

Research Article

Open Access

Jin Zhang, Yanyan Wang, Yuehua Zhao*, Xin Cai

Applications of inferential statistical methods in library and information science

<https://doi.org/10.2478/dim-2018-0007>

received December 13, 2017; accepted February 8, 2018.

Abstract: Research methods play an extremely important role in studies. Statistical methods are fundamental and vital for quantitative research. The authors of this paper investigated the research papers that used statistical methods including parametric inferential statistical methods, nonparametric inferential statistical methods, predictive statistical correlation methods, and predictive statistical regression methods in library and information science and examined the connections and interactions between statistical methods and their application areas including information creation, information selection and control, information organization, information retrieval, information dissemination, and information use. Both an inferential statistical method and graphic clustering visualization method were employed to explore the relationships between statistical methods and application areas and reveal the hidden interaction patterns. As a result, 1821 research papers employing statistical methods were identified among the papers published in six major library and information science journals from 1999 to 2017. The findings showed that application areas affected the types of statistical methods utilized. Studies in information organization and information retrieval tended to employ parametric and nonparametric inferential methods, while correlation and regression methods were applied more in studies in information use, information dissemination, information creation, and information selection and control field. These findings help researchers better understand the statistical method orientation of library and information science studies and assist educators in the field to develop applicable quantitative research methodology courses.

Keywords: Statistical methods; application area; research methods; library and information research.

1 Introduction

The growth of different research fields is primarily driven by studies in these fields. As these research fields develop and the number of studies increase, the scope of these fields expands. Since the quality of research is determined in part by the research methodologies used, utilization of these methodologies, both quantitative and qualitative, has become more and more important. Research methodologies play an indispensable role in several research fields.

As one of the main branches of quantitative methods, statistical methods are regarded as major factors in determining the quality of studies. There are multiple categories of statistical methods, such as inferential statistical methods, predictive statistical methods, parametric statistical methods, and nonparametric statistical methods. Each category contains a lot of statistical approaches. For instance, ANOVA test is an inferential statistical method, while Pearson's correlation is a predictive statistical method. These various statistical methods have different characteristics and are used differently in studies. Therefore, it is meaningful to investigate the use of statistical methods in studies. This study investigated the use of four types of statistical methods: parametric inferential statistical methods, nonparametric inferential statistical methods, predictive statistical correlation methods, and predictive statistical regression methods.

As the scope of research fields expands and new research areas develop, studies using statistical methods are applied to more application areas than ever before. This phenomenon also occurs in the field of library and information science. For example, with the improvement in social media, social tagging emerged as a new application area. New research questions arise in the new and different research areas. Their corresponding collected data are different, leading to the employment of different

*Corresponding author: Yuehua Zhao: University of Wisconsin Milwaukee, Milwaukee, WI United States, E-mail: yuehua@uwm.edu
Jin Zhang, Yanyan Wang, Xin Cai: University of Wisconsin Milwaukee, Milwaukee, WI United States

statistical methods as well. Six application areas were defined in this paper, including information creation (IC), information selection and control (ISC), information organization (IO), information retrieval (IR), information dissemination (ID), and information use (IU).

The primary research problems addressed in this paper are: to what research areas in library and information science are statistical methods applied; what statistical methods are used in the field; and what interactions occur between statistical methods and application areas. The investigated studies came from major scholarly journals in library and information science from 1999 to 2017. Both qualitative and quantitative research methods were applied in this study, such as coding method, Chi-square test, and graphic clustering visualization method.

Implications of this study included the following: a holistic picture of statistical methods used in the contexts of research topics and areas; an understanding and selection of appropriate statistical methods to assist researchers in their research topics; and help for educators who teach quantitative research methods and statistical methods in library and information science to develop appropriate teaching plans and syllabi.

2 Literature review

Research is the effort to discover new knowledge and explore answers to scientific problems. Kothari (2011) defined the research as an art of intelligent investigation into the unknown. The main motivation that encourages researchers to undertake research is the desire to find out the hidden truths in the world. On the voyage of knowledge exploration, scientists launch their studies based on the previous stock of knowledge and present their original research contributions in journal papers, conference presentations, books, etc. The increased amount of research advances almost every field in modern times. Methodology is usually a key component in conducting the research.

Research methodology presents the various steps adopted by a research team to study a specific research problem as well as the logic behind these steps (Kothari, 2011). From the perspective of physics researchers, Rajasekar, Philominathan, and Chinnathambi (2006) indicated that the study of research methodology provides the necessary training in choosing methods, materials, and scientific tools for students and future researchers. Recently, social science schools have required students to take research methodology courses to both learn the

major works in the field and acquire pragmatic skills (Berg & Lune, 2011).

The quantitative approach involves the rigorous quantitative analysis of the data in quantitative form (Kothari, 2011). Quantitative research methodology focuses on observing issues from a problem-solving perspective that is highly structured and that relies on quantification, measurement, and evaluation (Connaway & Powell, 2010). Mixed research methodology is a research strategy that uses both qualitative and quantitative methodologies.

Statistics refers to a branch of mathematical procedures which deals with organizing, summarizing, and interpreting information (Gravetter & Wallnau, 2013). Statistics is one of the most useful and powerful tools in data analysis for both academics and practitioners (Vaughan, 2001). The essence of statistics lies in the idea of inference (Johnson, 2009). As researchers collect sample data to answer specific quantitative research questions, statistical methods enable them to draw conclusions about a broader base of people, events, or objects compared with samples actually included in the study (Munro, 2005). Generalization and interpretation of the statistical results stemming from limited observations permit the researchers to infer or predict the characteristics of the larger population. Applications of statistical methods can be classified as inferential methods, predictive methods, and other methods.

Statistical method has been developed to deal with the research problems in several disciplines. A number of recent studies indicate an increasing employment of statistical inference in problem solving. It allows researchers to clarify and summarize a large body of information with hidden phenomenon. Significantly, social, biological, and physical sciences, all of which make use of observations of natural phenomena through sample surveys or experimentation, use statistics to develop and test new theories (Ott & Longnecker, 2008). Weisburd and Britt (2007) suggested that conducting studies about crime and justice would be virtually impossible without statistics. In the medical field, scientists apply statistics not only in medical issues but also in their scholarly journals. *Statistical Methods in Medical Research* provides a unique venue to discuss the use of statistics in medical research. In the library and information field, Williams and Winston (2003) examined 119 papers from five journals and proposed the importance of research methods and especially statistics in academic libraries' research.

The preferences of the research methods are different across various fields. When quantitative orientations are used, the studies are usually given more respect in many of the social sciences (Berg & Lune, 2011). This implies a

tendency of evaluating the value and strength of a study from its methodological aspects. An earlier study also supported this suggestion. Fisher (1936) identified that the application of statistics in a discipline may raise its rank from that of the social sciences to that of science. In the information era, the power of statistical methods along with the availability of specific statistical software (e.g. SPSS, SAS, and R) advances modern science.

Library and information science (LIS) literature can cast some light on the analysis of methodologies used in LIS research (Zhang, Zhao, & Wang, 2016; Zhang, Wang, & Zhao, 2017). Zhang, Zhao, & Wang (2016) focused on the comparison among the six major scholarly journals in LIS in terms of the characteristics of statistical methods used. Zhang, Wang, & Zhao (2017) investigated the status of statistical methods used in the field and the temporal change patterns during a 15-year period. The current study, however, examined the connections and interactions between statistical methods and their application areas by investigating the papers published in six LIS journals during a recent 19-year period. The data from 1999 to 2017 were analyzed. In other words, the data of four more years were added to this study. The research methods of these studies were different. A visualization method was employed to ascertain the connections and interactions.

Järvelin and his colleagues have conducted a series of investigations of the methodological evolution of LIS research based on journal publications (Järvelin & Vakkari, 1990; Järvelin & Vakkari, 1993; Tuomaala, Järvelin, & Vakkari, 2014). These studies examined the data analysis strategies and data collection methods, as well as data analysis methods, applied in LIS (Tuomaala et al., 2014). In addition to their studies, other researchers have provided insight into the overview of the research methods applied in a wide range of LIS studies in the early 21st century. With the taxonomies created by Jarvelin and Vakkari (1990), Hider and Pymm (2008) investigated 20 high-profile LIS journals in 2005 to compare the distribution of overall research strategies, the data collection methods, and the type of data analysis (quantitative or qualitative) in the librarianship journals and nonlibrarianship journals. The results showed that quantitative research methods accounted for more than half of the overall investigated papers (64.7%) and that these applications were more frequently used in the nonlibrarianship journals (69.1%) than in the librarianship journals (51.4%). In the most recent study, Tuomaala et al. conducted a longitudinal analysis of the evolution of the research topics and methodologies from 1965 to 2005. Similar to the findings of the previous studies, the authors also found that the most prevalent data analysis approach

was quantitative methods (58.4%). Moreover, through the examination of the LIS publications over the 40 years, the authors identified the quantitative core of LIS research as four areas: information storage and retrieval, scientific communication, library and information service activities, and information seeking.

