Comparison of the Level of Familiarity of IT Units' Staff and students with Big Data Analyzes

Elham Nazari
Mashhad University of Medical Sciences
Parnian Asgari
Mashhad University of Medical Sciences
hamed tabesh (tah2584@gmail.com)
Mashhad University of Medical Sciences

Research

Keywords: Big Data, Benefits, Challenges, Analyzes

Posted Date: February 4th, 2020

DOI: https://doi.org/10.21203/rs.2.22539/v1

Abstract

introduction The rapid development of technology in recent decades has led to the production of a huge amount of data. This type of data analysis that is called Big Data Analysis obtain Many benefits, including reducing costs. One of the challenges of these analyses is the lack of specialized expertise and knowledge in this area. The purpose of this study was to compare the familiarity of IT staff and students with big data analyzes at various universities and organizations.

Materials and method This analytical study was conducted on IT units' staff and students of different organizations and universities in Mashhad, Iran. A questionnaire was designed based on reviewing the texts published in PubMed, google scholar, science direct, and EMBASE databases and using the Delphi method and the attendance of 10 specialists in different disciplines. The designed questionnaire evaluated the participants' knowledge about the Big Data analyzes in two parts. The participants were 265 IT units' staff and students of different organizations, completing the designed questionnaire. Participants' opinion was evaluated using two descriptive and analytical approaches. The relationship between knowledge scores and individual characteristics such as gender, age, work experience, Field of study, degree, the average number of hours' scientific study and non-scientific study per week was examined. To investigate the synchronous and reciprocal effects GLM was used.

Results Scores earned by students and staff were 2.66 ± 1.13 and 2.28 ± 1.21 respectively that p =. 012 represented a significant correlation between the level of knowledge of students and staff. In other words, the level of knowledge of staff about big data was more than the level of knowledge of the students.The correlation of each of the variables was not significant with the score of the Big Data Analysis Knowledge.But There was a significant correlation between experience and gender with the knowledge scores.

Conclusions In general, the level of knowledge in analyzing big data in different groups of people was at a low level that implementing measures such as holding training courses in this field seems necessary.

Introduction

Today with the emergence of various technologies such as Smart Phone, Internet Of Things (IoT) and the rapid development of Internet ,data are produced in all industries. These data are called Big Data and it have certain characteristics such as diversity, high volume.

This type of data could not be managed alone and analyzed due to having such characteristics as high volume, diversity that it's analysis using classic methods impossible and require the use of related and appropriate methods. An analysis of Big Data is known for making the right decisions as to the Big Data analysis [1, 2].

Today, Big Data Analysis has become a hot topic in all industries and academic environments because these analyses have many benefits such as reducing cost, discovering useful data patterns, facilitating data interpretation, extracting important features, and summarizing and sharing data for critical and vital decisions making at the same time, can these benefits and capabilities of these analyzes can be found in various industries such as Health care, military, agriculture, banking, etc., and used them in important research fields [3–6].But these analyses have challenges that have created barriers. Some of these challenges were the lack of quality and sufficient data, the lack of equipment and Infrastructures necessary for analysis, lack of familiarity with the techniques needed and lack of expertise.

The most important challenge lack of sufficient knowledge and expertise in this field that will affect the benefits [7–9]. There are these challenges in Iran too and it is necessary to resolve these challenges by applying measures such as informing, holding training courses, conferences, etc.

Therefore, the importance of Big Data analysis and the need to pay attention to the purpose of this study was to compare the level of familiarity of IT Units ' Staff and students with Big Data Analyzes in Mashhad.The familiarity and awareness of students who were in the research stage were examined.

And staff who have background and experience in similar environments and software applications to evaluate them.

