ASM Sc. J., 13, 2020
https://doi.org/10.32802/asmscj.2020.sm26(2.15)

Malaysian Undergraduates' Perceptions of Learning Statistics: Study on Attitudes towards Statistics using Fuzzy Conjoint Analysis

Kathiresan Gopal[1*], Nur Raidah Salim[1] and Ahmad Fauzi Mohd Ayub[1,2]

[1]*Institute for Mathematical Research,*
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
[2]*Faculty of Educational Studies,*
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Statistics has become more important in today's world due to the emergence of big data. Accordingly, introductory statistics course plays the role of precursor to data science. However, learning of statistics has become difficult to Malaysian undergraduate students. In line with that, this paper evaluated the perceptions of learning statistics among students in a Malaysian public university based on their attitudes towards statistics. A survey was conducted on 293 students taking an introductory statistics course. The survey consisted of 28 attributes corresponding to 4 dimensions of attitudes. As perceptions of learning statistics are vague, fuzzy set approach was employed in this study by analysing the ratings using fuzzy conjoint analysis to evaluate students' perceptions. Results attested that students generally had negative perceptions of learning statistics. In particular, students were often anxious and frustrated during tests and solving problems (affect); were able to learn but had troubles in understanding (cognitive competence); acknowledged the importance of statistics but were still doubtful about the relevance (value); and admitted that statistics is difficult especially due to the technical aspects of statistics (difficulty).

Keywords: attitudes towards statistics, fuzzy conjoint analysis, perceptions, statistics learning

I. INTRODUCTION

Statistical and data analysis skills are vital to make data-driven decisions. Statistics has undeniably become salient in this big data era (Ridgway, 2016). Accordingly, in Malaysia, undergraduate students are equipped with statistical and data analysis skills through introductory statistics course. This course plays the role of precursor to data science (Horton *et. al.*, 2014). However, students coming from a less mathematics exposed background, often find it difficult to learn statistics (Kien *et. al.*, 2016). As such, most students failed to grasp statistical and data analysis skills effectively. Several factors influence the learning of statistics, such as students' attitudes towards statistics (Carver *et al.*, 2016; Schau *et. al.*, 2012).

A. Attitudes Towards Statistics

Chiesiand Primi (2018) asserted that bond attitudes towards statistics affect students' learning approaches and eventually impacts their achievement. Ultimately, attitudes have an effect on how students move towards statistics as a discipline (Chiesi & Primi, 2018).

Attitude towards statistics is students' dispositions to respond favourably or unfavourably to the attributes of statistics learning (Garcıa-Santillan, *et. al.*, 2013). It is a non-cognitive multidimensional construct encompassing four dimensions viz. **affect** (positive and negative feelings about statistics); **cognitive competence** (attitude about their intellectual knowledge and skills when applied to statistics); **value** (attitude regarding the usefulness, relevance, and worth of statistics in personal and professional life) and **difficulty** (attitude about the

*Corresponding author's e-mail: kathiresan@upm.edu.my

difficulty of statistics as a discipline) (Schau, *et. al.*, 2012; Bond *et. al.*, 2012; Emmioglu & Capa-Aydin, 2012). Positive *affect* reflects students' interest; positive *cognitive competence* reflects students' belief and confidence in understanding and ability; positive *value* reflects students' appreciation and relevance; and positive *difficulty* reflects students' understanding and ease to learn (Emmioglu & Capa-Aydin, 2012).Students' attitudes towards statistics resemble their perceptions of statistics learning.

B. Fuzzy Sets Approach

Words or sentences in human natural languages used to describe preferences are known as linguistic variables, L (non-numeric valued) (Turksen & Willson, 1994). The uncertainties and unclear boundaries inherited in linguistic variables such as *disagree* for a statement or *excellent* for a service are handled with ease using fuzzy sets approach (Zimmermann, 2001).

