An Investigative Study on University Students' Statistical Literacy in Pakistan

* Anisul Hassan, PhD Scholar
** Dr. Abdul Ghaffar, Assistant Professor
*** Dr. Amir Zaman, Assistant Professor

Abstract

The main aim of universities is to conduct research studies on different aspects of human life for the purpose of development and prosperity. In this regard the importance of statistical literacy cannot be denied in every discipline of knowledge. Therefore, the main objective of this study was to investigate the statistical literacy of the university students at BS level. Quantitative survey research design was employed where sample of the study comprised of randomly selected 360 BS level students from 9 public sector universities of Khyber Pakhtunkhwa (Pakistan). The data were collected through self-developed instrument namely Instrument for Statistical Literacy (ISL). The study demonstrated that students had low statistical literacy at BS level; it was therefore, recommended that Higher Education Commission of Pakistan may include a significant portion of practical work in the approved curriculum at university level for the enhancement of statistical literacy.

Keywords: Statistical Literacy; Instrument for statistical Literacy (ISL); SPSS

Introduction

In today's data-driven technological society all the citizens whether they are students, experts, researchers and business men also deal with statistics in all walks of life. Likewise, the university students also use statistics for statistical analysis purposes in research projects. In universities the prime aim is to conduct researches in different disciplines. Due to this importance research projects at BS, Master, M.Phil and PhD level are obligatory. High level skills and knowledge of statistics are needed for the completion of these research projects. Considering this need Higher Education Commission of Pakistan (HEC, 2013) approved curricula containing statistical knowledge and skills for BS level programs aimed that in conducting researches the students have to encounter statistics for data collection, presentation, analysis and interpretation purposes. For this purpose statistics offers different sampling techniques for data collection; tables and graphs for data presentation, descriptive statistics (mean, mode and variance), inferential statistics, hypotheses testing for data analysis and interpretation.

Statistical Literacy

The term statistical literacy is the amalgamation of two words "statistics" and "literacy". Statistics is the aggregate of numerical data or the unknown characteristic of sample data in plural sense. While in singular the procedure used for collecting, presenting, analysing and interpreting data is known as statistics (Shah, 2018). However, descriptive Statistics and inferential statistics are required for effective production, interpretation and examination of data (Wild & Pfannkuch, 1999). The procedure used for describing and summarizing data like tables, charts, graphs, mean, mode, median etc without taking conclusions about the whole population is known as descriptive statistics (Walpole, 1982); while using sample data and probability to make estimations, predictions, decisions and generalizations about the population is inferential statistics (Illowsky, 2017).

Literacy is the ability of reading, writing, speaking and listening (Inglis & Aers, 2008). In the present age of functional literacy the ability of evaluating and analysing the written materials is considered as literacy (Schield, 2002). However, to comprehend the literacy completely the context knowledge is necessary (Fransman, 2005). In these; cognitive skill context by Street (2004), daily life context by Gee (1999) and numeracy by Gal (2000) have been presented. According to Gal (2000) numeracy is the ability of interpreting and communicating the statistical information. This indicates that interpreting and communicating the statistical information is another context of literacy.

_____
*   Department of Education, Abdul Wali Khan University Mardan Email: **anisulhassan81@gmail.com**
**   Department of Education, Abdul Wali Khan University Mardan Email: abdulghafar@awkum.edu.pk
*** Department of Education, Abdul Wali Khan University Mardan Email: dr.amirzaman@awkum.edu.pk

The writing and speaking ability of someone about statistical messages is statistical literacy (Carmichael, 2010) but this definition has the scarcity of evaluating processes of statistical messages and other ideas (Yotongyos, Traiwichitkhun, & Kaemkate, 2015). Context knowledge is indispensible in comprehending the statistical literacy (Callingham, & Watson, 2005; Sanchez, 2007). In the daily life context Wallman (1993) considers statistical literacy as the understanding and critical evaluating ability of someone about statistical outcomes given in his/her daily life. On the other hand, Schield (1999) considers statistical literacy as competency more than just reading the statistical messages. Therefore, a statistically literate person must know and be able to differentiate between the statistical terminologies like sample statistic, population parameter, casual and non-causal relations. Watson and Callllingham (2003) say that statistical literacy is the accurate use of statistical terminologies, formulas and statistical information not mare the understanding of these. According to Townsend (2006) a statistical literate person has the abilities of understanding, communicating, evaluating, interpreting and application of statistical information. In research context Gal and Garfield (1997) take in to account the statistical literacy as the interpretation of the results given in the previous researches. In this paper statistical literacy is defined as the understanding of statistical symbols, terminologies (Garfield, 1999 and Watson, 1997) and evaluating critically the arguments using statistics as evidence (Schield, 2001).