Material And Methods

This cross-sectional study was conducted on IT units' staff and students of different organizations and universities in Mashhad, Iran. Mashhad is the largest city in the eastern of Iran with about 3 million people, located on the border with Afghanistan and Turkmenistan and on the way of Silk Road with more than 70 public organizations and private companies. Given that today, information technology (IT) is vital for organizations and necessary for the organization's improvement, in each organization, at least one IT expert provides services. In organizations, important tasks are assigned to IT staff, including managing existing networks, software, and hardware; maintaining existing software; developing and upgrading software; monitoring databases, etc. People who have this job are usually educated in software, hardware, network, and information technology. There are two major state universities in Mashhad, Ferdowsi University and the University of Medical Sciences. The former host's students from different fields of study including engineering and basic sciences. The latter host's students from medical fields of study such as medicine and biology. To evaluate students' knowledge and awareness of Big Data analysis in different fields of study in Mashhad universities, a questionnaire was developed. To assess the level of knowledge and awareness of IT staff of different organizations in Mashhad with Big Data analyzes, a questionnaire was designed. The questionnaire consisted of closed-ended items. The initial items of the questionnaire were prepared based on the reviewing the texts published in PubMed, google scholar, science direct, and EMBASE databases and then designed according to the Delphi method with the attendance of 10 specialists in different disciplines (medical informatics, biostatistics, and computer). This questionnaire contains five items concerned with one's knowledge of how to analyze Big Data. The relevant items can be seen in Table 1:

Table 1
Questionnaire items

Questions	Description
Knowledge Questions	
QK1	What is the definition of Big Data?
QK2	What are the hardware requirements for analysis?
QK3	What is the focus of Big Data analyzes?
QK4	What are the advantages of Big Data analyzes?
QK5	What are the disadvantages of Big Data analyzes?

The reliability and validity of the questionnaire were confirmed as a panel of 10 experts confirmed the validity and Cronbach's alpha was estimated to test reliability and was estimated at 81% and 73% for staff and students. Then, the required data were collected and it was made sure that all questionnaires were completed. Then the questionnaires were provided to 30 public and private organizations and present research attempted to include students of different fields of study. These included the following within Medicine, Computer Engineering, Pharmacy, Basic Sciences etc. from two major universities., The inclusion criterion for the selection of organizations was as follows: having independent IT units within their organizational chart and having staff with experience in working with different software; for example, the social security organization, hospitals, transportation organization, governorate, and the like were included in this study.These organizations provide services in the field of health care, transfer management, supervision of other organizations, and so on. Data were collected from these organizations, and it was ensured that the participants completed all the questionnaire items Out of 150 questionnaires sent by post to the IT staff working in these organizations, 123 questionnaires were completed. From among the initial 150 distributed questionnaires, 142 were completed and returned. T-test and ANOVA and GLM were used for the selected variables. Data entry and analysis were done using SPSS21 and Excell-2007.

Results

Individual characteristics of the participants can be seen in Table 2:

Table 2. Individual characteristics of the participants

Variables	Items	Frequency (percentage) of student)n=142)	Frequency (percentage) of staff (n=123)
Age	18-24year	7(4.9%)	0(0%)
	25-34 year	62(43.7%)	19(15.4%)
	35-44 year	63(44.4%)	61(49.6%)
	45-54 year	10(7.0%)	36(29.3%)
	55-64 year	0(.0%)	7(5.7%)
Sex	Male	58(40.8%)	80(66.7%)
	Female	84(59.2%)	40(33.3%)
	Missing	0(0%)	3(0.02%)
Experience_history	<=1 year	113(79.6%)	28(22.8%)
	>1 year	29(.20%)	95(0.67%)
Degree	BA	29(20.4%)	57(48.3%)
	MA	39(27.5%)	61(51.7%)
	Professional doctorate	45(31.7%)	0(.0%)
	PhD	29(20.4%)	0(.0%)
	Missing	0(0%)	5(0.4%)
Score	0	6(4.2%)	0(0%)
	20	37(26.1%)	18(14.6%)
	40	39(27.5%)	43(35.0%)
	60	34(23.9%)	32(26.0%)
	80	22(15.5%)	23(18.7%)
	100	4(2.8%)	7(5.7%)

As shown the amount of knowledge in the student group at age 25 – 44 year and the age range 35-44 year in the group of staff have had the highest score.

Also, the level of knowledge in the student group of female and male in the group of staff was higher. The work experience in the group of staff was also higher. Most of the staff have masters and most Ph.D. students. In both groups, they won the most points at 40 points.

Table 3: Mean and standard deviation of participants' hours of scientific and non-scientific studies across fields of study in two groups of students and staff.

Non-scientific-hours studying	Scientific-hours studying	Participant
Mean ± SD	Mean ± SD	
3.13±0.92	3.79±0.59	Student(n=142)
2.70±0.82	3.13±0.89	Staff(n=123)
0.001	<0.001	p-value

As can be seen, there was a significant correlation between the mean hours of scientific study and the mean hours of the non - scientific study of students and staff.

