

Wilcoxin Mann-Whitney Test

true

10-24-2021

This is an exploration into this non-parametric test insofar as I can manage.

The document details the following: Some information about non parametric tests and What is the wilcoxin-mann whitney test The forumla and an intuition for how the test works Some considerations when using the test A stress test for when the test is no longer reliable

Some information about Non-Parametric Test At some point you realize that the initial statistical golems (or tests) that you are introduced initially have quite a bit of constraints on them. It is not obvious that the constraints are just that, constraints. So you eventually realize that you can remove them, although it's another skill to manipulate the tests to have the properties you desire. That's what a statistician can do, myself still being a wanna be hack that occasioanlly gets glossy eyed when I come across these things.

In any case, the definitions or rather guiding requirements for a non-parametric test being that it doesn't satisfy some requirement in a list that would qualify for a parametric test turns out to be an unreliable way to classify these golems. The working definitions that I have choosen (for now) is that non-parametric tests use distributions that can't be described by a finite number of parameters when making inferences about the population which is to say these use golems that are infinite dimensional (all of non para larry). The counter part, parametric inference assumes some information about the population probability density function, such that it belongs to a particular family of distributions that is defined by a finite number of parameters (intro to non-para 2009).

The wilcoxin-mann whitney test (WMWT) does not use any priori information about the population density and falls into the category as non-parametric.

some intuition about how it works

There are two names written into the title being wilcoxin and mann-whitney. It seems like developed tests that are equivalent to each other with one being contained in the other and the results the same. Wilcoxin derived a statistic that will be referred to as W being

$$\sum c_i R_i$$

which is referrerd to as a linear rank statistic. The ranks given by

$$R_i$$

are added together and weighted by

$$c_i$$

in some fashion the satistician deems useful. In our case for the WMW all of the

$$c_i$$

are values of one, and we operate with just the ranks.

The mann whitney version of events i found to be the more intuitive; this looks like

$$\sum_{i=1}^n \sum_{j=1}^m \phi(x_i, y_j)$$

where

$$\phi(x_i, y_j)$$

is defined as follows: \$\$