Since perceptions of learning statistics are vague as well, therefore applying fuzzy set approach is appropriate in this study. For example, if a student rated 4 (*agree*) on an attribute that states '*statistics is difficult*', it does not really reflect the degree of the student's agreement on statistics being difficult, to conclude. Therefore, to deal with the fuzziness, strength of each ratings (agreement) are evaluated using a numerical value i.e. degrees of similarity (Sarala & Kavitha, 2017).

In view of the above discussion, the objective of this study is to evaluate the perceptions of learning statistics among Malaysian undergraduate students.

II. METHODOLOGY

A. Survey

A survey was conducted on 293 randomly selected students from non-mathematics programme taking an introductory statistics course in a Malaysian public university. This survey took place at end of the course. Survey questionnaire was adapted from the well-established and validated *Survey of Attitudes Toward Statistics-28* (SATS-28) instrument with 28 attributes developed by Schau *et. al.,* (1995). SATS was validated using confirmatory factor analysis and its' concurrent validity was verified based on the significant correlations with Wise's *Attitudes Toward Statistics* (ATS) scales (Schau *et. al.,* 1995). Item analysis on the instrument resulted with reliability indices of 0.85 for affect (6 attributes), 0.83 for cognitive competence (6 attributes), 0.85 for value (9 attributes) and 0.77 for difficulty (7 attributes). Rating on these attributes were obtained using standard 5-point Likert scale corresponding to the linguistics variables of agreement i.e. *strongly disagree*, *disagree*, *neutral*, *agree* and *strongly agree*.

B. Fuzzy Conjoint Analysis

Turksen and Wilson (1994) developed the fuzzy conjoint model (FCM) (Equation 1). Fuzzy sets arising from FCM are linear combinations of the attributes' weights (Sarala & Kavitha, 2017). The standard fuzzy sets F defined for ratings on attributes are the input to FCM (Sofian & Rambely, 2018). The approximate degree of membership for each domain element (*linguistic label*), $y_j$in the calculated overall preference fuzzy set R, $\mu_R(y_j,\, A_m)$for a particular attribute A_m is (Turksen & Willson, 1994):

$$\mu_R(y_j,\, A_m) = \sum_{i=1}^{j} W_{(r_i,\, A_m)} \cdot \mu_{F_i}(x_j) \qquad (1)$$

where:

- y_j and x_j are domain elements, j is the number of linguistics variables, j = 1, 2, ..., 5
- A_m is a particular attribute, m is the number of attributes, m = 1, 2, ..., c where c = 6 for affect and cognitive competence, c = 9 for value and c = 7 for difficulty
- $\mu_{F_i}(x_j)$ is the membership value of the linguistic rating, F_i at given linguistic level x_j (elements of the standard fuzzy set F at level x_j)
- $W_{(r_i,\, A_m)}$ is the fuzzified weight for linguistic rating r_i corresponding to attribute A_m
- $W_{(r_i,\, A_m)} = \frac{\sum r_i}{\sum_{k=1}^{j} r_{(k,A_m)}}$ is the sum of the particular rating$r$ throughout the respondents for attribute $A_m$and $\sum_{k=1}^{j} r_{(k,A_m)}$ is the sum of all the ratings throughout the attribute A_m
- $\mu_R \in [0,1]$

The membership values for each linguistic variable (or term), $\mu_{F_i}(x_j)$ are pre-defined values obtained from Zimmermann (2001), which is based on Zadeh's work on concept of linguistic variables and its application to approximate reasoning. The fuzzy sets, F representing the linguistics variables (*degree of agreement*) for L are as below:

F_1=(0.50/1, 1.00/2, 0.75/3, 0.25/4, 0.00/5)

F_2 = (0.50/1, 1.00/2, 0.75/3, 0.25/4, 0.00/5)

F_3 = (0.00/1, 0.50/2, 1.00/3, 0.50/4, 0.00/5)

F_4=(0.00/1, 0.25/2, 0.75/3, 1.00/4, 0.50/5)

F_5 = (0.00/1, 0.00/2, 0.50/3, 0.75/4, 1.00/5)

F is anchored to $L = (L_1, L_2, L_3, L_4, L_5)$. The notation a/b used in F is defined as a at level b. For instance, in F_1 that corresponds to L_1, the first element $(0.50/1)$ means the compatibility of rating '1' with L_1(*strongly disagree*) is 0.50.

