtained using the weighted average and the arithmetic mean have been compared. A total of 27 000 000 cases have been analyzed.

4. RESEARCH METHOD AND TEST RESULTS

The analysis of formulas of the weighted average and the arithmetic mean performed above allows for the hypothesis that the probability of a large difference between the analyzed averages is dependent on the distribution of attributes and the distribution of weight. For the purposes of the simulation the following assumptions have been adopted:

- both attributes and weights are integers,
- the interval of variability of attributes is within the range <0;100>,
- the interval of variability of weights is within the range <1;100>,
- difference in the averages below 0.05 on a scale of 0-100 has been assumed to be very small,
- difference in the averages above 1 on a scale of 0-100 • difference in average more than one on a scale of 0-100 was considered average,
- difference in average exceeding 5 on a scale of 0-100 has been assumed to be high.

In the case of the range of variation of weights it has been assumed that the minimum weight is 1, since the adoption of 0 would mean that the attribute is in general negligible. In the case of the attribute variation range one may allow for a selected attribute not to appear at all in any object, hence the minimum attribute value has been adopted to be 0.

It should be noted that, for the commonly used 5 and 7 degree scales, the difference of 1 on the average level of, on a 0-100 scale, is respectively 0.04 and 0.06.

Four distributions have been used to randomly choose both weights and attributes during the simulation: uniform distribution (UD), normal distribution with the following parameters: mean = 50, standard deviation = 33 - (ND) and asymmetric distributions (AD). In the case of asymmetric distributions, the number generation algorithm has been constructed in such a way that it can provide only the large (90-100) or small (1-10 for the weights and 0-10 for the attributes) values without intermediate values. One has iteratively increased by one the number of high values (90-100) at the expense of reducing by one the small values (1-10 for the weights and 0-10 for the attributes). Each distribution provided the randomly chosen 1000 numbers representing the n of the weights and 1000 numbers representing the n of attributes. Random selections have been carried out for n = 5, n = 10, n = 20 The next step in the simulation has been to calculate all the combinations of the weighted average (16 000 000 for each value of n). Subsequently, the difference between each weighted average and arithmetic mean. The results of the simulations are provided in tables 2-7 in the appendix to this article.

It may be noted that with the increase in the number of averaged attributes, the number of cases where there has been a big difference between the arithmetic mean and the weighted average (more than 5 on a scale of 0-100) significantly decreases. For n = 5 with 9 000 000 combinations 45.5% of such cases have been reported, for n = 10 37.57% of such corresponding cases have been reported and for n = 20 there have been 25.65%. At the same time for at least an average difference (more than 1 on a scale of 0-100) and for the very small differences (less than 0.05 on a scale of 0-100) there was no clear linear trend, which does not exclude a curvilinear dependence. Analyzing all combinations of the various features and weight distributions one may notice that for each of the study groups (n = 5, n = 10, n = 20), the most significant differences for 1 000 000 combinations have been reported for weights resulting from the asymmetric distribution and the attributes resulting from the asymmetric distribution. Most very small differences (less than 0.05 on a scale of 0-100) for n = 5 and n = 10 have also been observed in situations where weight and attributes were randomly selected from the asymmetric distribution - it has been respectively 4.42% and 2.22%. For n = 20 most very small differences have been observed for the attributes resulting from the normal distribution and the weights resulting from the asymmetric distribution. The least very small differences have been observed in the case where both the weights and the attributes have been derived from the uniform distribution, for n = 5 it has been 0.66%, for n = 10 it has been 0.82%, while for n = 20 it has been 1.1%.

5. CONCLUSIONS

The differences between the weighted average and the arithmetic mean are the higher, the higher the value of an important attribute gets, with low values of unimportant attributes. However, the more similar to each other the weights or values of the attributes are, the smaller difference between the arithmetic and the weighted mean. It has also been observed that with the increase in the number of attributes, the probability of the emergence of large differences between the arithmetic mean and the weighted average decreases.

The conclusions of the simulations are not limited to the SERVQUAL or SERVPERF methods, they can be generalized for all methods that use a weighted average. The analyzes do not provide evidence that the weighted average or arithmetic mean is better suited for multi-dimensional decisions, where each dimension may have different importance. They indicate only the conditions in which there is a high probability that the results will vary quite a lot, and when one may expect very similar results.

Appendix:

Tab.2. The number of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**5**

		Weights (1-100)			
		wa – a	UD	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	below 0.05	6606	24641	7603
		above 1	869221	746055	847624
		above 5	439456	525498	369258
	ASYMMETRIC DISTRIBUTIONS	below 0.05	15018	44153	17482
		above 1	777291	720126	753767
		above 5	519728	607189	476315
	NORMAL DISTRIBUTION	below 0.05	7937	29602	9157
		above 1	844807	731110	818877
		above 5	371093	484577	301500

Tab.3. The number of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**10**

		Weights (1-100)			
		wa – a	UNIFORM DISTRIBUTION	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	above 0.05	8161	18895	9265
		above 1	838263	799773	814532
		above 5	318974	453619	253155
	ASYMMETRIC DISTRIBUTIONS	below 0,05	12177	22152	13988
		above 1	807751	830295	793025
		above 5	422154	721977	375987

	NORMAL DISTRIBUTION	above 0.05	9650	21970	11025
		above 1	809519	783349	781442
		above 5	246222	404181	185277