Investigating the relationship of knowledge score with each variable was studied individually (simple analysis) and once analyzed in the GLM modeling (multiple analyses). These results are as follows:

Table 4. Comparison of the mean scores of knowledge of staff and students in terms of age group, gender, background, degree, field, number of hours of scientific and non-scientific study

scores of knowledge of participant

Variables	Items	Staff (n=123)			Students (n=143)			
		n_i	Mean	SD	n_i	Mean	SD	p-value[1]
Age	18-24 year	0			7	2.2857	1.38013	-
	25-34 year	19	2.4737	1.07333	62	2.0806	1.20516	0.206
	35-44 year	61	2.5902	1.16013	63	2.4286	1.20100	0.448
	45-54 year	36	2.8889	1.08963	10	2.7000	1.15950	0.635
	55-64 year		2.5714	.97590				–
p-value		.511			.284			
Sex								p-value[3]
	Male	80	2.6000	1.06260	58	2.2069	1.16617	0.041
	Female	40	2.7750	1.25038	84	2.3452	1.24662	0.075
p-value[4]		.425			.506			
experience_history								Pvalue[5]
	<=1year	28	2.4643	.99934	113	2.1947	1.17912	0.267
	>1 year	95	2.716	1.145	29	2.667	1.278	0.809
p-value[6]		.172			.140			
degree								Pvalue[7]
	BA	57	2.5965	1.09967	29	2.1034	1.26335	0.065
	MA	61	2.7213	1.15659	39	2.0256	1.08790	0.003
	Professional doctorate	0			45	2.6000	1.21356	
	PhD	0			29	2.3448	1.26140	
p-value[8]		550			.135			
scientific_hours_studying								Pvalue[9]
	<1	18			11			0.976
	1-3	35	2.7143	.04520	12	2.3333	1.30268	0.312
	3-5	20	2.7000	1.41793	3	2.3333	1.52753	0.683
	>5	49	2.6735	1.14360	46	2.2586	1.22383	0.044
Pvalue[10]		.987			.976			
Non-scientific_hours_studying								Pvalue[11]
	<1	44	2.568	1.246	33	2.879	1.166	0.270
	1-3	46	2.6957	1.19014	35	2.1714	1.07062	0.044
	3-5	22	2.7273	1.03196	15	2.0000	1.25357	0.062

	>5	20	2.1500	.74516	48	2.2083	1.25407	0.813
Pvalue[12]		0.128			0.839			

Note that:

1. To compare the mean scores of knowledge in the analysis of student data and staff data in each age group, the independent sample was used.
2. One-way ANOVA test was used to compare the mean scores of knowledge of big data analysis of each group in different age groups.
3. To compare the mean score of big data analysis of students and staff in any gender of the independent test sample was used.
4. To compare the mean score of big data analysis of each of the groups in two different genders of the independent test sample was used.
5. To compare the mean score in big data analysis of students and staff in each category of work experience of the independent test sample was used.
6. The independent tests were used to compare the mean scores of knowledge in the analysis of student data and staff data at each degree.
7. To compare the mean score in big data analysis of each of the groups in different sections. Of one-way ANOVA was used.
8. To compare the mean score in big data analysis of students and staff in each row from the scientific study hours of the independent test sample was used.
9. To compare the mean score of big data analysis of each of the groups in the scientific study hours of one-way ANOVA was used
10. To compare the mean score in big data analysis of students and staff in each row of non-scientific study of the independent test sample was used.
11. To compare the average score of big data analysis of each of the groups in non-scientific study groups hours of one-way ANOVA was used.

As seen in Table 4 a significant difference between the average score of staff and students was not seen in any of the levels of age group. Also, there was no significant difference between the mean scores of staff's knowledge at different levels of age group. Also for students although the effect of the age group on the comparison of the average score of knowledge of both groups of staff and students was adjusted with stratification. However, there was no significant difference between the mean scores of knowledge in these two groups.The higher average level of knowledge of the male staff is significant male students.The average of the knowledge score in staff of the master's degree was also observed compared to the master's students. Significant differences in staff knowledge score with more than 5 hours of scientific study in comparison with the knowledge of the students with more than 5 hours of scientific study are remarkable.