The final output of FCA is a fuzzy similarity measure, which is the sum of the Euclidean distance between corresponding elements in *R* and *F*. The *s* of corresponding elements in *R* and *F* for attribute A_m is (Turksen & Willson, 1994):

$$s_j(R, F) = \frac{1}{1 + \sqrt{\sum_{i=1}^{j}[\mu_R(y_j) - \mu_F(y_j, L_i)]^2}} \qquad (2)$$

where:

- $\mu_F(y_j, L_i)$ is the elements of F corresponding to linguistic term L_i (actual overall evaluation or response) and j is the number of linguistic variables, j = 1, 2, ..., 5
- $\mu_R(y_j)$ is the calculated membership degree using original ratings in surveyquestionnaire for the attributes from Equation 1.$s \in [0,1]$

Significance of attributes is determined using the ordinal information (rank) given by *s* (Sofian & Rambely, 2018). The maximum *s* of each attribute, denoted by *s** is used to obtain the linguistic term that reflects a perception's nature (Turksen & Willson, 1994). Nature of perceptions, i.e. **positive** or **neutral** or **negative** for an attribute is decided based on the attribute's sentiment and L that corresponds to *s** denoted as *L*(*s**).

Attributes are ranked in descending order of *s**. Attribute with highest *s** (rank = 1) is the most significant (important) attribute that influenced students' perceptions and otherwise. Significance of attributes is directly proportional to the magnitude of *s**. As *s**→ 1, significance of attributes is higher while if *s**→ 0, significance is lower. Example of similarity degree computation for attribute *Aff*1 is outlined as below:

1. *Aff*1's rating is collected:
number of rating '1' = 19; number of rating '2' = 25
number of rating '3' = 101; number of rating '4' = 130
number of rating '5' = 20

2. r_i is computed:
$r_1 = 19(1) = 19$; $r_2 = 25(2) = 50$
$r_3 = 101(3) = 303$; $r_4 = 130(4) = 520$
$r_5 = 18(5) = 90$

3. $\sum_{k=1}^{5} r_{(k,Aff1)} = 19 + 50 + 303 + 520 + 90 = 982$

4. $W_{(r_1,Aff1)}$is computed:

$$W_{(r_1,Aff1)}L_1 = 19 \div 982 = 0.019348$$
$$W_{(r_1,Aff1)}L_2 = 50 \div 982 = 0.050916$$
$$W_{(r_1,Aff1)}L_3 = 101 \div 982 = 0.308554$$
$$W_{(r_1,Aff1)}L_4 = 520 \div 982 = 0.529532$$
$$W_{(r_1,Aff1)}L_5 = 90 \div 982 = 0.091648$$

5. $\mu_R(y_j = 1, Aff1)$ is computed by multiplying with $\mu_{F_1}(x_j = 1)$:

$\mu_R(y_j = 1, Aff1) = 0.019348(1) + 0.050916(0.5) + 0.308554(0) + 0.529532(0) + 0.091648(0) = 0.044807.$

Similarly, for remaining four values of y_j and x_j, which resulted as:

$$\mu_R(y_j = 2, Aff1) = 0.352088$$
$$\mu_R(y_j = 3, Aff1) = 0.799389$$
$$\mu_R(y_j = 4, Aff1) = 0.765275$$
$$\mu_R(y_j = 5, Aff1) = 0.356415$$

6. $s_1(R, F)$ is computed using $\mu_R(y_{j=1})$ and $\mu_F(y_j = 1, L_1)$:

$$1 \div (1 + ([0.044807 - 1]^2 + [0.352088 - 0.75]^2 + [0.799389 - 0.5]^2 + [0.765275 - 0]^2 + [0.356415 - 0]^2 = 0.422192$$

