Advances and Applications in Statistics

Published Online: November 2014
Available online at http://pphmj.com/journals/adas.htm
Volume 42, Number 2, 2014, Pages 157-164

HOW RANDOMIZATION VALIDATES THE SIGNIFICANCE LEVELS OF THE FISHER EXACT TEST FOR 2×2 TABLES

David Sotres-Ramos

Programa de Estadística
Colegio de Postgraduados
carretera México-Texcoco, Km. 36.5, 56230
Texcoco, estado de México
México
e-mail: davidsotres@prodigy.net.mx

Abstract

It is shown that the Fisher Exact Test for a 2×2 table has valid significance levels under the sole assumption of random assignment of subjects to test treatment and control.

Introduction

The problem of deciding whether a proposed innovation constitutes an improvement over some standard procedure arises in many different contexts, such as industrial designed experiments, clinical research, agricultural experiments, etc. This paper focuses on the randomized experiment with two treatments: new and standard treatments, and when the response variable is dichotomous.

Received: September 3, 2014; Accepted: October 10, 2014
2010 Mathematics Subject Classification: 62F05.
Keywords and phrases: Fisher Exact Test, randomization, 2 × 2 table.

In the simplest case, randomization is a process by which each subject has the same chance of being assigned to either the test treatment or to control. An example would be the toss of a coin, in which heads (50%) indicates treatment group and tails (50%) indicates control group.

There are three very important consequences produced by randomization when applied in a study, see for instance Friedman et al. [2].

(i) Tends to produce study groups comparable with respect to known as well as unknown risk factors.

(ii) Removes investigator bias in the allocation of subjects, and

(iii) Guarantees that statistical tests will have valid significance levels.

Unfortunately, these remarkable properties of randomization are frequently absent in introductory statistical courses.

The main objective of this paper is to show that valid significance levels of the Fisher Exact Test for a 2×2 table, can be derived under the sole assumption of random assignment of subjects to test treatment and control.

The Experiment: A New Drug

Table 1. Respiratory outcomes of the clinical trial

Treatment	Respiratory Outcome		Row Totals
	Favorable	Unfavorable	
New treatment	10	2	12
Placebo	2	4	6
Column totals	12	6	18

An investigator in a hospital wishes to test the effectiveness of a new drug that is claimed to have a beneficial effect on some respiratory disorder. There are 18 patients in the hospital suffering from this disorder to about the same degree. Twelve patients are selected at random to receive the new drug, and the other 6 serve as controls: they are given a placebo, a harmless pill not

containing any active ingredients. After some time a visiting physician interviews the patients and record whether they have observed a favorable effect or not. Table 1 summarizes the information obtained from this randomized experiment with a dichotomous response variable.

The test medication group contains 12 patients and the placebo group contains 6 patients. In the hypothesis testing context, the question of interest is whether the rates of favorable responses for the test drug and placebo are the same. We can address this question by investigating whether there is a statistical association between treatment and outcome. The null hypothesis is stated as follows:

H: There is no association between treatment and outcome. (1)

A basis for testing the hypothesis *H* with valid significance level, guaranteed by randomization, is provided by the following consideration.

Suppose that the treatment has no effect, i.e., that a patient health is in no way affected by whether or not he receives the new drug. We shall refer to this assumption as the hypothesis *H* of no treatment effect. Since under the assumption of this hypothesis (for short, under *H*) the response of each patient is determined solely by his state of health, it is clear that the outcome of the patients does not depend on which of them receive the drug and which serve as controls. We may thus think of each patient outcome (favorable or unfavorable) as attached to him even before the assignments to treatment and control are made. The selection of 12 patients, among the 18 available patients, to receive the test drug then also selects 12 outcomes: those attached to the selected patients. Each possible such selection divides the outcomes into two groups: the number of favorable outcomes for the treated patients and of the controls. These divisions are displayed in Table 2 for all possible cases. Thus for example, the first box in Table 2 corresponds to the possibility that the 12 patients who eventually are awarded with favorable outcomes are those receiving the treatment (test drug).

As is seen from Table 2, the patients and hence their outcomes can be divided into two groups in 7 different ways. As will be discussed in the next section, the cases listed in Table 2 and its corresponding probabilities,

induced by the randomization process, provides a basis for assessing the significance of the observed outcomes in Table 1.

Table 2. All possible cases resulting from the new drug experiment

Table cell			
(1, 1)	**(1, 2)**	**(2, 1)**	**(2, 2)**
12	0	0	6
11	1	1	5
10	2	2	4
9	3	3	3
8	4	4	2
7	5	5	1
6	6	6	0

The Fisher Exact Test for a 2×2 Table

For comparing a new treatment with the standard method, N patients are divided at random into a group of n who will receive the new treatment and a control group of $m = (N - n)$ who will be treated by the standard method. For a dichotomous response variable, at the termination of the study, the number of patients treated with the new treatment with favorable outcomes (X) is obtained. The hypothesis H of no treatment effect is rejected, and the superiority of the new treatment is acknowledged, if this number of favorable outcomes (X) for the treated patients is sufficiently large. To complete the specification of the procedure, it is necessary to decide just when X is sufficiently large. The variable X is known as the test statistic.

The Hypothesis H is then rejected and the treatment judged to be effective when X is sufficiently large, say, when

$$X \geq c. \tag{2}$$

The test defined by (2) is known as the Fisher Exact Test, see for instance Fleiss et al. [1]. The constant c in (2), the critical value, is conventionally determined so that under H the probability of getting a value of X greater than or equal to c is equal to some specified small number α, the level of significance. Common choices of α are 0.01, 0.05. The constant c is thus determined by the equation

$$P_H\{X \geq c\} = \alpha, \tag{3}$$

where the subscript H indicates that the probability is computed under H, that is, under the assumption that the new treatment has no effect.