General Linear Model results

Table 5. Investigating the relationship between the variables group (staff, students), gender, age, work experience (less than one year and more than one year) degree, the number of scientific and non-scientific study hours daily with a score of big data analysis.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.*
	86.461[a]	51	1.695	1.280	.143
	24.911	1	24.911	18.809	.000
scientific_hours_studying	.520	1	.520	.393	.532
non_scientific_hours_studying	.015	1	.015	.012	.915
group	3.635	1	3.635	2.745	.100
age	.407	2	.203	.153	.858
sex	1.095	1	1.095	.827	.365
experience_history	3.600	3	1.200	.906	.441
degree	7.598	3	2.533	1.912	.132
group * experience_history	.072	1	.072	.054	.817
group * degree	.892	1	.892	.673	.414
age * sex	.338	2	.169	.128	.880
age * experience_history	5.093	4	1.273	.961	.432
age * degree	11.008	5	2.202	1.662	.150
sex * experience_history	8.579	2	4.289	3.239	.043
sex * degree	9.245	3	3.082	2.327	.079
experience_history * degree	2.248	4	.562	.424	.791
group * age * sex	.000	0	.	.	.
group * age * experience_history	.000	0	.	.	.

To investigate the effect of different variables such as age and gender, etc. on the level of knowledge of individuals about the big data the General Linear Model was used.

Table 5 shows that none of the variables studied individually did not have any effect on the points obtained in the field of knowledge, however, the interaction between gender and work experience is an influential factor.

That no significant for related main effects could not be presented interpretation for them.

Assessment of the rating according to the age group in the two groups is shown in the diagrams below.

Figure 1. Evaluated scores by age group in two groups.

According to the figure1 average score in the age group of 35 to 44 years old among men students less than other age groups, although there was no significant difference seems to have not shown the statistical tests.

The average score of knowledge in other age groups of women is not seen as much difference in students and staff.

Evaluate of knowledge scores in terms of the level of work experience in two groups.

Figure 2. evaluate of knowledge in terms of the level of work experience in two groups

According to Figure 2, the average score in people with work experience of one to three is less than the other age groups

Although statistical tests did not show any significant difference.

The average score of knowledge in other groups of work experience was not seen much difference in students and staff.

Evaluate of knowledge scores in terms of gender and level of experience in the two groups.

Figure 3. evaluate of knowledge in terms of gender and level of experience in the two groups.

According to left figure 3,the average score in people who have had work experiences of between one and three and were men was less than the rest,it seems that although statistical tests did not show significant differences.The average scores of knowledge in other groups of history and gender were not seen much difference in students and staff.

According to right figure 3, the average score in people who have a history of between 3 and 5 And have been female. The average score of knowledge in other groups of history and gender is not seen much difference in students and staff.

evaluate of knowledge by gender and age group in two groups.

Figure 4. evaluate of knowledge by gender and age group in two groups

According to left figure 4,the average score in people who have been ages 34-25 And female have been less than others.

Statistical tests did not show a significant difference, though.The average score of knowledge in other age groups and gender is not seen much difference in students and staff.

According to right figure 4, the average score for people aged between 54 and 35 was less than the rest.

It seems that although statistical tests do not show significant differences. The average score of knowledge in other age groups and gender is not seen much difference in students and staff.

Discussion

The rapid development of technology in recent decades has led to the production of a huge amount of data.

These data are called big data. The familiarity with Big Data analyzes is of great importance due to many benefits, including cost and error reduction and make decisions. This study examined the challenges of lack of knowledge and expertise in the two groups of students and staff as a barrier to exploiting the benefits. On average, the staff's knowledge about the concepts of Big Data was higher than the student's. In assessing the relationship between the level of knowledge of people in age groups and background Significant was reported. This means that depending on the gender and experience probably increases the amount of knowledge. Because it seems those male more than female are more interested in software engineering and management jobs and their work experience was a reason for gaining a higher knowledge score has been in them. It seems that challenges could be overcome by holding training courses, conferences, congress, recruiting specialist staff, etc.

Conclusion

The IT staff of organizations and companies should be familiarized with the concepts needed in the Big Data area. It is suggested that in future studies, students, doctors and other fields of study should be evaluated, and the challenges of the Big

Data analysis should be investigated from their viewpoint. because students can provide a base for familiarizing and applying useful analysis by performing new research in this area. In other businesses, finding out the extent of their familiarity with the Big Data analyzes could be useful in applying managerial and advertising policies. Big Data analyzes could have a constructive role in all industries, and today, these analyze have become widespread in most industries and businesses. Because given the growing trend of data production, big data analysis in the coming years would become a requirement for all industries and areas.