Research methods play a significant role in scholarly efforts in the LIS field (Chu, 2015), and many scholars have explored LIS research methodology since the 1970s (Atherton, 1975; Bernhard, 1993; Blake, 1994; Peritz, 1980; Kumpulainen, 2009; Hersberger, 2009; Westbrook, 1994; Eldredge, 2004). The state of statistical methods applied in the research in the LIS field has also been discussed for several decades (Vaughan, 2001; Enger, 2006; Dilevko, 2007; Van Epps, 2012; Doucette, 2017). Table 1 summarizes the previous studies on the use of inferential statistical methods in the LIS field. Atherton (1975) reported that 12.8% of LIS papers undertook inferential statistics during 1969–1971 (compared with Van De Water, Surprenant, Genova, and Atherton (1976) at 13.4%; Wyllys (1978) at 2.9%; Wallace (1985) at 6%; and Enger, Quirk, and Stewart (1989) at 11.1%). Enger et al. (1989) then concluded that the use of inferential statistical methods increased from 1981 to 1985. More recently, the percentage of inferential analysis use in LIS appeared to be up to 18.5% (Togia & Malliari, 2017). Togia and Malliari (2017) also suggested that researchers should more consider the use of inferential statistical method. In the field of health library and information science, Dimitroff (1992) conducted a content analysis on research between 1966 and 1990. Dimitroff (1992) reported that the most frequently used research method was survey research method (41%), followed by observation (20.7%) and bibliometrics (13.8%). Quantitative descriptive analysis methods (83.5%) were found as the most common analytical technique, while quantitative inferential analysis methods (1.9%) were applied. The findings of the studies conducted from the 1970s to 2010s showed that there was an increase in the use of inferential statistical methods in papers appearing in LIS journals.

In conclusion, previous studies have analyzed the subjects and themes of LIS journals, while others have examined the research strategies and methods utilized by researchers. In addition, previous research has demonstrated that the use of quantitative methodology, especially statistical methods, performs a meaningful role in LIS research. However, few studies have investigated the employment of different statistical methods within the various research areas. There is a need for researchers to explore the relationships between the methodological aspects and research topics in the LIS field. The current study builds upon the findings of previous contributions

Table 1: Previous studies on the use of inferential statistical methods in the LIS field.

Studies	Time period	Reported use of inferential statistics
Atherton (1975)	1969-1971	12.8%
Van De Water, Surprenant, Genova, & Atherton (1976)	1969-1971 and 1974	13.4%
Wyllys (1978)	1975	3%
Wallace (1985)	1981	6%
Enger, Quirk, & Stewart (1989)	1985	11.1%
Dimitroff (1992)	1966-1990	1.9%
Dilevko (2007)	2001-2005	14.5%
Togia & Malliari (2017)	2011-2016	18.5%

and intends to provide a holistic picture of statistical method applications in diverse LIS research topics and areas.

3 Research method

To investigate the abovementioned research problems in this study, the following null hypothesis was presented:

H0: There is no significant relationship between the defined statistical methods and the application areas in research studies in library and information science.

Statistical methods were classified into the following four categories based on the nature of the statistical methods and data types: (1) parametric inferential statistical methods, (2) nonparametric inferential statistical methods, (3) predictive statistical correlation methods, and (4) predictive statistical regression methods. The application areas, based on information cycle, were defined as follows: information creation (IC), information selection and control (ISC), information organization (IO), information retrieval (IR), information dissemination (ID), and information use (IU) (Hodge, 2000). Both statistical methods and application areas were expanded to a second level of schema during the research process. The independent variables were the type of statistical methods and application area. The dependent variable was defined as the frequency of a type of statistical methods used in an application area.

3.1 Data collection

The research papers investigated were collected from six major library and information science journals (*Journal of the Association for Information Science and Technology (JASIST)*, *Journal of Documentation (JD)*, *Journal of Information Science (JIS)*, *The Library Quarterly (LQ)*, *Library and Information Science Research (LISR)*, *Information Processing & Management (IPM)*) from 1999 to 2017.

The criteria for choosing the six journals were as follows: (1) the journals should be the top journals in the LIS field in terms of both domain experts' suggestions and impact factor; (2) they should be research-oriented journals; (3) they should have more than 19 years of history, which can allow the researchers to collect sufficient data; (4) the journals' research papers should have their full texts provided online so that the researchers can identify statistical methods used in each paper as well as the statistical methods' application areas.

After the six journals were determined, all research papers published in these journals from 1999 to 2017 were identified and examined by the researchers. For each of the examined research papers in a journal, its related fields such as title, publishing year, journal title, issue number, subject terms, abstract, and full text were recorded. After the content analysis on the abstract and full text was performed, the statistical method(s) were used and the corresponding application areas were identified and recorded.

When all the investigated journals were examined, the collected application areas and statistical methods were analyzed and categorized in a coding analysis process. The similar terms were combined, and semantically related terms were grouped. The corresponding statistical method(s) were specified, categorized, and recorded based on a statistical method classification system (Gravetter & Wallnau, 2009).

To identify the use of a specific statistical method in a paper, two researchers first reviewed the methodology of the paper to find whether it shows the use of any statistical methods and then looked for the corresponding results of those methods in the results part. All the statistical methods used in the paper were recorded.

After identifying the use of a statistical method and recording it, the researchers compared each method with others and mapped to a higher-level statistical approach. When a method was a special case of another one, this method was assigned to the latter. After several rounds of comparisons, all the special cases were assigned to the corresponding higher-level statistical approach groups.

These statistical approaches were then categorized into four types of statistical methods.

For each of the examined research papers, its application areas were also analyzed and categorized based on the six phases of information cycle (Hodge, 2000). In this paper, an application area means the research area where a statistical method was used. For the papers using statistical methods, their application areas were determined based on their subjects, titles, abstracts, and full texts by coding. A paper can have only one application area, and all the statistical methods utilized in a paper can have the same application area.

The coding method was utilized to generate the expanded application areas for the investigated papers. As mentioned earlier, each paper's subjects, titles, abstracts, and full texts were analyzed and the researchers generated subject notes to describe its application area of. Similar subject notes were categorized into one group, and an expanded application area was defined for each group. Several rounds of comparisons were conducted, and the final expanded application areas were identified. Each expanded application area was then classified into one of the six application areas.

The coding analysis results of statistical methods and application areas served as the detailed statistical approaches and expanded application areas in the later data analysis. They provided more detailed and thorough analyses for both statistical methods and application areas.

3.2 Inferential statistical analysis

To test and examine the proposed hypothesis mentioned in the previous section, Chi-square statistical analysis was applied to the collected data. The significance level (α) for this test was chosen to be 0.05. If the resultant p-value of the null hypothesis test was smaller than 0.05, the null hypothesis would be rejected. Otherwise, the hypothesis would be accepted. The inferential statistical analysis was carried out using SPSS (Version 22).

3.3 Information visualization analysis

A visualization analysis method was used in this study. The visualization method is not only intuitive but also a powerful data analysis method. It allows people to discover the interaction patterns of involved factors. The visualization method effectively visualizes the interaction

between the two factors and reveals the interaction patterns at different levels. It enables people to gain different perspectives of the interaction between the two factors.

The visualization method used in this study is called the "graphic clustering" visualization method (Rasmussen & Karypis, 2004). The visualization method generates a two dimensional space where one factor is clustered on the X-axis and another factor is clustered on the Y-axis. Each factor consists of a group of elements. In this study, one factor was the statistical research method, while the other factor was the application area. Elements referred to the detailed statistical methods and expanded application areas that were revealed in the coding analysis.

The visualization tool for this study was gCluto (Version 1.0), an open-source software. The software provides several clustering algorithms which can be selected by users. In this study, the repeated bisection clustering algorithm was used to cluster elements in each of the two factors. As a result, each factor corresponded to a hierarchical tree that was assigned to an axis of the visual space. After the two resultant hierarchical trees were assigned to their respective X-axis and Y-axis of the visual space, a matrix visual space was formed. The cells of the matrix are interactions between elements of one factor and elements of the other factor. Objects were projected onto the cells based on their relationships with the two factors in the visual space. As a result, related elements from both factors were clustered together to form an interaction area of interest. An area in dark red color in the visual space indicated that it was an interaction area of projected objects. An area in light color suggested that less or no objects were projected onto that area. In this study, objects were research papers using statistical methods. The visualization analysis was used to cluster the factors (expanded application areas and detailed statistical methods from these papers) in visualization displays to illustrate how statistical methods are associated with each other in the context of the application areas. In the same vein, visualization analysis was used to demonstrate how the expanded application areas were connected in the contexts of statistical methods in the visualization space:

1. The statistical methods vs. application areas
2. Statistical methods vs. expanded application areas
3. Detailed statistical methods vs. application areas
4. Detailed statistical methods vs. expanded application areas

4 Results and discussion

4.1 Coding result

The statistical methods applied in the investigated research papers were classified into four categories. According to Gravetter and Wallnau's (2013) opinion, there are two types of statistical methods: inferential statistical methods and predictive statistical methods. The statistical methods of the first type were further separated into two categories in terms of the analyzed data to which they were applied. These two categories were parametric inferential statistical method and nonparametric inferential statistical method. The methods that were suitable to both parametric and nonparametric data were assigned into the parametric inferential statistical method category. The statistical methods of the second type were also separated into two categories: the correlation method category and the regression method category. The four types of statistical methods and the corresponding statistical methods are summarized in Table 2.

There were six application areas that were identified according to the six phases of information cycle. Under each application area, there were multiple expanded application areas. As mentioned in the "Research Method" section, these expanded application areas were generated from the content of the investigated research papers by the coding method. Every research paper that applied at least one statistical method was manually reviewed and coded by two researchers, and the intercoder reliability kappa coefficient of the coding process was 0.671, between 0.61 and 0.80. This value implied a substantial agreement between the coders according to Viera and Carrett's (2005) work. This bottom-up coding process ensured that every paper's application area was considered by the researchers and can be covered by the final expanded application areas. The researchers were experts in LIS so that they were able to generate the reasonable expanded application areas for the papers. In this study, one research paper was assigned to only one expanded application area. The six application areas and the corresponding expanded application areas are listed in Table 3.