Formulas (2) and (3) make precise just when the test statistic X will be considered too large for the hypothesis of no treatment effect to remain tenable. On the one hand, values of X greater than or equal to c are very unlikely when H is true; in fact the probability of observing such values just by chance is then α. On the other hand, such values are expected when the treatment has the desired effect. The occurrence of such values therefore leads to the abandonment of H in favor of the alternative that the treatment is effective.

To determine c from equation (3) it is necessary to learn how to find the probability (under H) that X has any specified value. For the case $N = 18$, $n = 12$, $m = N - n = 6$ discussed in our example in Table 1, these probabilities are provided by the following consideration.

Note that the row totals in Table 1 (12, 6) are fixed by the treatment allocation process; that is subjects are randomly assigned to test and control. Also, the column totals (12, 6) can be regarded as fixed by the null hypothesis; there are 12 patients with favorable response and 6 patients with unfavorable response, regardless of treatment. Therefore, we may thus think the random assignment of 12 patients to the test treatment, as a sampling without replacement from a population of $N = 18$ elements which are divided in two groups: $D = 12$ favorable outcomes and $N - D = 6$ unfavorable outcomes. We then let X denote the number of favorable outcomes in a random sample of $n = 12$ elements. Since the X favorable

outcomes must come from the subgroup of $D = 12$ favorable elements in the population, and the $(n - X)$ unfavorable outcomes must come from the subgroup of $(N - D) = 6$ unfavorable elements in the population, the test statistic X has the hypergeometric distribution, see for instance Hogg and Craig [3], and is given by the following formula:

$$P_H\{X = x\} = \binom{12}{x}\binom{6}{12 - x}\Big/\binom{18}{12}. \tag{4}$$

In this way, one finds the probabilities (under H) of X taking on its various possible values, which are displayed in (5).

x	6	7	8	9	10	11	12
$P_H\{X = x\}$	0.0498	0.2560	0.4	0.2376	0.0533	0.0039	0.0001

(5)

These probabilities constitute the distribution of X under H. Since H is sometimes called the *null hypothesis* (it states that the treatment effect is zero or null), the distribution of X under H is called the *null distribution* of X.

From (5), it follows in particular, that

$$P_H\{X \geq 10\} = 0.0573. \tag{6}$$

Thus, if we set $\alpha = 0.0573,$ then the null hypothesis is rejected when $x = 10,$ 11 or 12. If we go back to our example in Table 1, we have obtained $X = 10,$ then our conclusion is to reject the null hypothesis and the new treatment is judged to be effective.

It is seen from (5) that the probability $P_H\{X \geq x\}$ takes on only a few values, namely 0.0001, 0.0040, 0.0573, 0.2943, 0.6943, 0.9503, 1.0. It is therefore not possible to find a critical value c satisfying (3) for every value of α but only for the values just listed.

Of course, the null distribution of X can be obtained quite generally by the same method used to calculate the distribution (4) for the case $n = 18,$

$n = 12$ and $D = 12$ = number of patients with favorable outcomes among the total number of N patients in the population. Therefore, in general (under H) X has a hypergeometric distribution, and write

$$P_H\{X = x\} = \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}. \tag{7}$$

For $x = 0, 1, 2, ..., n$; where $n \leq D$ and $n \leq N - D$.

Recall that a p-value is the probability of the observed data or more extreme data occurring under the null hypothesis. Therefore, to find the one-sided p-value of the Fisher Exact Test, you sum the probabilities as small or smaller than those computed for the value of X observed, in the direction specified by the one-sided alternative. In this case, it would be those values of X in which the test treatment had the more favorable response, or

$$p = 0.0533 + 0.0039 + 0.0001 = 0.0573.$$

Therefore, the hypothesis is rejected when $p \leq \alpha$, which of course is equivalent to the original rejection criterion $X \geq c$.

Fortunately, the one-sided p-value of the Fisher Exact Test may be calculated by several Statistical Computer Systems, such as Statistical Analysis System (SAS), Statistical Software for Exact Nonparametric Inference (StatXact). Several applications of the Fisher Exact Test are given in Stokes et al. [6].

Therefore, the main conclusion of this paper has been shown, which is: valid significance levels of the Fisher Exact Test for a 2×2 table, can be derived under the sole assumption of random assignment of subjects to test treatment and control.

It is important to note that the above derivation has the advantage of simplicity, and the required randomization is not difficult to carry out, thereby ensuring that the assumptions are satisfied. On the other hand, its narrow basis limits the scope of the resulting inference. Since no assumptions are made concerning the nature or provenance of the subjects, any inference

regarding the effectiveness of the treatment will refer only to the particular subjects in the study.

References

[1] J. Fleiss, B. Levin and M. Cho Paik, Statistical Methods for Rates and Proportions, 3rd ed., Wiley, Hoboken, New Jersey, 2003.

[2] L. M. Friedman, C. D. Furberg and D. L. DeMets, Fundamentals of Clinical Trials, PSG Publishing Company, Inc., Littleton, MA, 1985.

[3] R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics, 5th ed., Upper Saddle River, New Jersey, USA, 1995.

[4] SAS Institute Inc. SAS/STAT User's Guide, Version 8, SAS Institute Inc., USA, Cary, NC, 1999.

[5] Statistical Software for Exact Nonparametric Inference, StatXact User's Guide, CYTEL Software Corporation, Cambridge, MA, USA, 2005.

[6] M. E. Stokes, C. S. Davis and G. G. Koch, Categorical Data Analysis Using the SAS System, 2nd ed., SAS Institute Inc., Cary, NC, 2000.