Declarations

List of abbreviations

IOT: Internet Of Things

IT: Information Technology

Availibility of data and methods

All data generated or analysed during this study are included in this published article.

Competing interests

The author(s) declared no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research,authorship, and/or publication of this article.

Authors' Contributions

Study conception and design: Elham Nazari, Hamed Tabesh;

Acquisition of data: Elham Nazar, Parnian Asgari

Analysis and interpretation of data: Elham Nazari, Hamed Tabesh;

Drafting of the manuscript: Hamed Tabesh, Elham Nazari, Parnian Asgari

Critical revision: Hamed Tabesh;

Acknowledgement

The present study was carried out as a result of a research project approved by Medical Sciences University of Mashhad with the code 961731.

References

1 .Archenaa J, Anita EM. A survey of big data analytics in healthcare and government. Procedia Computer Science. 2015 Jan 1;50:408-13.

2. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. BioMed research international. 2015;2015. Häyrinen K, Saranto K, Nykänen P.

3. Kim GH, Trimi S, Chung JH. Big-data applications in the government sector. Communications of the ACM. 2014 Mar 1;57(3):78-85.

4. Raj P, Raman A, Nagaraj D, Duggirala S. Big Data Analytics for Healthcare. InHigh-Performance Big-Data Analytics 2015 (pp. 391-424). Springer, Cham.

Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics. 2007;31(4-5):198-211.

Acharjya DP, Ahmed K. A survey on big data analytics: challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications. 2016 Feb 1;7(2):511-8.

Kuo MH, Sahama T, Kushniruk AW, Borycki EM, Grunwell DK. Health big data analytics: current perspectives, challenges and potential solutions. International Journal of Big Data Intelligence. 2014 Jan 1;1(1-2):114-26.

8.Hermon R, Williams PA. Big data in healthcare: What is it used for? 2014.

9.Bossé É, Solaiman B. Information fusion and analytics for big data and IoT: Artech House; 2016.

Figures

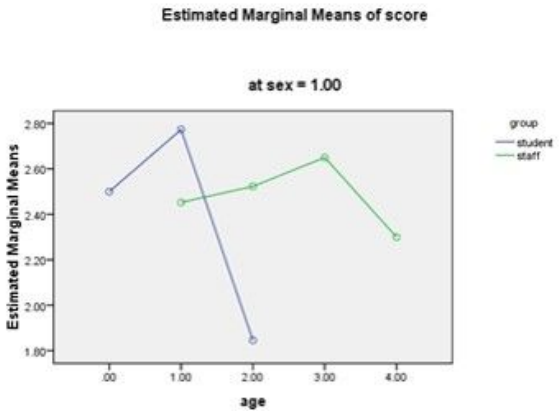

Figure 1

Evaluated scores by age group in two groups.