4.2 Descriptive result

In this study, six major scholarly journals (*JASIST*, *JD*, *JIS*, *LQ*, *LISR*, and *IPM*) were investigated. Research papers in these journals published from 1999 to 2017 were identified, analyzed, and examined. The number of investigated papers that employed statistical methods was 1821.

Table 2: The four types of statistical methods and the corresponding statistical methods.

Type of Statistical Methods	Statistical Approaches
1 Parametric inferential statistical methods	1.1 <i>t</i> -test
	1.2 ANOVA
	1.3 MANOVA
	1.4 Z-test
	1.5 Comparison test
	1.6 ANCOVA
2 Nonparametric inferential statistical methods	2.1 Chi-square test
	2.2 Wilcoxon's test
	2.3 Mann-Whitney U test
	2.4 Kruskal-Wallis test
	2.5 Kolmogorov-Smirnov test
	2.6 Kendall's W test
	2.7 Friedman test
	2.8 Fisher's test
	2.9 Sign test
	2.10 Binomial test
	2.11 McNemar test
3 Predictive statistical correlation methods	3.1 Correlation
	3.2 Pearson's correlation
	3.3 Spearman's correlation
	3.4 Order correlation
4 Predictive statistical regression methods	4.1 Regression
	4.2 Linear regression
	4.3 Logistic regression
	4.4 Multiple regression
	4.5 Hierarchical regression
	4.6 Cox regression

Keywords were extracted from titles, keywords, and abstracts of all of these investigated papers. After the validation process, a total of 991 keywords remained and the entire frequency of all the keywords was 12,769. In Figure 1, the X-axis is the appearance frequency of keywords and the Y-axis is the number of keywords which appeared as such frequency. The curve in the figure illustrates that most of the keywords had low frequencies and only a few keywords had high frequencies. The high frequency keywords reflected the focuses of the investigated papers.

The top 20 terms with the highest frequencies are listed in Table 4. It was not surprising that "information" and "library" were high on the keyword list. The fact that the second and third positions were occupied by "user" and "retrieval" demonstrated that the IU and IR attracted much of the research interest.

Table 3: The six application areas and the corresponding expanded application areas.

Types of Application Areas	Expanded Application Areas
1 Information creation (IC)	1.1 Research pattern
	1.2 Cover design
	1.3 Knowledge creation
	1.4 Research productivity
2 Information selection and control (ISC)	2.1 Publication evaluation
	2.2 Researcher and institution evaluation
	2.3 Information quality control
	2.4 Subscription selection
	2.5 Information source selection
	2.6 Impact of technology on information selection
	2.7 Information evaluation indicators
	2.8 Information privacy
3 Information organization (IO)	3.1 Indexing and abstracting
	3.2 Clustering
	3.3 Categorization
	3.4 Classification
	3.5 Labeling and tagging
	3.6 Metadata
	3.7 Internet webpage organization
4 Information retrieval (IR)	4.1 Multimedia retrieval
	4.2 Search behavior
	4.3 Query expansion
	4.4 Cross-language and bilingual retrieval
	4.5 Relevance judgment
	4.6 Information system and performance evaluation
	4.7 Information literacy
	4.8 Text summarization
	4.9 Retrieval algorithm and theory
	4.10 Web search
	4.11 Natural language processing
	4.12 Data mining
5 Information dissemination (ID)	5.1 User and community communication performance
	5.2 Information sharing
	5.3 Social media
	5.4 Publication choice
6 Information use (IU)	6.1 Information and content usage
	6.2 Information service
	6.3 Library service and usage
	6.4 User behavior
	6.5 Information management
	6.6 Web usage
	6.7 Citation behavior
	6.8 Learning and training performance
	6.9 Job satisfaction and development
	6.10 Information system usage

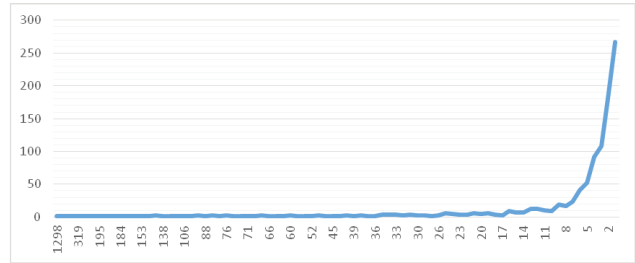


Figure 1: Keyword distribution across all the four statistical methods.

Table 4: Top 20 most frequent keywords across the four types of statistical methods.

Keyword	Count	Keyword	Count
information	1298	system	175
user	710	evaluate	170
retrieval	576	manage	153
behavior	319	seek	152
study	297	social	148
knowledge	273	web	148
search	195	research	138
analysis	194	performance	116
library	186	model	108
citation	184	communicate	106

When all the investigated papers were analyzed based on the types of statistical methods they adopted, the keywords from the papers were extracted and classified into the different types of statistical methods. After the validation process, 606, 523, 574, and 356 unique keywords were extracted from the papers that employed parametric inferential statistical methods, nonparametric inferential statistical methods, correlation methods, and regression methods, respectively. The total frequencies of all the keywords for these four types were 3206, 2534, 2914, and 1183, respectively. Table 5 lists the top 20 most frequent keywords in each of the four statistical methods. Regardless of the type of statistical methods, “information” occupied the first position on the keyword lists.

In this study, the total number of investigated statistical methods was 3046. The frequencies of the four types of statistical methods are shown in Figure 2. In the figure, the X-axis represents the type of statistical methods, and the Y-axis represents the number of uses of statistical methods. The number of parametric inferential statistical methods used in the investigated research papers was 1023, which was the most among the four method types,

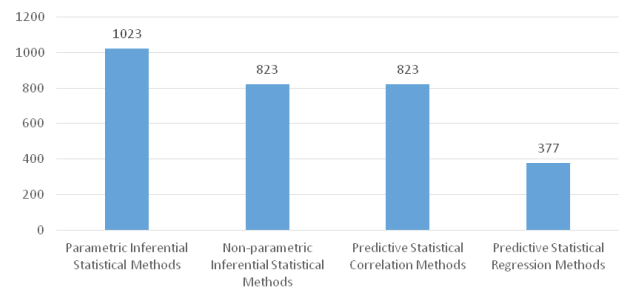
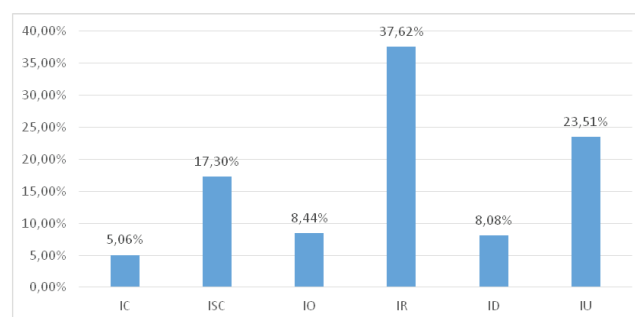
Table 5: Top 20 most frequent keywords in each of the four types of statistical methods.

Parametric inferential statistical methods		Nonparametric inferential statistical methods		Predictive statistical correlation methods		Predictive statistical regression methods	
Keyword	Count	Keyword	Count	Keyword	Count	Keyword	Count
information	382	information	254	information	266	information	110
retrieval	195	retrieval	145	user	144	user	57
user	169	user	140	retrieval	110	behavior	32
study	81	study	70	citation	79	citation	31
search	70	search	58	analysis	75	knowledge	29
behavior	66	behavior	50	behavior	65	retrieval	29
knowledge	46	web	39	study	65	library	22
library	44	library	36	evaluate	55	research	21
system	40	seek	35	knowledge	52	study	19
analysis	39	analysis	30	research	50	social	18
social	39	knowledge	28	impact	44	manage	17
evaluate	33	social	27	journal	42	web	16
research	32	citation	26	library	41	public	15
query	31	text	24	bibliometric	40	seek	15
online	29	evaluate	23	system	37	perform	14
seek	28	model	23	web	37	analysis	13
citation	27	system	23	perform	35	communicate	12
image	27	journal	21	social	35	search	12
web	26	image	20	search	33	bibliometric	11
health	24	academic	19	seek	26	evaluate	11

while the regression methods (377) were least employed in the investigated research papers.

The distribution of statistical methods in different application areas is shown in Figure 3. The X-axis in the figure represents the application areas, and the Y-axis represents the ratio of the number of statistical methods used in each area to the total number of statistical methods. Among the six application areas of the statistical methods, IR (37.62%), IU (23.51%), and ISC (17.30%) areas achieved first, second, and third places, respectively. IO (8.44%), ID (8.08%), and IC (5.06%) areas attracted less statistical methods than the other three areas and obtained fourth, fifth, and sixth places, respectively.

Figure 4 shows the frequencies of individual statistical methods used in the investigated research papers. The X-axis represents the individual statistical methods, and the Y-axis represents the frequencies of statistical methods used. Some authors did not state clearly, which specific statistical correlation or regression methods were used in their studies. As a result, it was difficult to classify these methods into specific statistical method groups, and they were put into a general correlation method or

**Figure 2:** Frequencies of the four types of statistical methods.**Figure 3:** Ratios of statistical methods used in the six application areas.

regression method as shown in Figure 4. *t*-test (480) clinched first place among all the methods. McNemar test (7) ranked in the lowest position.