$\therefore s_1(Aff1) = 0.422192$, similarly, others:

$s_2(Aff1) = 0.497284$ $s_3(Aff1) = 0.661637$

$s_4(Aff1) = 0.76866$ $s_5(Aff1) = 0.557491$

$\therefore s^*(Aff1) = s_4 = 0.76866$ and $L(s^*) = L_4$

III. RESULTS AND DISCUSSIONS

A. *Affect*

The most significant attribute of affect was *Aff2*, as shown in Table I revealed that students felt insecure when they have to do their statistics problems. Attribute*Aff3* depicted that students were frustrated going over statistics tests in class. *Aff1* disclosed that students liked statistics. *Aff6* indicated that students were scared by statistics. *Aff4* and *Aff5* were rated as *neutral* (L_3), which showed that students were unsure if they were under stress during statistics classes; and if they enjoyed taking the statistics course.

Students' perceptions with respect to affect were mainly negative. These perceptions were due to anxiety and frustration when doing statistics problems and during tests. Apparently, there is no positive perceptions found, while neutral perceptions were found for attributes relating to their learning environment.

Table I. Similarity degree between fuzzy sets R and F for affect (*Aff*) attributes

Attribute	L_1	L_2	L_3	L_4	L_5	s^*	$L(s^*)$	Rank
*Aff*1	0.422192	0.497284	0.661637	0.76866	0.557491	0.76866	L_4	3
*Aff*2	0.41892	0.49101	0.644543	0.789017	0.568707	0.789017	L_4	1
*Aff*3	0.425518	0.49113	0.618613	0.771599	0.597992	0.771599	L_4	2
*Aff*4	0.47807	0.584721	0.719994	0.627671	0.504986	0.719994	L_3	5
*Aff*5	0.434934	0.522899	0.717647	0.705185	0.525292	0.717647	L_3	6
*Aff*6	0.431408	0.506495	0.659803	0.741551	0.566885	0.741551	L_4	4

B. Cognitive Competence

As seen in Table II, the most significant cognitive competence's attribute was *Cog4*, in which students agreed that they can learn statistics. Attribute *Cog5*revealed that students were able to understand statistics equations. *Cog3*indicated that students made a lot of math errors in statistics. *Cog1* revealed that students had trouble understanding statistics due to their way of thinking. Next, *Cog2*showed that students were unsure if they had idea of what was going on in statistics. *Cog6* indicated that it was difficult for students to understand statistics concepts. More than half of the students agreed that they can learn statistics, contributing to positive perceptions. Additionally, they were able to understand statistics equations.

Nonetheless, students had negative perceptions when it comes to the mathematics calculations in statistics, understanding statistics in the usual way and understanding statistics concepts. Besides that, students were uncertain if they were clueless about statistics, contributing to neutral perception.

Table II. Similarity degree between fuzzy sets R and F for Cognitive Competence (*Cog*) attributes

Attribute	L_1	L_2	L_3	L_4	L_5	s^*	$L(s^*)$	Rank
*Cog*1	0.421765	0.494575	0.650609	0.775128	0.568281	0.775128	L_4	4
*Cog*2	0.464187	0.571703	0.75546	0.635872	0.50001	0.75546	L_3	5
*Cog*3	0.416675	0.483817	0.628148	0.794774	0.588995	0.794774	L_4	3
*Cog*4	0.405403	0.467888	0.603979	0.852906	0.598655	0.852906	L_4	1
*Cog*5	0.413714	0.484976	0.6448	0.799032	0.565486	0.799032	L_4	2
*Cog*6	0.428044	0.503231	0.659442	0.753841	0.564296	0.753841	L_4	6