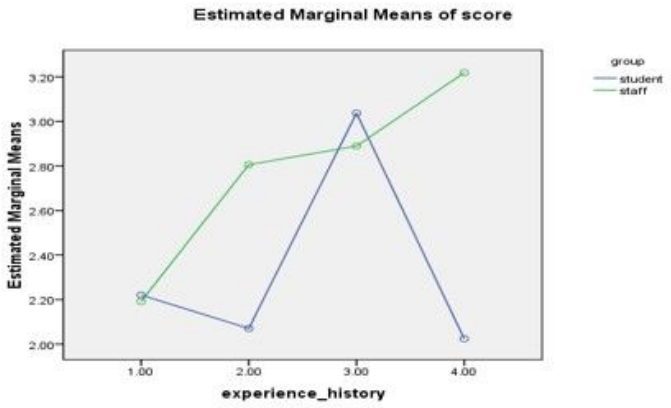

Figure 2

evaluate of knowledge in terms of the level of work experience in two groups

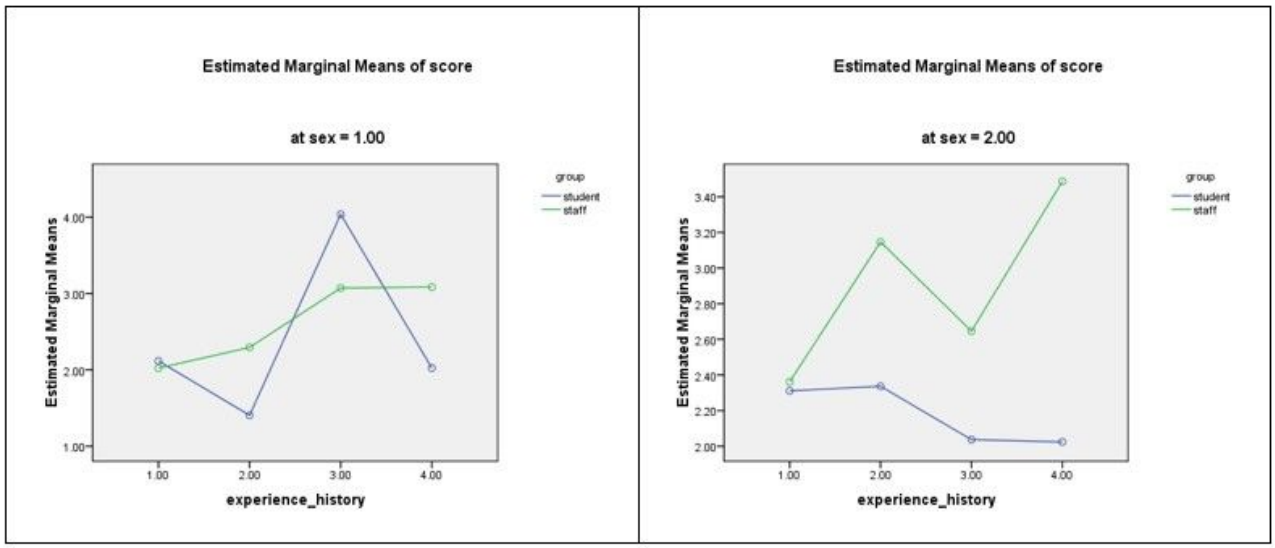

Figure 3

evaluate of knowledge in terms of gender and level of experience in the two groups.

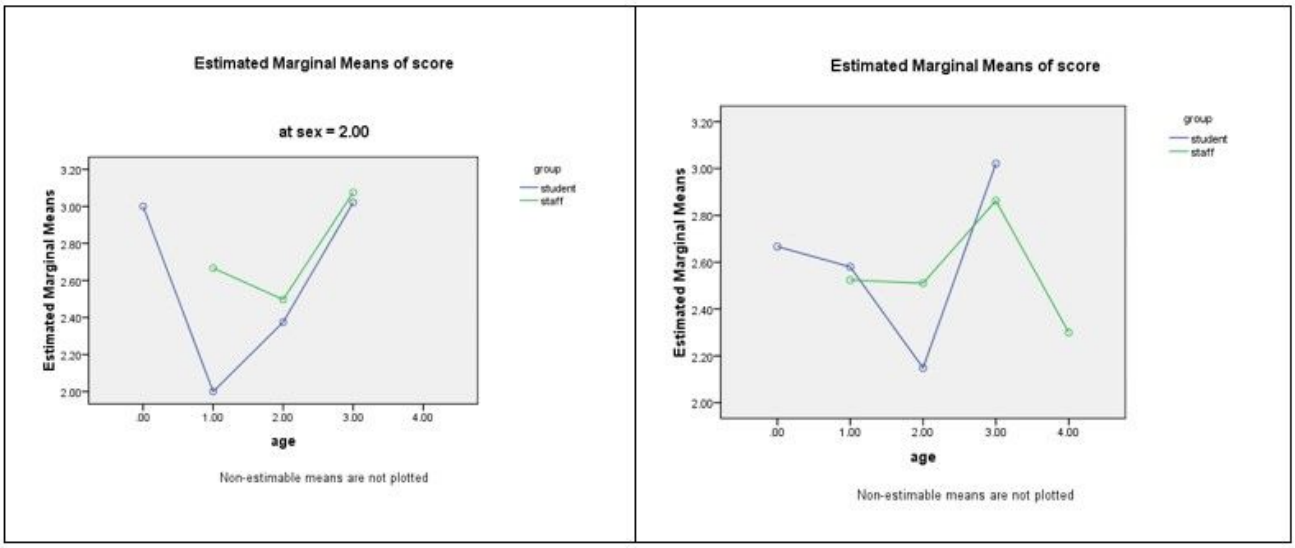

Figure 4

evaluate of knowledge by gender and age group in two groups