Figures 5–10 show the frequencies of the four types of statistical methods used in the expanded application areas of IC, ISC, IO, IR, ID, and IU, respectively. In these figures, the X-axis of each figure represents the expanded application areas, and the Y-axis represents the frequency of a specific type of statistical methods.

Figure 5 shows the usage of the four types of statistical methods in the expanded application areas of the IC area. In the research pattern area, correlation methods (22) were used the most among the four statistical methods, and the mean frequency of the four statistical methods was 12.75. In the cover design area, the nonparametric inferential statistical methods (3) were used the most, and the mean frequency of the four statistical methods was 1.25. In the knowledge creation area, parametric inferential statistical methods (19) were used most frequently, and the mean frequency of the four statistical methods was 17. In the research productivity area, correlation methods (9) were used the most, and the mean frequency of the four statistical methods was 7.5.

The frequencies of statistical methods used in the ISC area are shown in Figure 6. Parametric inferential statistical methods (27) were used the most in the information quality control area. In the subscription selection area, both the parametric inferential statistical methods and the correlation methods (8 for each) were utilized the most among the four types of methods. The impact of technology on information selection area used parametric inferential statistical methods (10) more than the other three types of statistical methods. Correlation methods dominated in the other five expanded areas. The means of the frequencies for publication evaluation, research and institution evaluation, information quality control, subscription selection, information source selection, impact of technology on information selection, information evaluation indicators, and information privacy were 26, 35.25, 20.75, 6.5, 9.5, 6, 21, and 6.5, respectively.

Figure 7 shows the usage of statistical methods in the IO area. The most used statistical methods in the indexing and abstracting (19), clustering (19), classification (17), and labeling and tagging (27) areas were parametric inferential statistical methods. The most used statistical methods in the categorization area were nonparametric and parametric inferential statistical methods and correlation methods (6 for each). The metadata area used nonparametric inferential statistical methods 8 times, which was more than the other method types. The most used methods in the internet webpage organization area

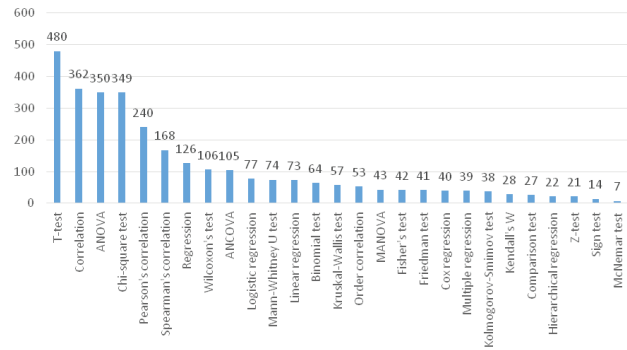


Figure 4: Frequencies of each individual statistical method.

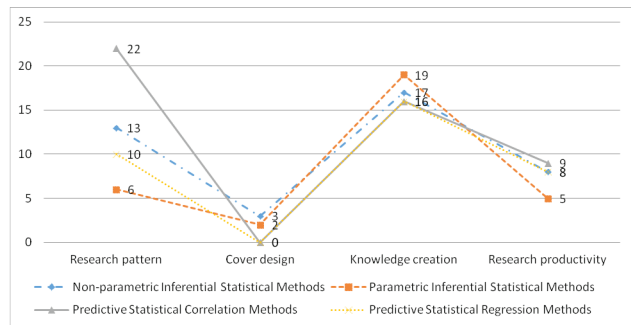


Figure 5: Frequencies of the four types of statistical methods used in the IC area.

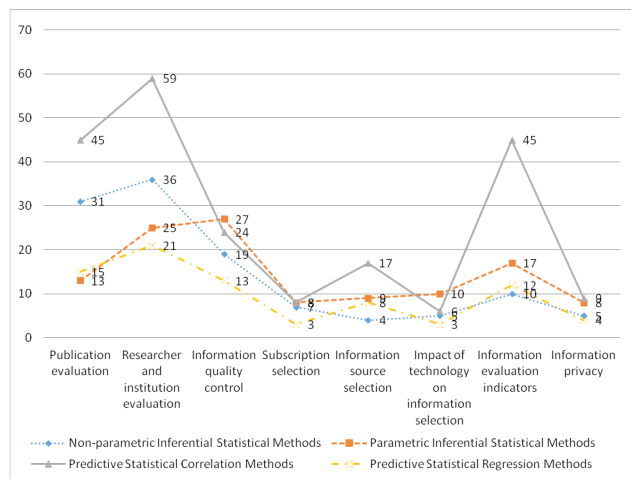


Figure 6: Frequencies of the four types of statistical methods used in the ISC area.

were nonparametric and parametric inferential statistical methods (14 for each). The least used methods in all the seven expanded areas were regression methods. The means of the frequencies for the indexing and abstracting, clustering, categorization, classification, labeling and tagging, metadata, and internet webpage organization

areas were 9.25, 10.75, 5, 9.25, 15.25, 4.25, and 10.5, respectively.

The usage of the four types of statistical methods in the IR area is shown in Figure 8. Parametric inferential statistical methods were used the most among the four types of statistical methods in 11 expanded areas. The exception was the search behavior area that utilized non-parametric inferential statistical methods (99) more than twice the parametric inferential statistical methods (97). The means of the frequencies for the multimedia retrieval, search behavior, query expansion, cross-language and bilingual retrieval, relevance judgment, information system and performance evaluation, information literacy, text summarization, retrieval algorithm and theory, web search, natural language processing, and data mining areas were 15.75, 70, 15, 6, 20.75, 38.5, 11, 8, 27.75, 24.25, 9.5, and 40.25, respectively.

Figure 9 shows the frequencies of the investigated statistical methods used in each expanded application area of the ID area. There were four expanded areas in the ID area. The most used methods in user and community communication performance, information sharing, and social media areas were parametric inferential statistical methods and correlation methods. The most used methods in the publication choice area were nonparametric inferential statistical methods (5) and regression methods (3). The least used methods in all the four expanded areas were regression methods. The means of the frequencies for the user and community communication performance, information sharing, social media, and publication choice areas were 21.75, 20.25, 15.75, and 3.75, respectively.

The numbers of each type of statistical methods utilized in the IU area are shown in Figure 10. The most utilized statistical methods in information and content usage (36), information service (25), library service and usage (33), user behavior (52), and information system usage (26) areas were parametric inferential statistical methods. The most used methods in the information management (18), web usage (23), and citation behavior (30) areas were correlation methods. The most used methods in learning and training performance were parametric and nonparametric inferential statistical methods (three for each). The most used methods in job satisfaction and development were parametric and nonparametric inferential statistical methods and regression methods (eight for each). The means of the frequencies for information and content usage, information service, library service and usage, user behavior, information management, web usage, citation behavior, learning and training performance, job satisfaction and development, and information system usage

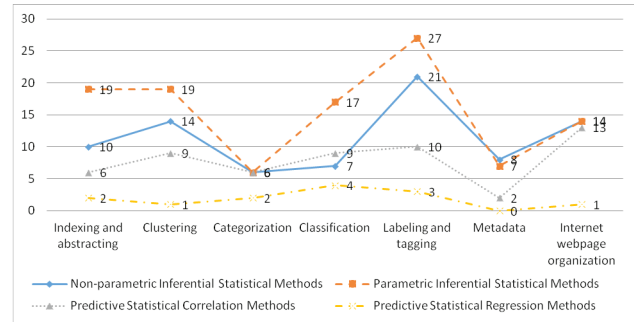


Figure 7: Frequencies of the four types of statistical methods used in the IO area.

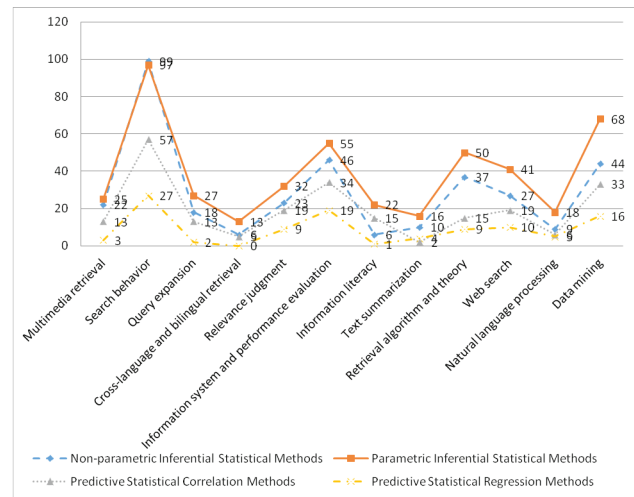


Figure 8: Frequencies of the four types of statistical methods used in the IR area.

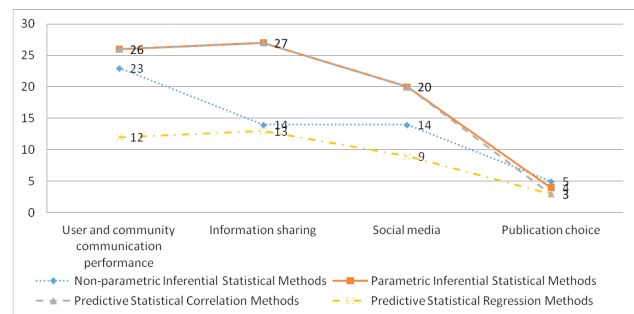


Figure 9: Frequencies of the four types of statistical methods used in the ID area.

were 28.25, 15.75, 27, 32.75, 14.75, 15.75, 16.75, 2, 6.75, and 19.25, respectively.

