Statistical software for medical professionals

Olivera Djordjevic,
Rehabilitation Clinic “dr Miroslav Zotovic”,
Medical faculty, University of Belgrade, Belgrade, Serbia
odordev@eunet.rs

Statistical software for medical professionals

Medical statistics has been introduced in medical research by sir Austin Bradford Hill in the late 1930s .

During the 1940s, 1950s and 1960s, the use of formal statistical methods in medical research grew and statistical data analysis subsequently became an inherent part of medical research method.

Statistical software for medical professionals

- The optimal situation in medical research environment
- Unfortunately, the optimal is not the most usual situation

Statistical software for medical professionals

The question of choice of the statistical software for handling medical data is wide and depends on many rational and some not so rational and yet, potentially much influential factors

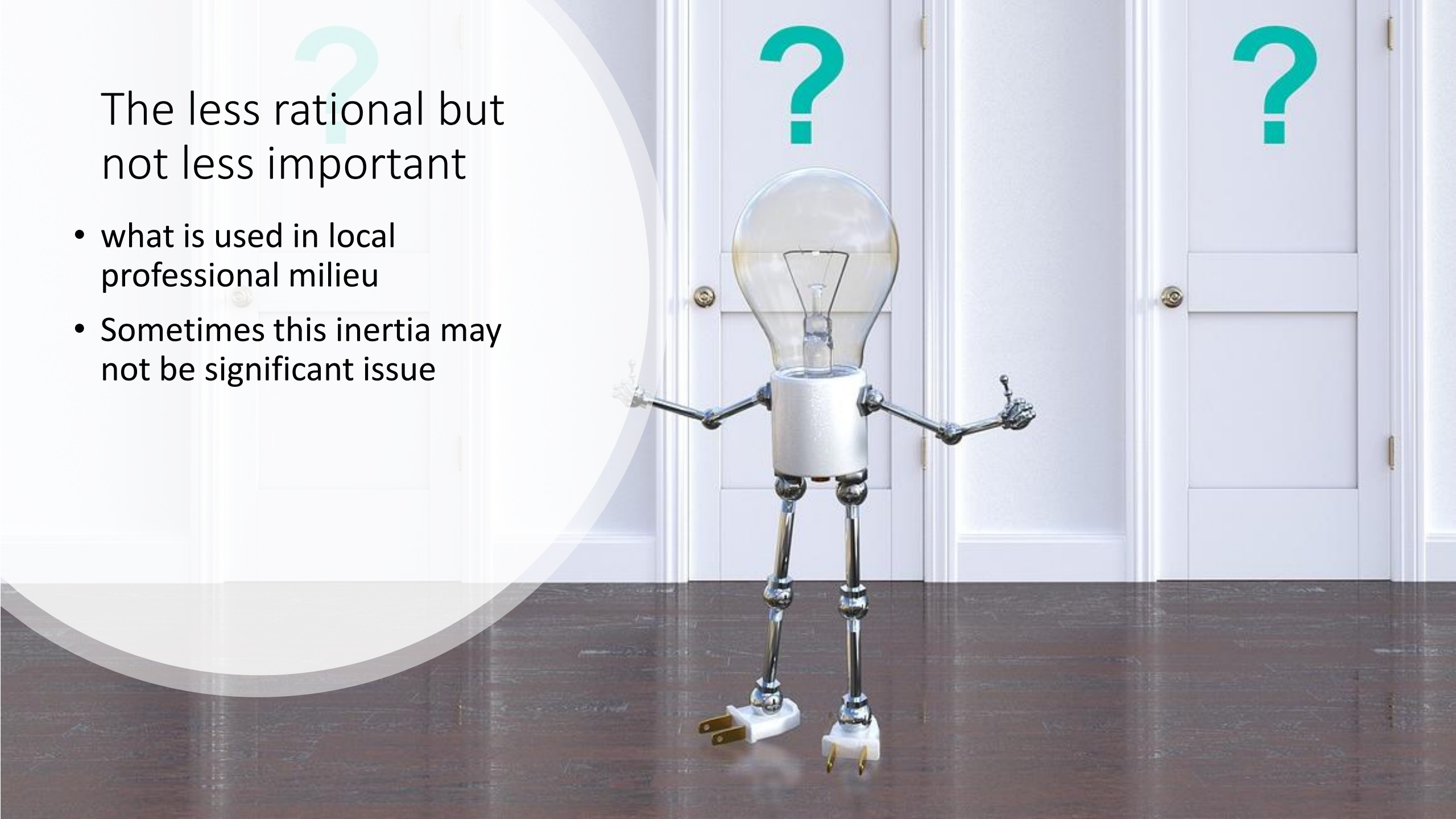
How do we choose data analysis tool for medical research

