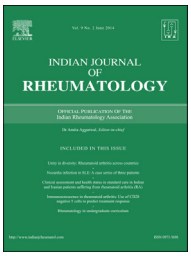
ScienceDirect

Available online at www.indianjrheumatol.com and www.sciencedirect.com



Nuts and Bolts in Clinical Research

Demystifying statistics: How to choose a statistical test?

*Subramanian Shankar**, *Ramakant Singh*

Department of Internal Medicine, Armed Forces Medical College, Pune 411040, India

ARTICLE INFO

Article history:
Received 2 April 2014
Accepted 3 April 2014
Available online 9 May 2014

Keywords:
Chi square
t test
Regression analysis

ABSTRACT

The young researcher today is confronted with a choice of hundreds of statistical tests, both while reading research works or while planning own research. The principles guiding the choice of statistical tests are simple and this article, aimed at the young researcher, aims to demystify the same.

Copyright © 2014, Indian Rheumatology Association. All rights reserved.

1. Introduction

With the advent of "evidence-based medicine", it became important for medical professionals to be well versed with the statistical methods not only to understand and critically analyse various research works but also to plan their own studies.[1] However, statistics is often regarded as a weak spot for many medical researchers, possibly because of the way it is taught in most medical schools with a lot of emphasis on technical details. Use of wrong or inappropriate statistical test is thus very commonly observed in various research works published in biomedical journals.[2,3]

Unlike in the past, the researchers of today neither have to perform statistical calculations manually nor do they have to be familiar with the mathematical principles of a statistical test. There are excellent softwares which can perform any statistical test in a very short time. The softwares however cannot select an appropriate statistical test to be used in a given situation. Therefore, it all boils down to understanding the basis of selecting a statistical test for analysis of a given data set, which remains a pure human effort. Although, there are well over 100 statistical tests in use, the great majority of research questions can be tackled by using only a handful of them. The test to be used is decided on the basis of the type of the research question, the type of data being analysed and the number of groups or data sets involved in the study.[4]

Choosing an appropriate test can be best compared to the process of learning how to drive a car. One can become an excellent driver without understanding much about the chassis, engine or fuel injection system of the car. In a similar way, the choice of statistical tests is independent of most of the theory governing various statistical tests.

This article is aimed at the young researcher and will focus on just one question; namely, which statistical test is to be used in which situation. It expects them to have no background understanding of statistical theory and practice. It is assumed

* *Corresponding author.*
E-mail address: shankarsid@gmail.com (S. Shankar).
http://dx.doi.org/10.1016/j.injr.2014.04.002
0973-3698/Copyright © 2014, Indian Rheumatology Association. All rights reserved.

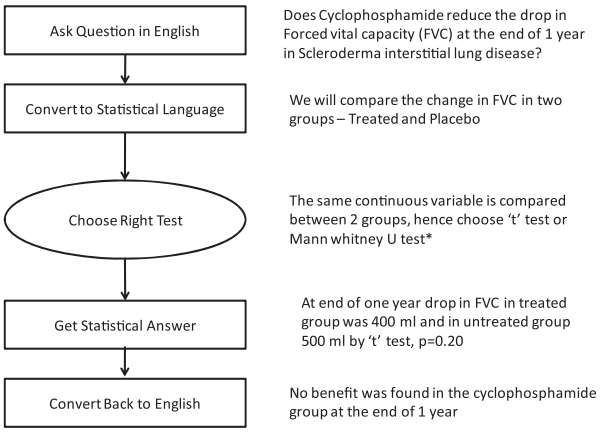


Fig. 1 – Five steps to choosing the correct statistical test with an example.

that the reader will eventually expand his/her knowledge over time about the various tests discussed in this article.

2. A step-by-step approach

To start with any research work, a research question is to be formulated in a language the researcher understands, namely, in plain English. Once the question is formulated, it has to be translated into statistical language which helps us in choosing the right sets of statistical tests. The answer that is obtained after applying the appropriate test is converted back to the original language to provide the final result (Fig 1).

The secret to choosing the correct statistical test therefore lies in articulating it correctly in plain English. It is often this first step that is not given adequate importance and which results in wrong choice of tests. The choice can best be visualized as a bilingual translation between English and statistical language.

This brings us to the next question. What is the statistical language? Statistics understands just one language: the language of "Variables".

3. What is a variable?

A variable is any feature or measurement in a given study that is to be analysed. Anything and everything we measure is a variable. For e.g. height, weight, smoking status, alcohol status, whether a person had a stroke or not; everything that can be measured is a variable. In all studies, we are usually doing one of the three things.

a. Assessing the characteristic of a variable
b. Comparing and contrasting 2 variables
c. Searching for associations between multiple variables

Variables are classified in 2 different ways. Based on their inherent nature/characteristic they are classified as Nominal (NV), Ordinal (OV) and Continuous (CV).[5] Based on the relationship they have with other variables they are classified as Dependant or Independent variable.

a. ***Dependant or response variable*** (DV): It is the variable of primary interest to the researcher. Each study has at least 1 dependent variable.
b. ***Independent or explanatory variable*** (IV): These are the variables that can influence/explain some fraction of DV. There can be many IV affecting a single DV.

Let us take a simple example. In a study that looks at multiple factors like age, smoking, diabetes, hypertension and hypercholesterolaemia affecting development of myocardial infarction, identify the DV and IV? Please pause and identify the DV and IV before going ahead.

The DV in this case is Myocardial infarction and the IVs are the other 5 factors. It is also apparent from the above example that both the DV and IV can be nominal, ordinal or continuous. After deciding the type of a variable, we will have to determine the type of analysis that can be done on a given set of observations (variables). Broadly, the analysis can be univariable, bivariable or multivariable. In some cases it gets influenced by time factor (survival analysis) and in others we modify it to compare with other tests (e.g., screening tests) (Fig 2).

The choice of tests also gets influenced by how the data is distributed in the population. For a normally distributed data,

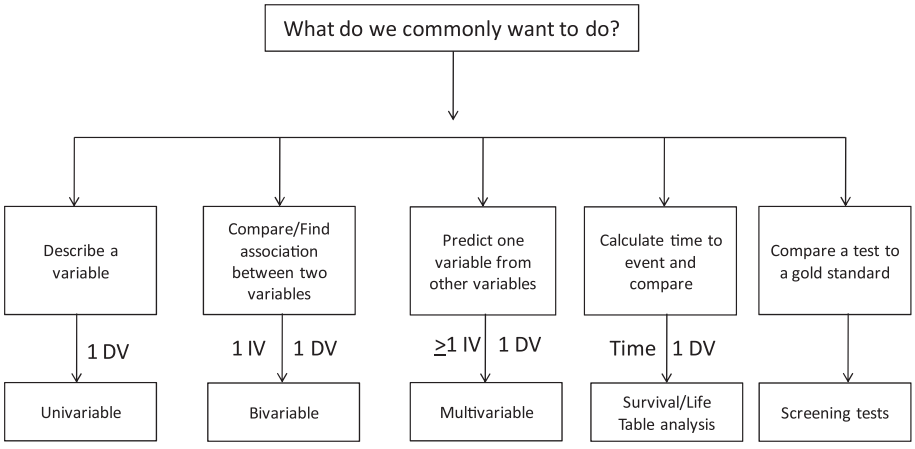


Fig. 2 – Various types of analysis. DV: dependent variable; IV: Independent variable.

Download English Version:

https://daneshyari.com/en/article/3357022

Download Persian Version:

https://daneshyari.com/article/3357022

Daneshyari.com