C. Value

Table III shows that the most significant attribute of value was *Val2*, with students agreeing that statistics should be a required part of professional training. The second most significant attribute was *Val3*, for which students agreed that statistical skills will make them more employable. As for the third most significant attribute *Val9*, students were unsure about the relevance of statistics. The next attribute *Val6*indicated that students were uncertain if they used statistics in daily life. Attribute *Val4* disclosed that students were not sure if statistics is useful for typical professional. Next, attribute *Val1*showed that students were iffy about the worth of statistics. The following attribute, *Val5* indicated that students were unsure about the application of statistical thinking outside of their classroom. Attribute *Val8*disclosed that students felt statistics will not be applied in their future profession. Finally, attribute *Val7* revealed students' agreement on the insignificance of statistics conclusions in daily life.

Overall, positive perceptions were found for value as students acknowledged the *value* or importance of statistics. However, the neutral perceptions found suggest that a large number of students were still doubtful regarding the relevance, worth and application of statistics outside this course. As for the negative perceptions, students failed to see the application of statistics in daily life.

Table III. Similarity degree between fuzzy sets R and F for Value (*Val*) attributes

Attribute	L_1	L_2	L_3	L_4	L_5	s^*	$L(s^*)$	Rank
*Val*1	0.498128	0.623856	0.741314	0.584324	0.476821	0.741314	L_3	6
*Val*2	0.404737	0.464994	0.598493	0.842715	0.610935	0.842715	L_4	1
*Val*3	0.416442	0.487219	0.647257	0.785931	0.570867	0.785931	L_4	2
*Val*4	0.496124	0.628342	0.757224	0.57802	0.469506	0.757224	L_3	5
*Val*5	0.465045	0.568456	0.73955	0.643772	0.505999	0.73955	L_3	7
*Val*6	0.47438	0.59003	0.76509	0.615071	0.486975	0.76509	L_3	4
*Val*7	0.435052	0.51981	0.702435	0.713989	0.536645	0.713989	L_4	9
*Val*8	0.438943	0.521437	0.687326	0.715534	0.545893	0.715534	L_4	8
*Val*9	0.470697	0.584526	0.77082	0.618099	0.48944	0.77082	L_3	3

D. Difficulty

Based on Table IV, the most significant difficulty's attribute was *Dif4* for which students agreed that learning statistics requires a great deal of discipline. *Dif5*disclosed that students felt statistics involves massive computations. *Dif7*revealed students' agreement on the need to learn a new way of thinking to do statistics. *Dif6*showed that students think statistics is highly technical. *Dif2*indicated that statistics is a complicated subject for these students.

Next, *Dif3*revealed that students were unsure if statistics is quickly learn-able. *Dif1* indicated that students were uncertain if statistics formulas were easy to understand.

Clearly, students had more negative perceptions for difficulty and none of the attributes were perceived positively. Students agreed at a high rate that statistics is difficult. Most students faced complications with the technicality present in statistics, possibly due to their poor background in mathematics. Neutral perceptions were present for attributes related to ease of learning statistics.

Table IV. Similarity degree between fuzzy sets R and F for Difficulty (*Dif*) attributes

Attribute	L_1	L_2	L_3	L_4	L_5	s^*	$L(s^*)$	Rank
*Dif*1	0.444215	0.536388	0.722265	0.688526	0.521454	0.722265	L_3	7
*Dif*2	0.426712	0.49485	0.629211	0.767135	0.588917	0.767135	L_4	5
*Dif*3	0.453398	0.554401	0.748042	0.659042	0.505335	0.748042	L_3	6
*Dif*4	0.398692	0.451736	0.567405	0.837065	0.649282	0.837065	L_4	1
*Dif*5	0.40639	0.465182	0.590761	0.832559	0.62222	0.832559	L_4	2
*Dif*6	0.416328	0.486047	0.643713	0.784579	0.576297	0.784579	L_4	4
*Dif*7	0.410728	0.475686	0.620364	0.815069	0.593109	0.815069	L_4	3

IV. CONCLUSIONS

Findings revealed that students' perceptions of learning statistics are generally negative. The dominance of negative perceptions indicates that the nation will see a shortage of qualified data-related expertise in the future. This would impact the nation's progress in many ways.