In literature much focus has been given on the investigation of statistical literacy at school (Watson, 2003; Doyle, 2008) as well as at college level (Rumsey, 2002; Del Mas, 2002) in statistics courses. In Pakistan BS education acts like a gateway for research work in different fields and is an effective mean of developing skills and knowledge for doing researches. Similarly at BS level the students have non-statistical and diverse educational backgrounds. In non-statistical disciplines statistics and research methodologies are required as basic courses. That's why this paper tries to investigate the statistical literacy of the BS students.

Objectives of the Study
1. To measure the level of statistical literacy of the university students
2. To find out the difference between the statistical literacy of the applied science and social Science students at university level

Research Question
1. What is the level of statistical literacy of the University students

Research Hypothesis
1. There is no significant difference between the statistical literacy of the applied science and social Science students at university level.

Methodology
Participants
In Khyber Pakhtunkhwa according to HEC (2018) there were 35 universities and (DAI) Degree Awarding Institutions (24 are public & 11 are private). As the study investigated the statistical literacy of university students hence it was necessary to investigate the statistical literacy of those BS students who studied statistics course recommended by HEC (2013). Keeping this in mind six departments three from applied science (Biotechnology, Chemistry, Zoology) and three from social science (Education, Psychology and Sociology) have been taken. Only 9 public sector universities have the mentioned six departments. There was 6112 number of students at BS level in these 9 universities. Firstly, 6 out of 9 universities were randomly selected and then 360 students from these six universities (10 from each department) were randomly selected. The sample size is reasonable (Creswell, 2012). The sample of students were taken from 8th semester only the reason behind this was that (HEC, 2013) offers statistics courses in different semesters (See table 1) so this was obligatory that the respondents under observation must have studied the recommended statistics courses.

Table 1

Detail of statistics courses offered in different semesters

S.No	Discipline	Statistics course	Semester
1	Sociology	Mathematics (Statistics)	1st
2	Zoology	Biostatistics	2nd
3	Chemistry	Statistics	2nd
4	Bio-technology	Probability and Biostatistics	4th

5	Psychology	Statistics in Psychology	5th
6	Education	Introduction to Statistics	6th

Instrument

Instrument for Statistical Literacy (ISL) developed by the researcher was used for data collection. ISL consists of 20 multiple choice questions as recommended by Yolcu, (2012). From the content analysis; the common topics were dig out from the curricula recommended by (HEC, 2013) for the selected six departments and question were erected for these. Knowledge elements with SPSS (Statistical Package for Social Sciences) addition of the Gal (2004) model were used. ISL is categorized in five components namely; Statistical Symbols (SS), Central Tendencies (CT), Descriptive Statistics (DS), Inferential Statistics (IS) and SPSS. The Cronbach's $\alpha = 0.91$ of ISL was calculated which according Cohen, Manion, & Morrison (2007) shows high reliability.

Analysis and Results

The five components of ISL were analysed with the help of SPSS version 20.The component wise and overall descriptive results of statistical literacy are given.

Table 2:

The component wise and overall descriptive statistics of ISL

Component of ISL	N	Minimum	Maximum	Mean	% age	Std. Deviation
Statistical Symbols (SS)	360	0	1	0.466	46	0.499
Central Tendencies (CT)	360	0	4	1.455	36	1.213
Descriptive Statistics (DS)	360	0	7	2.927	48	2.341
Inferential Statistics (IS)	360	0	4	1.619	29	1.211
SPSS	360	0	3	0.769	19	0.952
Over all (ISL)	360	0	17	6.897	34	4.481

Table 2 shows the descriptive analysis of the five components of ISL. In the first component "Statistical Symbols" the mean scores (0.466) (Minimum= 0, Maximum= 1) with standard deviation (0.499) shows below average understanding of the university students about the statistical symbols. Only 46% of the university students can understand the symbols and terminology used for sample and population mean, variance and for inferential statistics like ANOVA, z-statistic and t-statistic. In second component the mean score (1.455) (Minimum= 0, Maximum= 4) with standard deviation (1.213) shows that less number (36%) of university students can understand and calculate the meaning of mean, median and variance. In the descriptive statistics component the mean scores (2.927) (Minimum= 0, Maximum= 7) with standard deviation (2.341) demonstrates below average understanding of the descriptive statistics. Data shows that 48% of the universities students can; understand graphical representation of data, differentiate different types of variables, recognize the correlation and its types like negative, positive and no correlation, and calculate the percentage. The mean scores (1.169) (Minimum= 0, Maximum= 4) with standard deviation (1.211) in the inferential statistics component validates that less number (29%) of students can comprehend the probability and sampling techniques for the selection of sample units and inferential statistics like z and t etc. In SPSS component; very low 19% literacy about operating SPSS software has been identified for the university students as the mean scores (0.769) (Minimum= 0, Maximum= 3) with standard deviation (0.952) displays . The overall mean scores (6.897) (Minimum= 0, Maximum= 17) with standard deviation (4.481) of ISL indicates the low 34% statistical literacy of the university students.