Tab.4. The number of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**20**

			Weights (1-100)		
		wa – a	UNIFORM DISTRIBUTION	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	Below 0.05	11033	17669	12789
		above 1	780882	792933	747919
		above 5	171898	351828	115297
	ASYMMETRIC DISTRIBUTIONS	below 0.05	13324	12868	15207
		above 1	786853	891085	761301
		above 5	266145	728552	208728
	NORMAL DISTRIBUTION	below 0.05	13178	21229	15298
		above 1	742320	760484	703472
		above 5	109928	290060	66123

Tab.5. The percentage of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**5**

			Weights (1-100)		
		wa – a	UNIFORM DISTRIBUTION	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	below 0.05	0.66%	2.46%	0.76%
		above 1	86.92%	74.61%	84.76%
		above 5	43.95%	52.55%	36.93%
	ASYMMETRIC DISTRIBUTIONS	below 0.05	1.50%	4.42%	1.75%
		above 1	77.73%	72.01%	75.38%
		below 5	51.97%	60.72%	47.63%

	NORMAL DISTRIBUTION	below 0.05	0.79%	2.96%	0.92%
		above 1	84.48%	73.11%	81.89%
		above 5	37.11%	48.46%	30.15%

Tab.6. The percentage of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**10**

		Weights (1-100)			
		wa – a	UNIFORM DISTRIBUTION	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	poniżej 0.05	0.82%	1.89%	0.93%
		powyżej 1	83.83%	79.98%	81.45%
		powyżej 5	31.90%	45.36%	25.32%
	ASYMMETRIC DISTRIBUTIONS	poniżej 0.05	1.22%	2.22%	1.40%
		powyżej 1	80.78%	83.03%	79.30%
		powyżej 5	42.22%	72.20%	37.60%
	NORMAL DISTRIBUTION	poniżej 0.05	0.97%	2.20%	1.10%
		powyżej 1	80.95%	78.33%	78.14%
		powyżej 5	24.62%	40.42%	18.53%

Tab.7. The percentage of cases where the difference between the arithmetic mean and the weighted has been less than 0.05, and the number of cases where the differences have been between above 1 and above 5 for n=**20**

		Weights (1-100)			
		wa – a	UNIFORM DISTRIBUTION	ASYMMETRIC DISTRIBUTIONS	NORMAL DISTRIBUTION
Attributes (0-100) distribution	UNIFORM DISTRIBUTION	poniżej 0.05	1.10%	1.77%	1.28%
		powyżej 1	78.09%	79.29%	74.79%
		powyżej 5	17.19%	35.18%	11.53%
	ASYMMETRIC DISTRIBUTIONS	poniżej 0.05	1.33%	1.29%	1.52%

		powyżej 1	78.69%	89.11%	76.13%
		powyżej 5	26.61%	72.86%	20.87%
	NORMAL DISTRIBUTION	poniżej 0.05	1.32%	2.12%	1.53%
		powyżej 1	74.23%	76.05%	70.35%
		powyżej 5	10.99%	29.01%	6.61%

REFERENCE LIST

1. Ali Asghar Anvary R. (2009). Toward understanding conflicts between customers and employees' perceptions and expectations: evidence of Iranian bank. *Journal of Business Economics & Management*, *10*(3), pp. 241-254.
2. Blery E. D. & Gilbert, D. (2006). Factors influencing customer retention in mobile telephony: A Greek study. *Transformations in Business & Economics Journal*, *5*(2), 10–30.
3. Buttle F. (1995). SERVQUAL: review, critique, research agenda: European Journal of Marketing *30*(1), pp. 8-32.
4. Corneliu B. (2012) Concepts of service quality measurement in banks, Annals of the University of Oradea, *Economic Science Series*, *21*(2), pp. 889 - 894.
5. Corneliu B., Concepts of service quality measurement in banks, Annals of the University of Oradea, *Economic Science Series*, *21*(2), 889-894.
6. Cronin J., Taylor S. A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, *56*(3), pp. 55-67.
7. Mazis M. B., Ahtola O. T., Klippel, R. E. (1975). A comparison of four multi-attribute models in the prediction of consumer attitudes. *Journal of Consumer Research*, (2), pp. 38–52.
8. Parasuraman, A, Zeithaml V. A., Berry L. L. (1988). SERVQUAL: A Multi-item Scale Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, *64*(1), pp. 12-37.
9. Parasuraman, A., Zeithaml, V. and Berry, L.L. (1991). Refinement and reassessment of the SERVQUAL scale:*Journal of Retailing*, *67*(4), pp. 420-50.
10. Parasuraman, A., Zeithaml, V. and Berry, L.L. (1994). Reassessment of expectations as a comparison standard in measuring service quality: implications for future research:*Journal of Marketing*, *58*, January, pp. 111-24.
11. Paul, D. P. (2003). An Exploratory Examination of "SERVQUAL" Versus "SERVPERF" for Prosthetic Dental Specialists. *Clinical Research & Regulatory Affairs*, *20*(1), pp. 89-101.
12. Rosser M., Basic Mathematics for Economists, Routledge, 2003, p. 3.
13. Sanjay J. K., Gupta G. V. (2004). Measuring Service Quality: SERVQUAL vs. SERVPERF Scales. *The Journal for Decision Makers*, *29*(2), pp. 25-37.
14. Teas, K. R. (1993). Expectations, performance evaluation and consumers' perceptions of quality. *Journal of Marketing*, *57*(4), pp. 18–24.