The choice of a particular software package for a particular medical research depend on

1. the study's specific analytical needs,
2. the nature of our data,
3. the suitability of a particular software application for a specific analysis,
4. the choice of the packages that are available to us the investigators' skills and experience,
5. the budget for statistical software if we have to buy it,
6. the time that has to be spent on learning the software

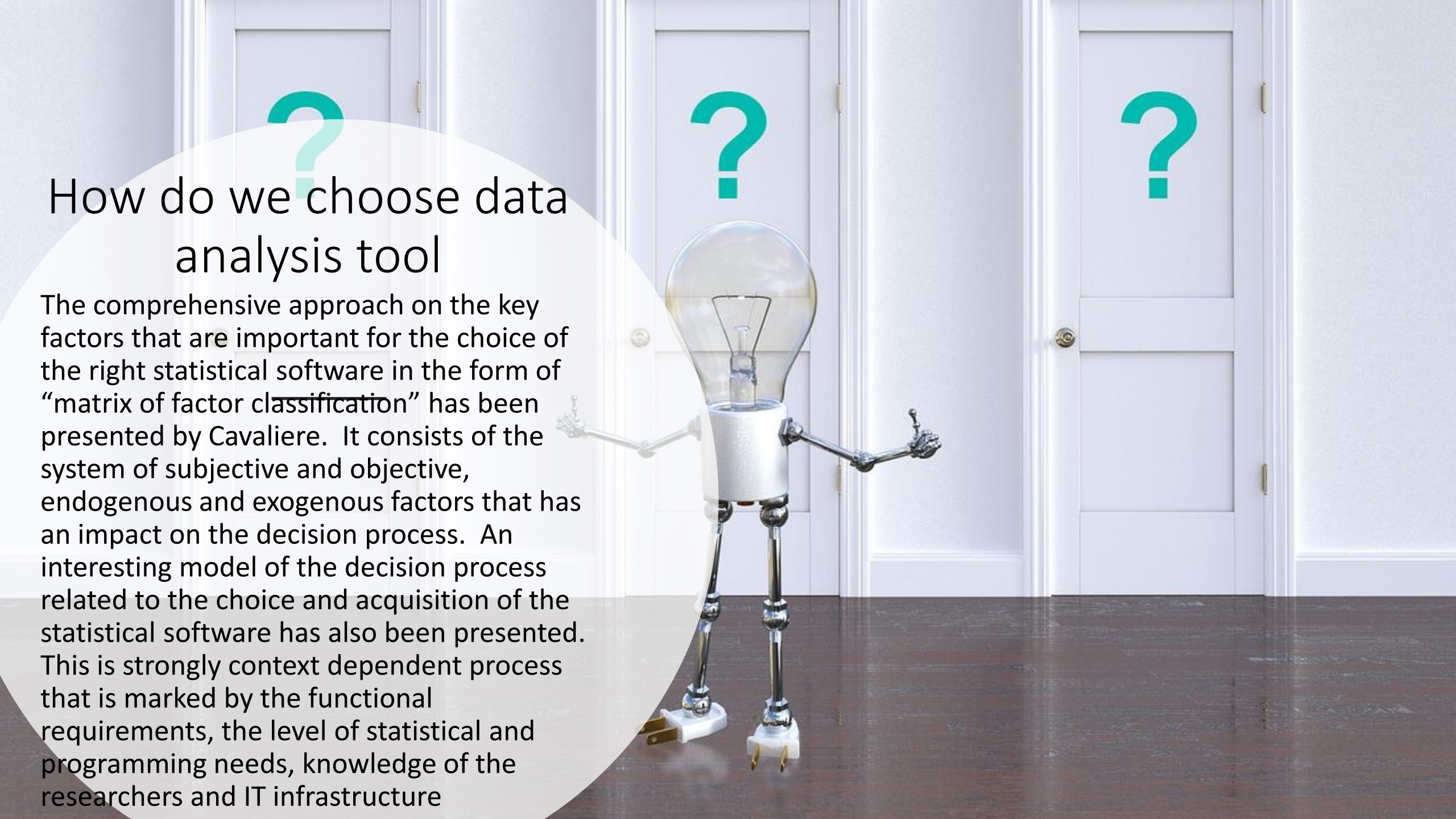
The less rational but not less important