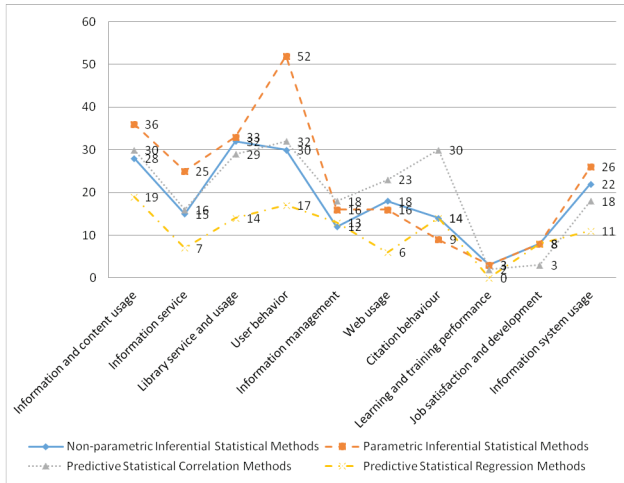


Figure 10: Frequencies of the four types of statistical methods used in the IU area.

4.3 Inferential result

The hypothesis (H0) was tested using a Chi-square test. In the Chi-square test, the result includes both a χ^2 -value and a p -value. The χ^2 -value is presented as χ^2 (df). Here df stands for the degrees of freedom. The result of the Chi-square test on H0 is summarized in Table 6. The result indicated that there was a significant relationship between the defined statistical method and the application area of studies in library and information science ($\chi^2(15) = 165.734$, $p\text{-value} = 0.000 < 0.05$). H0 was rejected. Therefore, the application areas of studies do affect which type of statistical methods is used. It means that studies of a specific application area tend to employ certain types of statistical methods.

4.4 Visualization result

When all the data were plugged into gCluto, two clusters were grouped based on the repeated bisection algorithm. In Figure 11, the X-axis represents the four types of statistical methods, and the Y-axis represents the six categories of application areas. The hierarchical trees in both axes depicted the relationships between objects by displaying the order in which objects were merged in the agglomerative process. As shown in Figure 11, in the X-axis, parametric inferential methods and nonparametric inferential methods were clustered together and correlation methods and regression methods were clustered together. The grouping results were consistent with the standard statistical method classification: parametric inferential methods and nonparametric inferential methods belong

Table 6: Results of H0 from the Chi-square test.

	χ^2 -Value	df	P -value
Pearson's Chi-square test	165.734	15	0.000

ged to inferential methods, while correlation methods and regression methods belonged to predictive methods. In the Y-axis, two high-level clusters emerged in the clustering process: IO and IR fell into the first cluster and the other four application areas formed another cluster. The hierarchical tree also showed that in the agglomerative process IU and ID were clustered together, and IC and ISC were connected together, and then these four application areas formed the second group of application areas. The grouping sequences identified the similarities between the application areas based on their employment of the statistical methods. In that sense, IO and IR, IU and ID, and IC and ISC showed more similarities to each other than to the others. The analogies of the application patterns of statistical methods between IO and IR, IU and ID, and IC and ISC also reflected the connections of the research between these fields.

The visualized matrix presented the interactions between application areas and statistical methods. The colors of the intersections were used to represent the values of the matrix. Considering the interactions as shown in Figure 11, IO and IR together presented mostly parametric inferential methods and rarely any regression methods. This suggested that papers that focused on IO and IR tended to employ parametric inferential methods instead of regression methods. IU and ID together applied mostly parametric inferential methods and correlation methods. IC focused on correlation methods and nonparametric inferential methods. ISC utilized more correlation methods than other methods. From the perspective of statistical methods, parametric inferential methods were widely applied in all application areas, especially in IO and IR. Nonparametric inferential methods were evenly utilized in different application areas. Correlation methods were employed more by the second cluster of application areas than the first cluster, and especially applied in ISC research. Similar to the results as shown in Figure 2, regression methods were not frequently adopted in the LIS field, but Figure 11 shows that these types of statistical methods were useful in the IC area.

More detailed data were then analyzed and visualized by gCluto. The matrix visualization in Figure 12 shows insights into the relationship between the application areas and specific statistical methods. In Figure 12, the X-axis represents the six categories of application

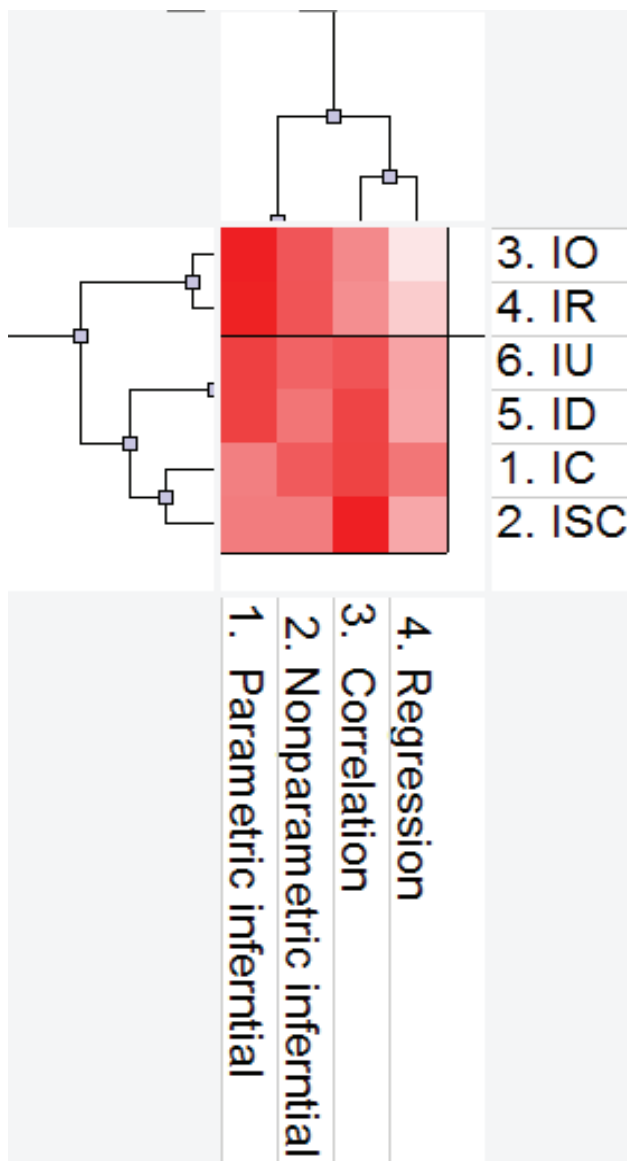


Figure 11: Matrix visualization of the interaction between application areas and statistical methods.

areas, and the Y-axis represents the 27 specific statistical methods applied in the investigated papers. Based on the applications of the detailed statistical methods, IU and ISC were grouped together and then grouped with ID and IC, while IR and IO were grouped together. It can be seen that the grouping results of the six application areas were slightly different from the outcomes as shown in Figure 11. The IR and IO still showed similar patterns in adopting detailed statistical methods, but the ID area was aligned with IU and ISC when it came to further method applications. On the Y-axis, 27 specific statistical methods were categorized into four clusters based on their utilizations in the six application areas. The grouping outcomes of the

specific methods did not follow the standard statistical method classification. Methods from the first and second cluster were primarily adopted by IR area. The first cluster included only nonparametric inferential methods (Binomial test, Wilcoxon's test, Mann–Whitney U test, Friedman test, and Fisher's test) which meant that these methods were commonly applied in the same application area, IR. The second cluster was composed of four of the most prevalent parametric inferential methods (*t*-test, ANOVA, MANOVA, and comparison test) and several nonparametric inferential methods (Kruskal–Wallis test, Kendall's W test, sign test, and McNemar test). The third cluster, occupied by correlation methods (correlation, Pearson's correlation, Spearman's correlation, and order correlation) and three of the most popular regression methods (regression, liner regression, and logistic regression), also included one of the parametric inferential methods (Kolmogorov–Smirnov test). The connections between the principle correlation methods and regression methods corresponded to their mathematical classification, and the appearances of one inferential method revealed that this test was usually applied with the predictive methods instead of other inferential methods. The fourth cluster indicated that the prevalent nonparametric inferential method (Chi-square test) was frequently applied in the same areas with several regression methods and parametric methods. The most widely used nonparametric inferential method (Chi-square test) was clustered with two parametric methods (*z*-test and ANCOVA) and three regression methods (multiple regression, hierarchical regression, and Cox regression) together in this fourth cluster.