Table 3:

Inferential statistics of the Instrument for Statistical Literacy (ISL) of the two groups

Groups	N	Mean	Std Deviation	Independent sample t-test $\alpha=0.05$		
				Df	T	Sig
Applied Sciences	180	7.17	4.610	358	1.165	0.245
Social Sciences	180	6.62	4.346			

Table 3 confirms the inferential analysis of the respondents of the two groups. (7.17) mean score of applied science students displays better performance than the mean score (6.62) of the social science

respondents however the variation (4.610) in the applied science data is higher than the variation (4.346) in the social science data. The significant two tailed t value 0.245 > 0.05 interprets no significance difference in the two groups. Hence from the inferential statistics it is inferred that there is no significant difference between the applied science and social science students' statistical literacy at university level.

Findings

Findings of the study evidenced that

1.   46% of the university students have the ability to know the statistical terminologies and symbols used for population, sample and inferential statistics.
2.   Less number of university students i.e. 36% can understand the meaning of central tendencies (mean, median and variance) and can calculate them.
3.   University students have below average understanding of the descriptive statistics. Results showed that 48% of the universities students can understand the graphical representation of data, correlation and its types.
4.   University students have low 29% literacy regarding inferential statistics like sample & sampling techniques, probability & probability distributions (t-distribution, normal distribution etc.).
5.   Only 19% of the university students have the ability to operate SPSS software for statistical data analysis purposes.
6.   34% of the university students are statistically literate.
7.   At BS level no significant difference was there in the statistical literacy of the applied science and social Science students.

Discussions and Interpretations

The results of this study showed below average understanding of the university students about symbols and terminologies used in statistics for distinct purposes. Watson and Kelly, (2008) also probed low achievements of the students in this respect. Students exhibited low performance in central tendencies while the studies of (Koleza, & Kontogianni, 2013; Mokros & Russell, 1995) showed better performance of the students. About half of the university students replied correctly in descriptive statistics same as the results of (Schield, 2006; Toluk-Uçar & Akdoğan 2009). In concepts of inferential statistics like probability, sampling techniques, t-test, chi-square and ANOVA very low understanding of the university students is investigated while (Watson & Moritz, 2000 and Yolcu, 2012) showed high understanding of the students in inferential statistics. This study investigated the very low understanding of the university students about operating SPSS software for statistical data analysis in contrast, (Gonulal Loewen, & Plonsky, 2017) examined high performance of the students in operating the SPSS software for statistical data analysis purposes.

Recommendations

1.   As the results showed that university students have low literacy regarding statistics. In today's data-driven technological society it is not up to the mark. Therefore, it is recommended that the curriculum approved by higher education commission of Pakistan (HEC) may have essential allocation of practical work which could improve the literacy of the university students regarding statistics.
2.   As majority of the university students don't know how to operate the SPSS software for statistical analysis. Therefore it is recommended that universities authorities may provide training opportunities and may arrange sessions, seminars, workshops etc with the support of (LID) Learning Innovation Division of HEC (Pakistan) in order to develop the operational skills about SPSS and other software like Stata, R, Nvivo etc. Additionally, it is recommended that (CRC) Curriculum Review Committee of HEC (Pakistan) may add a distinct portion in the curricula about statistical data analysis software.
3.   The results displayed low literacy of the university students about statistics so it is recommended for future researchers to investigate the reasons of low statistical literacy at university level.

References

Callingham, R. & Watson, J.M. (2005). Measuring statistical literacy. Journal of Applied Measurement, 6, 19-47.

Carmichael, C. S. (2010). The Development of Middle School Children's Interest in Statistical Literacy. Unpublished Doctoral Dissertation, University of Tasmania, Tasmania. Pp. 10-15.

Cohen, L., Manion, L., & Morrison, K. (2007). Research methods in education. Sixth edition 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN.

Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative. 4th Edition Upper Saddle River, NJ: Prentice Hall. Pp. 146-166

Del Mas, R. C. (2002). Statistical literacy, reasoning, and learning: A commentary. Journal of Statistics Education, 10(3). *Journal for Research in Mathematics Education*, *26*, 20–39.

Doyle, P. G. (2008). *Developing statistical literacy with students and teachers in the secondary mathematics classroom* (Doctoral dissertation, The University of Waikato).

