Chapter III

RESEARCH DESIGN

This chapter deals with the hypothesis, methodological description, nature of sampling and method of data collection, description and relevance of statistical tools applied and limitations.

Hypotheses

The following hypotheses are formulated, keeping the content and coverage of the framed objectives.

1. There is a significant increase in the growth of literature during the period of study.

2. There is a significant intercontinental and inter country variation in the growth and concentration of research output in Gender Studies in Informatics.

3. The cognitive structure of Gender Studies in Informatics varies significantly among the selected continents in the study.

4. There is a significant variation with respect to institutions, producing research output in Gender Studies in Informatics.
Methodology

This study aims at analyzing the trends in the development of Gender Studies in Informatics. This study traces out the past trends in the area of research in gender studies on the bases of cross section data. This study evaluates the contribution of countries to the growth pattern and development of research on this discipline during the last two decades. The aim is to build a comprehensive data base on women in informatics and to map the status of women in this discipline. This study in undertaken in two dimension. The first dimension relates to statistical analysis with respect to growth of literature, area wise research output, institution wise output source wise research output, country wise research output, authorship pattern, collaboration in research and so on. The second dimension relates to classification of documents into clusters and building scatter diagrams and maps. This study is partly exploratory and partly evaluatory in nature.

Sampling

Among the disciplines, the researcher has selected Gender Studies in Informatics. Many studies have highlighted gender issues in various dimensions and none of the studies concentrate on women's involvement in I.T.

Data collection

The relevant data are obtained from various sources such as journal articles, conference proceedings, reports, projects, online discussion news items
and so on. The relevant data are obtained from the abstracts and relevant data are downloaded from the online resources.

**Statistical Tools**

(i) Co-efficient of Variation

In order to study the inequality in distribution of literature, the Lorenz curve model is applied. Further inequality in the distribution of literature is studied with the help of coefficient variation model.

It is written as,

$$C. V = \frac{\sigma}{\chi'} \times 100$$

Where,

- $C.V$ = Coefficient variation
- $X'$ = Mean
- $\sigma$ = Standard deviation

**ANOVA Two Way Model**

In order to study the institution wise variation and subfield wise variation, the ANOVA two way model is applied. It is written as,

<table>
<thead>
<tr>
<th>Sources of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Sum of Squares</th>
<th>Ratio of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Samples</td>
<td>SSC (c-1)</td>
<td>MSC=SSC/(c-1)</td>
<td>MSC/MSE</td>
<td></td>
</tr>
<tr>
<td>Between Rows</td>
<td>SSR (r-1)</td>
<td>MSR=SSR/(r-1)</td>
<td>MSR/MSE</td>
<td></td>
</tr>
<tr>
<td>Residual or Error</td>
<td>SSE (c-1)(r-1)</td>
<td>MSE=SSE/(r-1)(c-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SST n-1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SSC = Sum of Squares between Columns
SSR = Sum of Squares between Rows
SSE = Sum of Squares due to Error
SST = Total Sum of Squares

The sum of squares for the source 'Residual' was obtained by subtracting from the total sum of squares, the sum of squares between columns and rows, i.e.,

$$\text{SSE} = \text{SST} - (\text{SSC} + \text{SSR})$$

The total number of degrees of freedom = n -1 or r -1

Where c refers to number of columns, and r refers to number of rows,

Number of degrees of freedom between columns = (c -1)
Number of degrees of freedom between rows = (r -1)
Number of degrees of freedom for residual = (c -1) (r -1)

The total sum of squares, sum of squares between columns and sum of squares between rows were obtained in the same way as before.

Residual or error sum of square = Total sum of squares - sum of squares between columns - sum of squares between rows.

The F values were calculated as follows:

$$F(v_1,v_2) = \frac{\text{MSC}}{\text{MSE}}$$

Where,

$$v_1 = (c -1) \text{ and } v_2 = (c -1) (r -1)$$
MSR
F(V_1V_2) = \frac{---}{MSE}

Where,
V_1 = (r-l) and V_2 = (c-l)(r-l)

It is noted that the F value may not be the same in both cases, viz., V_1 = (c-l) and V_1 = (r-l).

The calculated values of F were compared with the table values and logical conclusions were drawn.

It would be clear from above that in problems involving two-way classification, 'Residual' is the measuring rod for testing significance. It represents the magnitude of variation due to forces called 'chance'.