The high interactions between the detailed statistical methods and the application areas are represented by the dark red cells in Figure 12. Research in the IC area focused on the adaptation of the three methods: multiple regression, linear regression, and regression. The concentrations of these three regression methods confirmed the strong interactions between IC and regression methods as shown in Figure 11. IO research mainly applied the McNemar test. For instance, Névél, Rogozan and Darmoni (2006) applied the McNemar test to the automatic indexing research. Regarding the IR area, all the methods belonging to the first and second cluster were frequently applied in IR research. Compared with the other five areas, the IR area widely adopted most of the statistical methods. As shown in Figure 11, the IR area mainly used the 2 types of inferential methods, but the detailed matrix (Figure 12) also showed that several correlation methods (e.g., order correlation, Pearson's correlation, and Spearman's correlation) and regression methods (e.g., linear regression, hierarchical regression, and multiple regression) were still fre-

quently adopted in IR research. For example, researchers used the Pearson's correlation in the natural language processing research (Bollegala, Goto, Duc, & Ishizuka, 2013) and employed the Spearman's correlation in the document summarization research (Ouyang, Li, Zhang, Li, & Lu, 2013). The IU area tended to apply methods in the fourth cluster frequently. Interactions between the IU area and the detailed methods appeared with Cox regression, ANCOVA, z-test, hierarchical regression, multiple regression, and Chi-square test. Most of these methods came from the regression category. This differed from the interactions between IU and regression methods as shown in Figure 11 in which outcomes were less obvious compared with other types of methods. In the ISC field, the investigated papers tended to commonly adopt the following methods: Spearman's correlation, logistic regression, Pearson's correlation, and regression. All of them can be categorized as regression methods or correlation methods. The highly interactive regression cells illustrated that some regression methods were still adopted for the ISC research although the papers in that field did not usually apply other regression methods as shown in Figure 11. As shown in Figure 12, ID research was strongly related to order correlation and hierarchical regression in contrast to other methods. In the ID area, for example, hierarchical regression was used in the communication behavior study (Robbin & Buente, 2008) and the interpersonal knowledge transfer study (Kang & Kim, 2010).

Figure 13 shows the four types of statistical methods in the 45 expanded application areas. The X-axis represents the four types of statistical methods, and the Y-axis represents the 45 research areas discussed in the investigated papers. Among all the detailed areas, parametric inferential methods were widely employed in many research areas and especially in the topics belonging to the IO and IR application areas. This corresponded to the apparent interactions between the parametric inferential statistical methods and IO as well as IR as shown in Figure 11.

Another type of inferential statistical methods, the nonparametric inferential statistical methods, was most frequently applied in the following areas: cover design, metadata, learning and training performance, search behavior, multimedia retrieval, and publication choice. Moreover, the nonparametric inferential statistical methods were also the preferred methods in those areas. Although nonparametric inferential statistical methods did not highly interact with ID, these types of methods were commonly used in the expanded area studies of ID: publication choice and social media. It suggested that even within one research area different research topics held diverse preferences on statistical methods used.

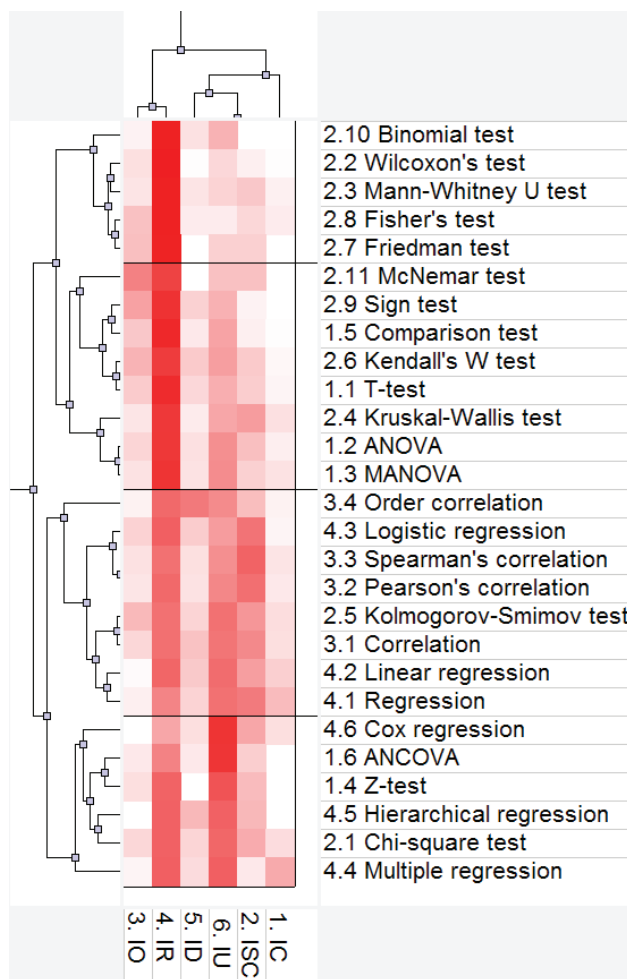


Figure 12: Matrix visualization of the interaction between the application areas and the detailed statistical methods.

Correlation methods were prevalent in the following expanded areas: researcher and institution evaluation, publication evaluation, research pattern, citation behavior, information evaluation indicator, and information source selection. Yet, the majority of the weak interactions appeared in the expanded areas within IO and IR. Correlation methods were frequently adopted by IO's expanded area of categorization and IR's expanded area of information literacy research.

Regression methods were rarely utilized in any of the application areas except for job satisfaction and development studies and research productivity. Furthermore, although different forms of the other three types of statistical methods dominated in several expanded application areas, the regression methods were predominant only in the abovementioned two areas.

A total of 45 expanded application areas were partitioned into six clusters grouped by the use of statistical

methods. The first cluster consisted of two detailed areas from IC and IO. The interactions as shown in Figure 13 indicated that these two expanded areas commonly used non-parametric and parametric inferential methods as their statistical methods. Then, 10 expanded application areas were attributed to each of the second and third cluster. Most of these 20 areas were from IR, IO, and IU. The fourth cluster not only represented IU research but also included eight expanded application areas from ISC, ID, and IO. The fourth cluster appeared to be the largest cluster and included 13 expanded areas stemming from the ISC, IO, ID, and IU application areas. In the fourth cluster, parametric and nonparametric inferential methods and correlation methods were prevalent. The fifth cluster included four expanded areas that belonged to the IC, ID, and IU application areas. All the four types of statistical methods were almost evenly applied in the fifth cluster. Correlation methods clearly predominated in the sixth cluster. The research pattern and citation behavior research were originally assigned to IC and IU, but this topic was later grouped with other ISC expanded application areas with similar utilizations of statistical methods.

Figure 14 illustrates the interaction between the expanded application areas and the detailed statistical methods. The X-axis represents 27 specific statistical methods, and the Y-axis represents 45 expanded application areas identified in the investigated papers. Not surprising, the intersections were much sparser compared with the previous figures. On the one hand, the white areas suggested that most of the expanded application areas employed only a few statistical methods in their studies. For example, among the investigated papers, the learning and training performance research utilized only three statistical methods: Chi-square test, *t*-test, and correlation. On the other hand, the white spaces as shown in Figure 14 also demonstrate that most of the statistical methods were not explored by all the areas but commonly were applied to some specific topics. For instance, binomial test was solely used in information system and performance evaluation, search behavior studies, information literacy, and citation behavior.

The cells indicated in bold as shown in Figure 14 represent the comparatively frequent use of a specific method in an application area. For example, Wilcoxon's test was highly utilized to deal with the text summarization research (Kishida, 2008). The research pattern research employed the correlation method to detect the relationship between citation errors and library anxiety (Jiao, Onwuegbuzie, & Waytowich, 2008). Regarding publication evaluation research, regression was used to evaluate publications in biology and psychology (Tang, R. &

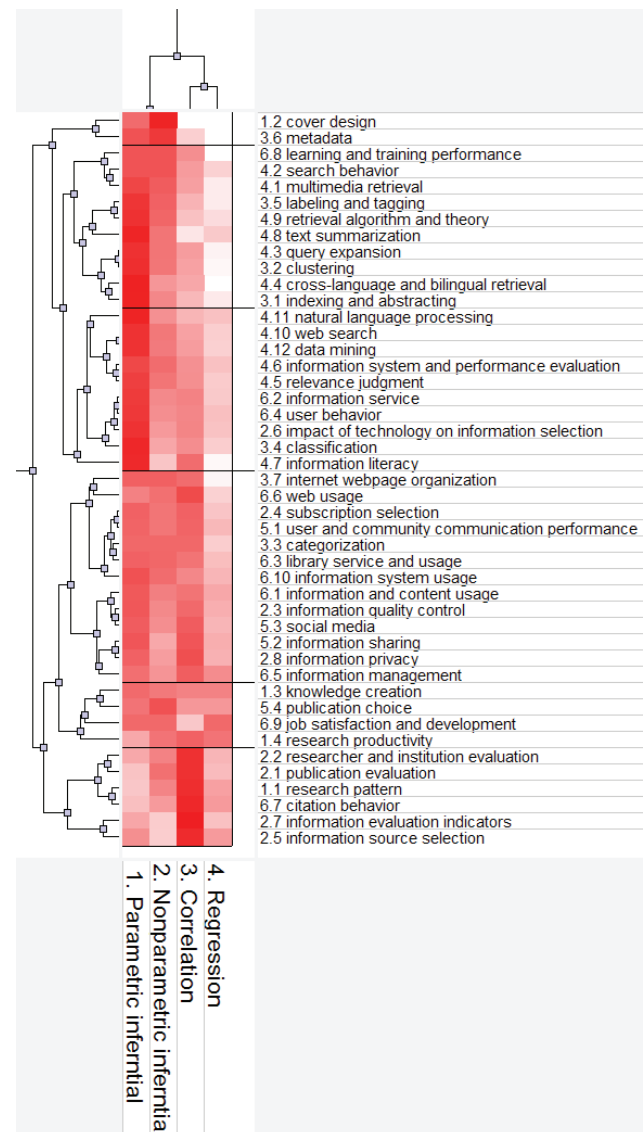


Figure 13: Matrix visualization of the interaction between the expanded application areas and the types of statistical methods.