Fransman, J. (2005). Understanding literacy: a concept paper. The EFA Global Monitoring Report, Literacy for Life, 31.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. *The assessment challenge in statistics education*, 1-13.

Gal, I. (2000). Adult Numeracy Development: Theory, Research, Practice. Series on Literacy: Research, Policy, and Practice. Hampton Press, Inc., 23 Broadway, Cresskill, NJ 07626.

Gal, I., (2004) "Statistical Literacy: Meanings, Components, Responsibilities", in Zvi, D., and J. B., Garfield, 2004, The Challenge Of Developing Statistical Literacy, Reasoning, and Thinking, Kluwer Academic Publishers, Dordrecht, The Netherlands. pp. 47-78

Garfield, J. (1999). Thinking about statistical reasoning, thinking, and literacy. First Annual Roundtable on Statistical Thinking, Reasoning and Literacy (STRL-1).

Gee, J. 1999. An Introduction to Discourse Analysis: theory and method. Routledge: London Falmer Press: London.

Gonulal, T., Loewen, S., & Plonsky, L. (2017). The development of statistical literacy in applied linguistics graduate students. *ITL-International Journal of Applied Linguistics*, *168*(1), 4-32.

https://hec.gov.pk › Documents › 2012-2013 ›

https://hec.gov.pk › Documents › 2018 ›

Inglis, F., & Aers, L. (2008). *Key concepts in education*. Sage.

Illowsky,B. (2017). Introductory Statistics Open Stax Rice University 6100 Main Street MS-375 Houston, Texas 77005. P.5.

Koleza, E., & Kontogianni, A. (2013, January). Assessing statistical literacy: What do freshman know. In *Proceedings of the 8th Congress of the European Society for Research in Mathematics Education, CERME* (Vol. 8).

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. Journal of Statistics Education, 10(3).

Sanchez, J. (2007). Building statistical literacy assessment tools with the IASE/ISLP. International Association for Statistical Education (IASE)/International Statistics Institute (ISI) Satellite.P.1

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. of Significance Journal. The Association of Public Data Users, 1(1).Pp. 15-20.

Schield, M. (2001, August). Statistical Literacy: Reading Tables of Rates and Percentages. In ASA Proceedings of Statistical Education Section.

Schield, M. (2002). Three kinds of statistical literacy: what should we teach? ICOTS6. Durban, South Africa.

Schield, M. (2006). Statistical literacy survey analysis: Reading graphs and tables of rates and percentages. In *Proceedings of the Sixth International Conference on Teaching Statistics*. Ciudad del Cabo: International Statistical Institute and International Association for Statistical Education.

Shah, J. (2018). A Text book of Statistics for class XI. Khyber Pakhtunkhwa Text Book Board Peshawar Pakistan. P.1

Street, B. 2004 Understanding and defining literacy. Background Paper for EFA Global Monitoring Report 2006.

Townsend, M. (2006). Developing statistical literacy in youth: Statistics Canada's education outreach program. Statistika–Journal for Economy and Statistics, 5.

Uçar, Z. T., & Akdoğan, E. N. (2009). 6-8. Sınıf öğrencilerinin ortalama kavramına yüklediği anlamlar. İlköğretim Online, 8(2), 391-400.

Walpole, R. S. (1988). Introduction to Statistics Third Edition (Pengantar Statistika, Edisi ke-3),(Alih bakasa Bambang Sumantri). *Penerbit PT Gramedia Pustaka, Jakarta*.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. Journal of the American Statistical Association, 88(421), 1-8.

Watson, J. M. (1997). Assessing statistical thinking using the media. The assessment challenge in statistics education, 12, 107-121.

Watson, J. M. (2003). Is statistical literacy relevant for middle school students? *The Bulletin of the International Statistical Institute 54th Session*, *60*, 68-71.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal, 2(2), 3-46.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for research in mathematics education*, 44-70.

Watson, J. M., & Kelly, B. A. (2008). Sample, random and variation: The vocabulary of statistical literacy. *International Journal of Science and Mathematics Education*, 6(4), 741-767.

Wild, C. J. & Pfannkuch, M. (1999).Statistical thinking in empirical enquiry (with discussion). International Statistical Review, 67(3), 223–265.

Yolcu, A. Y. Ş. E. (2012). An investigation of eighth grade students' statistical literacy, attitudes towards statistics and their relationship. Master's Thesis, Middle East Technical University, Ankara. P.12

Yotongyos, M., Traiwichitkhun, D., & Kaemkate, W. (2015). Undergraduate Students' Statistical Literacy: A survey study. Procedia-Social and Behavioral Sciences, 191, 2731-2734.