**Relative Growth Rate**

The relative growth rate is the increase in the number of publications/pages per unit of time. Here, one year is taken as the unit of time. The mean relative growth rate \( R'(1-2) \) over a specified period of interval can be calculated from the following equation suggested by Mahapatra.

\[
R'(1-2) = \frac{W_2 - W_1}{T_2 - T_1}
\]

Where,
\( R' \) = mean relative growth rate over the specific period of interval;
\[ W_1 = \log W_1 \text{ (Natural log of initial number of publications/pages);} \]
\[ W_2 = \log W_2 \text{ (Natural log of initial number of publications/pages);} \]
\[ T_2 - T_1 = \text{The unit difference between the initial time and final time.} \]

Therefore,
\[ R'(a) = \text{Relative growth rate per unit of publications per unit of time (year)} \]
\[ R'(P) = \text{Relative growth rate per unit of pages per unit of time (year)} \]

Doubling Time

A direct equivalence exists between the relative growth rate and doubling time. If the number of publications/pages of a subject doubles during a given period, then the difference between the logarithms of the numbers at the beginning and at the end of the period must be the logarithms of the number. This difference has a value of 0.693. Thus, the corresponding doubling time for publication and pages can be calculated by the following formula:

\[ \frac{0.693}{R'} \]

Therefore,
\[ \frac{0.693}{R'(a)} \]

Doubling time for publications \( D_t(a) \) =


**Activity Index**

The proportion of the contribution of a country is seen from two different angles even though both analyze the productivity of a country in terms of growth. i.e., (i) volume of papers produced by a country in different years and thus the extent of increase; and (ii) volume of growth in different years in relation to the world productivity.

To gauge the productivity, the performance index is applied in the study.

\[
PI = \left\{ \frac{CAO}{AOC} \right\} \times 100
\]

**PI** = Performance Index  
**CAO** = Country's Average Output  
**AOC** = Average Output of Continent  
**Yr** = Total years of the study period

The performance index is used to measure the productivity of a country in relation to the continent output. The activity index of a country is measured by:

A similar performance index is constructed to analyze the relative position of the country at the world level. The performance index is constructed as follows:
\[
\text{PI} = \{\frac{\text{CAO/yr}}{\text{WAO/yr}}\} \times 100
\]

- **PI** = Performance Index
- **CAO** = Country's Average Output
- **WAO** = World Average Output
- **Yr** = Total years of the study period

**Coword Method**

Coword method is one of the tools of scientometrics. Articles from journals related to the study area collected for last two decades. Keywords used to describe the contents of an article which are the basic building blocks of a research field structure. A cluster of keywords can be understood as a short description of a research theme. A research field is then described as a structure of mutually connected research themes.

A research theme can be identified by using the information about common occurrences of keywords in some articles. Let's calculate an association index as:

\[
e_{ij} = \frac{f_{ij}^2}{(f_i \cdot f_j)}
\]
Where, \( f_{ij} \) is the number of common presence of keywords \( i \) and \( j \) in an article, \( f_i \) is the total number of occurrences of the world \( i \) in all the articles can understand an association index as a measure of strength of ties between, keywords in a research field. This measure is then used for clustering the key words into research themes.

Each research theme obtained in this process has two parameters. The first one, called density measures the strength of internal ties among all keywords describing the research theme. We can understand this parameter as a measure of the theme development. The second one, called centrality, measures the strength of external ties to other themes. We can understand this parameter as a measure of importance of a theme in the development of the entire analyzed field.

Both median and mean values for density and centrality can be used in classifying themes into four groups. Thereafter, a research field can be understood as a set of research themes, mapped in a strategic diagram - graph made by plotting themes according to their centrality and density rank values or values along two axes, x- are centality, y-axe density. Strategic diagrams with rank values are used more commonly than the ones with values, because of their legibility.

Themes in the first quadrant are both well developed and important for the structuring of a research field. Themes in the fourth quadrant have well
developed internal ties but unimportant external ties and so are only of a marginal importance for the field. Themes in the third quadrant are both weekly developed and marginal themes in the second quadrant are important for a research field but are not developed. This study use the clock wise convention of quadrant numbering and we hope it won't cause any misunderstandings later in the text.

In this paper, a strategic diagram and its changes over periods of time are the main tools for the analysis of dynamics of the artificial neural networks field.

**Degree of Collaboration**

The degree of collaboration is defined as the ratio of the number of collaborative research papers to the total number of research papers in the discipline during a certain period of time. The formula suggested by Subramanyam is used. It is expressed as:

\[
C = \frac{Nm}{Nm + Ns}
\]

Where, \( C \) is the degree of collaboration in a discipline. \( Nm \) is the number of multi-authored research papers in the discipline published during a year. \( Ns \) is the number of single authored papers in the discipline published during the same year. Using this formula the degree of collaboration is determined.
Concepts

Scientometries

Quantitative analysis of bibliographic data of scientific publications with the objective to find characteristics of research performance.

Mapping Techniques

Mappings are constructed to see the growth of research output in the area of Gender Studies in Informatics.

Gender studies

It explains the research studies conducted throughout the globe pertaining to women's education in information technology, and income and employment generation through Information Technology.

Cognitive structure

It explains the development of area wise research output in Gender Studies in Informatics, institution wise research performance and co classification analysis of subjects.

Limitations

The findings of this apply only to Gender Studies in Informatics and it does not apply to other fields. This study covers gender issues only with respect to women's I.T. Education and income and employment generation through information technology and other aspects of women development programmer.
are not brought under the purview of the study, because studying of all aspects of women development is not possible by the level of an individual researcher, due to constraints imposed by money, time, energy and efforts. This study does not take into account citations and their impacts, key word analysis, since it is a new area of research. This study is pursued with certain degree of caution.