

Which Statistical Hypothesis Test Should I Apply? A Simple Guide for Beginners

Inaamul Haq, Aanisa Nazir¹

Department of Community Medicine, Government Medical College, Srinagar, Jammu and Kashmir, India, ¹Center for DNA Fingerprinting and Diagnostics, Laboratory of Bacterial Genetics, Hyderabad, Telangana, India

Correspondence to:

Dr. Inaamul Haq, Department of Community Medicine, Government Medical College, Srinagar - 190 010, Jammu and Kashmir, India.

E-mail: haqinaam@yahoo.co.in

How to cite this article: Haq I, Nazir A. Which statistical hypothesis test should I apply? A simple guide for beginners. *Int J Prev Med* 2016;7:81.

INTRODUCTION

Researchers in the field of medical and allied sciences (MAS) are generally mathematically faint at heart. The evolution from a student to a researcher is abrupt. Even though statistical methods are taught to an MAS student at the postgraduate level in most Indian universities, practical application of the same is mostly lacking. Once these students are thrown into the realm of research, they fail to apply even the basic methods in analytical statistics. A researcher works hard and honestly to collect good data but may report the wrong findings because an incorrect statistical test was used for data analysis. With this study, we intend to present a simpler way of answering the common question, “Which statistical hypothesis test should I apply?”

THE BUILDING BLOCKS

There are three essential aspects which we must understand before we can start analyzing any data.

The type of variable used

“Any aspect of an individual that is measured, like blood pressure, or recorded, like age or sex, is called a variable.”^[1] A “variable” may take different values in different individuals (or animals, objects, organisms, and populations) or in an individual at different times. Common examples include age, sex, weight, height, caste, religion, income, education, colony count, bacterial strain, antibiotic sensitivity, bacterial motility, bacterial morphology, and bacterial growth rate. A researcher must be very clear about the variable(s) being recorded or measured in the research. For the purpose of this study, we shall define four types of variables: “dichotomous,” “polychotomous,” “ordinal score,” and “scale.”

Dichotomous

Dichotomous variables have only two categories or levels. Variables with values such as “yes” or “no,” “present” or “absent,” “test” or “control” are dichotomous. Gender is a dichotomous variable with two values - “male” and “female.” Disease status (present/absent), exposure status (present/absent), residence (urban/rural), antibiotic sensitivity (sensitive/resistant), type of bacterial strain (wild type/mutant), and motility (motile/immotile) are examples of dichotomous variables.

Polychotomous: Polychotomous variables are categorical variables with more than two categories. Examples include caste, religion, blood group, political affiliation, type of organism, socioeconomic status, severity of disease or symptom, and Likert scale. The categories of a polychotomous variable may or may not have an inherent order.

Ordinal score

Sometimes, polychotomous variables have more than just few ordered categories based on a ranking or scoring system. A typical example is the commonly used scale for measuring pain - the visual analog scale.^[2] It has at least 10 ordered categories. Other examples include Clinical Global Impression^[3] and Yale-Brown Obsessive Compulsive Scale.^[4] For this study, we shall consider all polychotomous variables with seven or more ordered categories as “ordinal scores.”

Scale

These are quantitative variables. The characteristic that the variable indicates can be counted or measured. The measured ones usually have some unit of measurement (years, min, kg, cm, g, etc.). Examples include weight, height, age, clinical biochemistry parameters, pulse rate, bacterial doubling time, and bacterial colony size (in mm).