Safer, M. A., 2008) and publications in biomedicine (Bornmann & Daniel, 2007). The search behavior research, as the most popular expanded application area, included 203 papers that used several statistical methods. The following methods were utilized by this topic: MANOVA assisted the user characteristic assessment studies (Zhang & Chignell, 2001); Mann–Whitney U test helped to explore the biomedical information search behavior (Vanopstal, Stichele, Laureys, & Buysschaert, 2012); the binomial test was used to investigate end users' search behaviors in Medline database (Sutcliffe, Ennis, & Watkinson, 2000); and the Fisher's test was employed in search task research (Li, 2009).

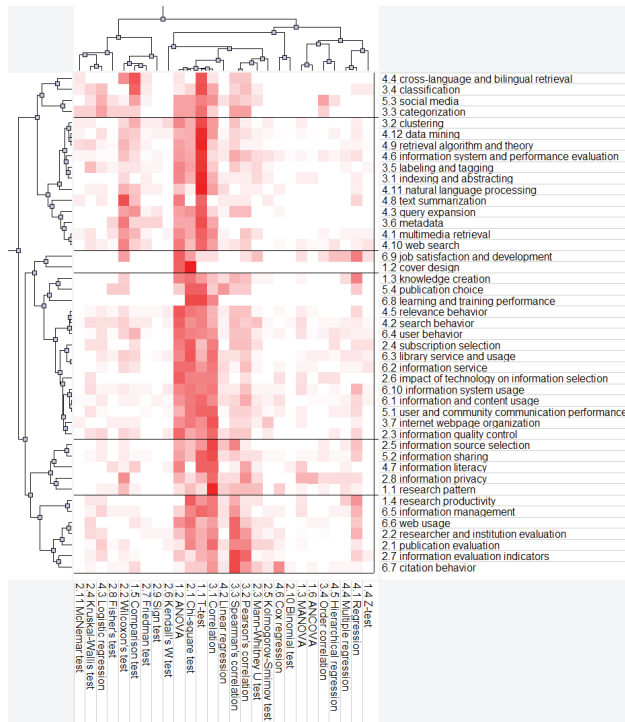


Figure 14: Matrix visualization of the interaction between the expanded application areas and the detailed statistical methods.

Based on the employment of 27 specific statistical methods, 45 specific application areas were grouped into six clusters. Areas in the second and fourth cluster widely used several methods, whereas areas in other clusters adopted several more specific methods. Moreover, the second and fourth cluster contained more areas than the other clusters. The red blocks in the second, fourth, and fifth cluster were spread across the detailed statistical methods, and the bold indications for these blocks were concentrated in the following four specific statistical methods: ANOVA, Chi-square test, *t*-test, and correlation. It suggested that these areas can be analyzed by various methods. The expanded application areas in the first and second cluster frequently adopted *t*-test in their studies. Correlation was widely used by five expanded application areas in the fifth cluster. It can be seen that the most popular tests such as ANOVA and *t*-test were generally utilized by most areas.

One of the research trends in the LIS field is that more and more studies have applied statistical methods and mixed research methods to domain problems. For instance, Enger (2006) revealed that descriptive statistical methods and inferential statistical methods were frequently used in the LIS field. Statistical methods are used to accept or reject proposed hypotheses in studies, and they also indicate the degree to which the hypotheses are

accepted and rejected. In general, quantitative research methods and qualitative research methods have different paradigms and characteristics in terms of research design and implementation. Each has its strength and weakness. Quantitative research methods fit confirmatory studies while qualitative research methods fit exploratory studies. Statistical methods such as *t*-tests and ANOVA tests are widely employed in several evaluation-related studies in the field such as information retrieval system evaluations, user behavior evaluations, program effectiveness evaluations, project evaluations, and so on. Furthermore, statistical methods such as regression analysis methods are powerful and effective means to solve complicated problems in emerging research areas in the field such as big data and social media.

The findings of this study will help researchers in the field not only understand the applications of statistical methods in the specific areas of the field but also identify proper statistical methods for their own studies.

5 Conclusion

Statistical methods are widely utilized in several quantitative studies. They are used for analyzing data and drawing an inferential conclusion from the collected data. Statistical methods are often used as a foundation for research methodology to communicate research design, explain research findings, test hypotheses, and give the degree to which research results are reliable. It is exceedingly important for researchers and also consumers of research (such as educators, students, and practitioners in library and information science) to understand how statistical methods are used and to which research areas statistical methods are applied.

In this study, six major scholarly journals (*The Journal of the Association for Information Science and Technology*, *Information Processing and Management*, *The Library Quarterly*, *The Journal of Information Science*, *Library and Information Science Research*, and *The Journal of Documentation*) were investigated. Research papers in these journals published from 1999 to 2017 were identified, analyzed, and examined using a Chi-square test and the graphic clustering visualization method to determine the following: what are the research areas in library and information science to which statistical methods are applied; what are the statistical methods used in the field; and what are the interactions between statistical methods and application areas. The number of investigated papers that employed statistical methods was 1821. Four types of statistical methods (parametric inferential statistical

methods, nonparametric inferential statistical methods, predictive statistical correlation methods, and predictive statistical regression methods) and six application areas (information creation, information selection and control, information organization, information retrieval, information dissemination, and information use) were defined. The four types of statistical methods were expanded to 27 statistical approaches, and the six application areas were expanded to 45 expanded application areas. A total of 27 different statistical methods were applied to 45 expanded application areas.

Findings of this study showed that there was a significant relationship between the defined statistical method and the application area of studies in library and information science. Parametric inferential methods were most used in the application areas of search behavior, data mining, information system and performance evaluation, and user behavior. Nonparametric inferential methods were most applied in the research of search behavior, information system and performance evaluation, and data mining. Correlation methods were most employed by the studies of researcher and institution evaluation, search behavior, publication evaluation, and information evaluation indicators. Regression methods were most used in search behavior research. Studies in information organization and information retrieval tended to utilize parametric and nonparametric inferential methods, while correlation and regression methods were employed by studies in information use, information dissemination, information creation, and information selection and control field. Based on the use of the methods in the application areas, parametric inferential methods and nonparametric inferential methods were grouped together and correlation methods and regression methods were grouped together. This complied with the inferential statistics and predictive statistics classification.

The findings of the study provide a detailed picture of statistical method uses in the contexts of research topics and application areas and offer a better understanding of library and information science research. They assist researchers in understanding a research problem in a specific research area and potential statistical method(s) to solve the problem. Therefore, they help the researchers make an appropriate decision on statistical method selection for their research problems. The findings of this study can also be used to aid educators who teach quantitative research methods and statistical methods in library and information science to develop and design syllabi which include popular statistical methods, introduce examples for each of the statistical methods, identify important and frequently used statistical methods for research

topic in the field, and understand the interaction between research areas and statistical methods. These are critical in order for students to master statistical methods.

The limitation of this study is that only the numbers of statistical methods used in different application areas were collected and compared. The percentage of papers using statistical methods in each application area to all the papers in the corresponding area was not calculated. Since the total number of papers in a specific area might affect the number of papers using statistical methods in the area, the percentage can reflect the importance of statistical methods in the area more than the number of papers. The future study will collect the percentages of papers using statistical methods in the identified application areas and analyze the importance of statistical methods in each area.

Future research directions on this study include, but are not limited to, increasing the number of the investigated journals in library and information science to gain a larger picture of the application of statistical methods in the field and conducting a temporal analysis on the interaction between statistical methods and their application areas.

References

- Atherton, P. (1975). Research in Information Science: An Assessment. In *Perspectives in Information Science* (pp. 665–683). Dordrecht: Springer. https://doi.org/10.1007/978-94-011-7759-7_43
- Berg, B. L., & Lune, H. (2011). *Qualitative research methods for the social sciences* (8th ed.). Boston: Pearson.
- Bernhard, P. (1993). In search of research methods used in information science. *Canadian Journal of Information and Library Science*, 18(3), 1–35.
- Blake, V. L. P. (1994). Since Shaughnessy. *Collection Management*, 19(1–2), 1–42.
- Bollegala, D., Goto, T., Duc, N. T., & Ishizuka, M. (2013). Improving relational similarity measurement using symmetries in proportional word analogies. *Information Processing & Management*, 49(1), 355–369. <https://doi.org/10.1016/j.ipm.2012.05.007>
- Bornmann, L., & Daniel, H.-D. (2007). Multiple publication on a single research study: Does it pay? The influence of number of research articles on total citation counts in biomedicine. *Journal of the American Society for Information Science and Technology*, 58(8), 1100–1107. <https://doi.org/10.1002/asi.20531>
- Chu, H. (2015). Research methods in library and information science: A content analysis. *Library & Information Science Research*, 37(1), 36–41. <https://doi.org/10.1016/j.lisr.2014.09.003>
- Connaway, L. S., & Powell, R. R. (2010). *Basic research methods for librarians* (5th ed.). ABC-CLIO.

