MEDLINE®

Life Sciences Usages and Impacts
In
2015-2016

About MEDLINE®

Medline is the US National Library of Medicine (NLM)database that contains more than 25 million references to journal in life sciences. The journal in Medline are selected based of on the recommendation the Literature Selection Technical Review committee and initiated review by NLM. The review takes the form of consultation with internal and external experts.

This study will focus on the knowledge discovery in the database and examine the usages and Impacts of Medline database between 2015 to 2016

- The data has 19 headers and 31287 rows (including books, book series, conference proceedings and journals) spanning 219 subject areas and 237 sub-topics from over 500 institutions in 2596 journal titles (i.e. publishers). The data has both categorical variables (e.g. type of publication,journal title etc.), and quantitative variable (e.g. Number of views, Number of citations the article received, Journal normalized impact indicator etc.)
- The data is normally distributed for normalised impact factor and racking (i.e. SJR 2017) and the data size is large for analysis of statistical estimate for further evaluation of the research questions.
- file:///C:\Users\Jimmmygreit\OneDrive\Classnote\609%20C\Project\Medline.xlsx

Statistics Summary

SNIP 2017	
Mean	1.34
Standard Error	0.01
Median	1.12
Mode	1.11
Standard Deviation	0.94
Sample Variance	0.89
Skewness	4.43
Range	14.80
Count	30,931.00
Confidence Level(95.0%)	0.01

SNIP 2017 stands for Journal normalized impact indicator

SJR 2017	
Mean	2.33
Standard Error	0.02
Median	1.46
Mode	1.16
Standard Deviation	2.89
Sample Variance	8.34
Skewness	4.52
Range	34.90
Count	30,931.00
Confidence Level(95.0%)	0.03

SJR 2017 stands for Journal rank

Statistics Summary

Views	
Mean	19.94
Standard Error	0.15
Median	14.00
Mode	6.00
Standard Deviation	26.30
Sample Variance	691.46
Skewness	14.24
Range	1,341.00
Count	31,286.00
Confidence Level(95.0%)	0.29

Views stands for Number of views to an article

Citations	
Mean	11.12
Standard Error	0.13
Median	6.00
Mode	2.00
Standard Deviation	22.82
Sample Variance	520.87
Skewness	23.35
Range	1,701.00
Count	31,286.00
Confidence Level(95.0%)	0.25

Citations stands for number of citations an article received

Possible questions that could be addressed:

a.) Is there relationship between document type and the citation?
b.) Is there any relationship between document type and journal impact (SNIP)?
c.) Which field of life sciences are productive?
d.) Which Journal title are productive?
e.) Is there relationship between view counts and citation counts to research documents?
f.) Are papers published in higher impact journals, more cited?

Is there relationship between document type and the citation?

Variables:

The **explanatory variable** in this question is **document type** and **response variable** is **citation**.

Role type classification:

Categorical explanatory and quantitative response

Document type:

Article, Review, Note
Conference Paper
Chapter, Erratum
Letter, Short Survey
Editorial, Article in Press, &
Tombstone

Hypothesis:

Document type has mean of citation

Ho:μ=11.12

Ha:μ≠11.12

Statistical method:

Boxplot, supplemented with descriptive statistics(T-test will be calculated for each document type)

The result will help to know which document type is being cited .

Is there any relationship between document type and journal impact (SNIP)?

Variables:

The **explanatory variable** in this question is **document type** and **response variable** is SNIP 2017.

Role type classification:

Categorical explanatory and quantitative response.

Document type:

Article, Review, Note
Conference Paper
Chapter, Erratum
Letter, Short Survey
Editorial, Article in Press, &
Tombstone

Hypothesis:

Document type has mean of SNIP 2017?

Ho:μ=1.34

Ha:μ≠1.34

Statistical method:

Boxplot, supplemented with descriptive statistics(T-test will be calculated for each document type).

This will help us to know which document type has usage

Which field of life sciences are productive?

Variables:

The **explanatory variable** in this question is **life sciences** and the **response variable** is SNIP 2017.

Role type classification:

Categorical explanatory and quantitative response

Life sciences Field :

2700 - General Medicine
1306 - Cancer Research
1303 - Biochemistry
2723 - Immunology and Allergy
1304 - Biophysics
2734 - Pathology and Forensic Medicine
2720 - Hematology
2722 - Histology
1312 - Molecular Biology
1703 - Computational Theory and Mathematics

Life Sciences has 219 fields

Objectives:

Identify the fields in life sciences within 75 percentile above and 50 percentile with the response variable.

The field with 50 percentile and above can be described as productive.

Statistical method:

The descriptive statistics will be used to rank the 219 fields in life sciences to determine the productive field.

Which Journal title are productive?

Variables:

The **explanatory variable** in this question is **journal title** and the **response variable** is SNIP 2017.

Role type classification:

Categorical explanatory and quantitative response

Journal Titles:

Neoplasia
Molecular Oncology
Journal of Forensic Medicine
Immunobiology
Zeitschrift fur Medizinische Physik
European Journal of Cell Biology
Methods

There over 2000 journal titles in the dataset

Objectives:

Identify the journal title within 75 percentile above and 50 percentile with the response variable.

The journal titles with 50 percentile and above can be described as productive.

Statistical method:

The descriptive statistics will be used to rank the journal title to determine their productivity

The result will help to classify journal title as productive or not.

Is there relationship between view counts and citation counts of research documents?

Variables:

The **explanatory variable** in this question is **views** and the **explanatory variable** is citation.

Role type classification:

Quantitative explanatory and quantitative explanatory

Hypothesis:

Ho : There is a relationship between view and citation

Ha: There is no relationship between view and citation

Statistical method:

ANOVA will be used to determine the relationship between views and citations

The result will help to know if there is possibility of usage because of view and nature of the relationship.

Are papers published in higher impact journals, more cited?

Variables:

The **explanatory variables** in this question are **Impact indicator** (SNIP 2017) and **citation** and the **response variable** is journal title.

Role type classification:

Quantitative explanatory and quantitative explanatory

Hypothesis:

Ho : The citation mean of high impact journal is 11.12

Ha: The citation mean of high impact journal is difference from 11.12

Statistical method:

ANOVA will be used to determine the relationship between citations and normalized impact indicator

Constraints & Interpretations

A publication was co-authored by 200 plus authors

Constraints:

i.) Where a journal was co-authored from different institutions, the institutions received a ranking, citation, or impact indicator per journal. As such, analyzing institutions along the variables is not done because of time.

ii.) Where a journal was co-authored from different countries, the countries received a ranking, citation, or impact indicator per journal. As such, analyzing countries along the variables is not done because of time.

Interpretations:

The T-Test will be used because the standard deviation of the sample(i.e. Medline 2015-2016 data) will be used. The level of significance is 95%. The p-value will be determined and where the p-value is small, the data present enough evidence to reject H0 (and accept Ha). If the p-value is not small, the data do not provide enough evidence to reject H0.

The point estimate (mean) will be used for ranking of the unit of analysis (journal title). The mean of the unit of analysis will be compared with the sample mean to establish a cutoff point. Therefore; pointing to the productivity of the unit of analysis.

Thank you