Although the issue in statistics learning is almost similar in every Malaysian public university, however, the findings of this study are not generalizable since the sample size is not considerably high and not representative enough for the whole country. This study can be extended by taking equal sized samples from all the public universities in the country.

Application of FCA provided an overview of students' perceptions of learning statistics based on their attitudes towards statistics. Additionally, attributes that had the most significant effects on students' rating tendency were also identified. Identification of negatively and neutrally perceived attributes are useful to education stakeholders to understand the difficulties faced by students in learning statistics. This would help to improve the overall teaching-learning process of introductory statistics, particularly in Malaysian public universities. This approach would be more efficient, as it narrows down to specific aspects of statistics learning within the dimension of attitudes rather than looking at this issue as a whole.

V. REFERENCES

Bond, M. E., Perkins, S. N., and Ramirez, C. 2012, Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. Statistics Education Research Journal, 11(1), 6–25.

Carver, R., Everson, M., Gabrosek, J., Horton, N.J., Lock, R.H., Mocko, M., Rossman, A., Rowell, G.H., Velleman, P., Witmer, J.A., and Wood, B. 2016, Guidelines for assessment and instruction in statistics education (GAISE). Washington, DC: American Statistical Association. Retrieved from http://www.amstat.org/education/gaise.

Chiesi, F. and Primi C. 2018, What happens when attitudes toward statistics change (increase vs decrease) during the course?. In M. A. Sorto, A. White, and L. Guyot (Eds.), Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Kyoto, Japan: International Statistical Institute.

Garcia-Santillan, A., Venegas-Martinez, F., and Escalera, M. E. 2013, Attitude toward Statistic in College Students (An Empirical Study in Public University). Journal of Statistical and Econometric Methods, 2(1), 43-60.

Emmioglu, E., and Capa-Aydin, Y. 2012, Attitudes and achievement in statistics: A meta-analysis study. Statistics Education Research Journal, 11(2), 95–102.

Horton, N. J., Baumer, B. S., and Wickham, H. 2014, Teaching precursors to data science in introductory and second courses in statistics. In K. Makar, B. de Sousa, and R. Gould (Eds.), Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA. Voorburg: International Association of Statistics Education.

Kien-Kheng, F., Azlan, N., Ahmad, S. N. D., Leong, N. L. H., and Mohamed, I. 2016, Relationship between Cognitive Factors and Performance in an Introductory Statistics Course: a Malaysian Case Study. Malaysian Journal of Mathematical Sciences, 10(3), 269-282.

Ridgway, J. 2016, Implications of the data revolution for statistics education. International Statistical Review, 84(3), 528-549.

Sarala, N. and Kavitha, R. 2017, Fuzzy conjoint model in measuring students' expectation and teachers' beliefs on learning mathematics. International Journal of Advanced Trends in Engineering, Science and Technology, 2(2), 6–10.

Schau, C., Millar, M., and Petocz, P. 2012, Research on attitudes towards statistics. Statistics Education Research Journal, 11(2), 2-5.

Schau, C., Stevens, J., Dauphinee, T. L., and Vecchio, A. D. 1995, The development and validation of the survey of attitudes toward statistics. Educational and psychological measurement, 55(5), 868-875.

Sofian, S. S. and Rambely, A. S. 2018, The effectiveness of game and recreational activity to motivate high achievers and low achievers: Evaluation using fuzzy conjoint analysis. In AIP Conference Proceedings, volume 1940, page 020128. AIP Publishing.

Turksen, I. B. and Willson, I. A. 1994, A fuzzy set preference model for consumer choice. Fuzzy Sets and Systems, 68(3), 253–266.

Zimmermann, H.-J. 2001, Fuzzy set theory–and its application. Springer Science+Business Media, New York, NY, 4th edition.