- Dilevko, J. (2007). Inferential statistics and librarianship. *Library & Information Science Research*, 29(2), 209–229. <https://doi.org/10.1016/j.lisr.2007.04.003>
- Dimitroff, A. (1992). Research in health sciences library and information science: A quantitative analysis. *Bulletin of the Medical Library Association*, 80(4), 340–346.
- Doucette, L. (2017). Quantitative Methods and Inferential Statistics: Capacity and Development for Librarians. *Evidence Based Library and Information Practice*, 12(2), 53–58. <https://doi.org/10.18438/B82940>
- Eldredge, J. D. (2004). Inventory of research methods for librarianship and informatics. *Journal of the Medical Library Association: JMLA*, 92(1), 83–90.
- Enger, K. (2006). Understanding the Development of Disciplines and the Ways they Contribute to Knowledge and Reflect Practice: An Analysis of Articles Published in Higher Education and Library and Information Science. In *Advances in Library Administration and Organization* (Vol. 24, pp. 1–51). Emerald Group Publishing Limited. [https://doi.org/10.1016/S0732-0671\(06\)24001-X](https://doi.org/10.1016/S0732-0671(06)24001-X)
- Enger, K., Quirk, G., & Stewart, J. A. (1989). *Statistical Methods Used by Authors of Library and Information Science Journal Articles*, by Enger KB, Quirk G, Stewart JA (SSRN Scholarly Paper No. ID 1365280). Rochester, NY: Social Science Research Network.
- Fisher, S. R. A. (1936). *Statistical methods for research workers*. Oliver & Boyd. gCluto (Version 1.0). Minneapolis, MN: Karypis Lab.
- Gravetter, F.J. & Wallnau, L.B. (2009). *Statistics for the behavioral sciences*. Wadsworth Cengage Learning. Gravetter, F., & Wallnau, L. (2013). *Essentials of statistics for the behavioral sciences*. Cengage Learning.
- Hersberger, J. A. (2009). The Current State of Public Library Research in Select Peer-Reviewed Journals: 1996–2000. *North Carolina Libraries*, 59(1), 10.
- Hider, P., & Pymm, B. (2008). Empirical research methods reported in high-profile LIS journal literature. *Library & Information Science Research*, 30(2), 108–114. <https://doi.org/10.1016/j.lisr.2007.11.007>
- Hodge, G. M. (2000). Best practices for digital archiving: An information life cycle approach. *D-Lib Magazine: the Magazine of the Digital Library Forum*, 6(1). <https://doi.org/10.1045/january2000-hodge>
- Järvelin, K., & Vakkari, P. (1990). Content analysis of research articles in library and information science. *Library & Information Science Research*, 12(4), 395–421.
- Järvelin, K., & Vakkari, P. (1993). The evolution of library and information science 1965–1985: A content analysis of journal articles. *Information Processing & Management*, 29(1), 129–144. [https://doi.org/10.1016/0306-4573\(93\)90028-C](https://doi.org/10.1016/0306-4573(93)90028-C)
- Jiao, Q. G., Onwuegbuzie, A. J., & Waytowich, V. L. (2008). The relationship between citation errors and library anxiety: An empirical study of doctoral students in education. *Information Processing & Management*, 44(2), 948–956. <https://doi.org/10.1016/j.ipm.2007.05.007>
- Johnson, R. A. (2009). *Statistics: Principles and methods*. John Wiley & Sons.
- Kang, M., & Kim, Y.-G. (2010). A multilevel view on interpersonal knowledge transfer. *Journal of the American Society for Information Science and Technology*, 61(3), 483–494.
- Kishida, K. (2008). Prediction of performance of cross-language information retrieval using automatic evaluation of translation. *Library & Information Science Research*, 30(2), 138–144. <https://doi.org/10.1016/j.lisr.2007.09.003>
- Kothari, C. R. (2011). *Research methodology: Methods and techniques*. New Age International.
- Kumpulainen, S. (2009). Library and Information Science Research in 1975: Content Analysis of the Journal Articles. *Libri*, 41(1), 59–76.
- Li, Y. (2009). Exploring the relationships between work task and search task in information search. *Journal of the American Society for Information Science and Technology*, 60(2), 275–291. <https://doi.org/10.1002/asi.20977>
- Munro, B. H. (2005). *Statistical methods for health care research*. Lippincott Williams & Wilkins.
- Névéol, A., Rogozan, A., & Darmoni, S. (2006). Automatic indexing of online health resources for a French quality controlled gateway. *Information Processing & Management*, 42(3), 695–709. <https://doi.org/10.1016/j.ipm.2005.01.003>
- Ott, R., & Longnecker, M. (2008). *An introduction to statistical methods and data analysis*. Cengage Learning.
- Ouyang, Y., Li, W., Zhang, R., Li, S., & Lu, Q. (2013). A progressive sentence selection strategy for document summarization. *Information Processing & Management*, 49(1), 213–221. <https://doi.org/10.1016/j.ipm.2012.05.002>
- Peritz, B. C. (1980). The Methods of Library Science Research: Some Results from a Bibliometric Survey. *Library Research*, 2(3), 251–268.
- Rajasekar, S., Philominathan, P., & Chinnathambi, V. (2006). Research Methodology. arXiv:physics/0601009. Retrieved from <http://arxiv.org/abs/physics/0601009>
- Rasmussen, M., & Karypis, G. (2004). *gCLUTO - An interactive clustering, visualization, and analysis system*. Retrieved from <http://glaros.dtc.umn.edu/gkhome/node/174>
- Robbin, A., & Buente, W. (2008). Internet information and communication behavior during a political moment: The Iraq war, March 2003. *Journal of the American Society for Information Science and Technology*, 59(14), 2210–2231. <https://doi.org/10.1002/asi.20925>
- Sutcliffe, A. G., Ennis, M., & Watkinson, S. J. (2000). Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51(13), 1211–1231. [https://doi.org/10.1002/1097-4571\(2000\)9999:99993.0.CO;2-5](https://doi.org/10.1002/1097-4571(2000)9999:99993.0.CO;2-5)
- Tang, R., & Safer, M. A. (2008). Author-rated importance of cited references in biology and psychology publications. *The Journal of Documentation*, 64(2), 246–272. <https://doi.org/10.1108/00220410810858047>
- Togia, A., & Malliari, A. (2017). Research Methods in Library and Information Science. <https://doi.org/10.5772/intechopen.68749>
- Tuomaala, O., Järvelin, K., & Vakkari, P. (2014). Evolution of library and information science, 1965–2005: Content analysis of journal articles. *Journal of the Association for Information Science and Technology*, 65(7), 1446–1462. <https://doi.org/10.1002/asi.23034>
- Van De Water, N., Surprenant, N., Genova, B. K. L., & Atherton, P. (1976). Research in information science: An assessment. *Information Processing & Management*, 12(2), 117–123. [https://doi.org/10.1016/0306-4573\(76\)90042-X](https://doi.org/10.1016/0306-4573(76)90042-X)
- Van Epps, A. (2012). *Librarians and Statistics: Thoughts on a Tentative Relationship*. *Practical Academic Librarianship*. The International Journal of the SLA Academic Division.

- Vanopstal, K., Stichele, R. V., Laureys, G., & Buysschaert, J. (2012). PubMed searches by Dutch-speaking nursing students: The impact of language and system experience. *Journal of the American Society for Information Science and Technology*, 63(8), 1538–1552. <https://doi.org/10.1002/asi.22694>
- Vaughan, L. (2001). *Statistical methods for the information professional: A practical, painless approach to understanding, using, and interpreting statistics* (1st ed.). Medford, N.J: Information Today Inc.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360–363.
- Wallace, D. P. (1985). The use of statistical methods in library and information science. *Journal of the American Society for Information Science*, 36(6), 402–410. <https://doi.org/10.1002/asi.4630360610>
- Westbrook, L. (1994). Qualitative research methods: A review of major stages, data analysis techniques, and quality controls. *Library & Information Science Research*, 16(3), 241–254. [https://doi.org/10.1016/0740-8188\(94\)90026-4](https://doi.org/10.1016/0740-8188(94)90026-4)
- Weisburd, D., & Britt, C. (2007). *Statistics in criminal justice* (3rd ed.). College Park, MD: Springer.
- Williams, J. F., II, & Winston, M. D. (2003). Leadership competencies and the importance of research methods and statistical analysis in decision making and research and publication: A study of citation patterns. *Library & Information Science Research*, 25(4), 387–402. [https://doi.org/10.1016/S0740-8188\(03\)00050-1](https://doi.org/10.1016/S0740-8188(03)00050-1)
- Wyllys, R. E. (1978). Teaching Descriptive and Inferential Statistics in Library Schools. *Journal of Education for Librarianship*, 19(1), 3–20. <https://doi.org/10.2307/40368591>
- Zhang, J., Wang, Y., & Zhao, Y. (2017). Investigation on the statistical methods in research studies of library and information science. *The Electronic Library*, 35(6), 1070–1086. <https://doi.org/10.1108/EL-02-2016-0042>
- Zhang, J., Zhao, Y., & Wang, Y. (2016). A study on statistical methods used in six journals of library and information science. *Online Information Review*, 40(3), 416–434. <https://doi.org/10.1108/OIR-07-2015-0247>
- Zhang, X., & Chignell, M. (2001). Assessment of the effects of user characteristics on mental models of information retrieval systems. *Journal of the Association for Information Science and Technology*, 52(6), 445–459. [https://doi.org/10.1002/1532-2890\(2001\)9999:99993.0.CO;2-3](https://doi.org/10.1002/1532-2890(2001)9999:99993.0.CO;2-3)