- what is used in local professional milieu
- Sometimes this inertia may not be significant issue

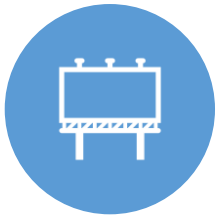


How do we choose data analysis tool

The comprehensive approach on the key factors that are important for the choice of the right statistical software in the form of “matrix of factor classification” has been presented by Cavaliere. It consists of the system of subjective and objective, endogenous and exogenous factors that has an impact on the decision process. An interesting model of the decision process related to the choice and acquisition of the statistical software has also been presented. This is strongly context dependent process that is marked by the functional requirements, the level of statistical and programming needs, knowledge of the researchers and IT infrastructure



Overview of prevalence of statistical software in medical research



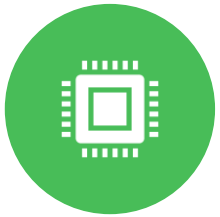
JOB ADVERTISEMENTS,

SCHOLARLY ARTICLES,

SURVEY OF USE, BOOKS,

BLOGS,

DISCUSSION FORUM ACTIVITIES,



PROGRAMMING POPULARITY MEASURES,

SALES AND DOWNLOADS,

COMPETITION USE GROWTH IN CAPABILITY.

Overview of prevalence of statistical software in medical research

According to *scientific articles*:

SPSS is still the most dominant package.

- It has been so for over 20 years.
- The balance between power and ease of use.

R is in the second place with around half as many articles.

- big power, but less ease of use.

Some trends of change…

Overview of prevalence of statistical software in medical research

Blogs , discussion forums:

- Research Gate: SPSS, R
- … SAS, GraphPad, XLSTAT
- …but also Datamelt *

*free, open-source, used in natural sciences, multi-platform and is very well documented

Overview of prevalence of statistical software in medical research

- The frequency of requirements for certain data analysis software in *job advertisements*, give us an estimation of how much is that software demanded.
- Job advertisements are comprehensively designed with the detailed description of the needed software skills and are backed by money. Therefore, they should be a trustful source of the need and popularity of certain data analysis software (Python, SQL, Java, Amazon ML,R…)
- However, medical research in our local community are seldom sponsored on a regular basis.

Overview of prevalence of statistical software in medical research

Surveys of use give us different estimations on the popularity of statistical software, may be a bit too strongly dependent of the undertaker of the survey itself.

- Rexer Analytics Survey (R, SPSS, SAS…)
- Lavastorm Analytics Community Group, Data Science Central and Kdnuggets (Excell, R, SAS, MS Access, SPSS…)
- Labome survey (Prism, SMP, StatView, Excell….SPSS and R are in the lower half of 10- item list)

Overview of prevalence of statistical software in medical research

Books

- The frequency of showing up a software ‘s name in the book title may give us an estimation on the software’s popularity.
- According to the number of books that include a software’s name in its title, the most popular software are SAS, SPSS Statistics, R, JMP…

Overview of prevalence of statistical software in medical research

The rise of software that uses the workflow (or flowchart) style of control has been observed as a trend, recently.

- easy to learn as menu driven software
- time- saving
- the wide use of this interface is allowing non-programmers to make use of advanced analytics,
- we have not observed significant rise in popularity of these potentially useful packages in medical research

What medical professionals need?

Successful data analysis is based on both statistical knowledge, mustering of the statistical software and the ability to interpret results.

Although medical professional seek a comprehensive statistical software, if the research team do not have fully designated statistical engineer or statistically educated team member, most of “lay” statistician seek software which is:

- intuitive
- cross-platform
- menu-driven
- do not have steep learning curve (or does not take considerable time and effort to learn).

Conclusion

- Although the popularity ranking of each package varies depends on the criteria used, we can still see major trends: SPSS , R, SAS, and Stata tend to always be in the top.
- Medical professionals who are doing statistics on their own in the research strive to the ease-of-use in a software and that is how quickly and effortlessly they can find out how to do what they want without time consuming prior instruction, consultation of manuals or third-party help. Trustful free software which fulfil these requirements would probably meet the needs of the most of them