Upsala J Med Sci 85: 97–102, 1980

Significance, Importance and Equality— Three Basic Concepts in the Analysis of a Difference

Adam Taube

Department of Statistics, University of Uppsala, Uppsala, Sweden

ABSTRACT

By means of data from ficticious cross over trials, it is first demonstrated that a statistically significant difference is not necessarily of a practically important order of magnitude. This fact is of special interest when the number of observations is large. Second, a statistically non significant difference does not prove the hypothesis about equality between, say, treatment effects. This fact is of special interest when the number of observations is small. For investigating whether equality is plausible, confidence intervals are more useful than non significant results from tests of significance.

INTRODUCTION

The purpose with this little note is to give an illustration of how elementary statistical methods can be applied in order to answer two principally different questions about a comparison between means or between proportions. The first question is whether a difference is statistically significant and the second whether it is of an important order of magnitude. If both these questions lead to negative answers, it is natural to ask whether it can be said that the two groups (treatments etc) being compared, are equal. The concepts "significance", "importance" and "equality" relate to three different aspects of a comparison. These aspects are all of interest both in observational studies and in experimental situations.

MATERIAL AND METHODS

As an example, we consider a cross over trial, where two drugs, A and B, are being compared. For simplicity, we assume that the results concerning each of them can be given in the form "improvement" or "no improvement". Thus, the data are of the kind illustrated in Table 1.

Table 1. Frequencies in a cross over trial.

Drug B	Drug A		
	No improvement	Improvement	Total
No improvement	a	b	a+b
Improvement	c	d	c+d
Total	a+c	b+d	n

The standard analysis concerning the statistical significance of the effect difference, is the well known McNemar test (2)

$$\chi^2 = (b - c)^2/(b + c) \qquad 1 \text{ d. fr.}$$

which is based only on those individuals who give different judgements of the two drugs. If this test statistic gives a sufficiently large value, the conclusion is that the two treatments have different effects. The result of this test is unaffected by the magnitude of the two frequencies a and d.

It is obvious that if the frequencies b and c constitute a very minor fraction of the total number of observations n, the result of the above test might be of little or no interest. This can be so, even if the χ^2-value is very large, indicating a strong statistical significance. Instead, one ought to study the rates of improvement

$$\hat{P}_A = (b+ d)/n \quad \text{and} \quad \hat{P}_B = (c + d)/n$$

and we notice that their difference, $\hat{D}$, can be written as

$$\hat{D} = \hat{P}_A - \hat{P}_B = (b - c)/n$$

This is the appropriate measure of whether the difference is important or not.

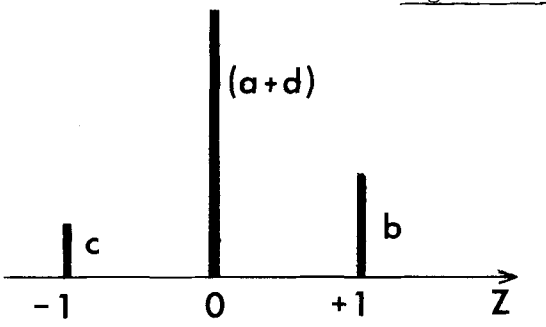
Let us now define a variable x^A such that

$$x^A = \begin{cases} 1 & \text{if improvement with drug A} \\ 0 & \text{if no improvement with drug A} \end{cases}$$

and similarly for x^B. For each of the n individuals, we can form the difference $z_i = x_i^A - x_i^B$ where $i = 1, 2, \ldots, n$. If, for a certain individual, there is no difference in effect between drug A and drug B we get $z = 0$. If A is better than B we get $z = 1$ and if drug B is better than drug A we get $z = -1$. Obviously, the variable z will have the values -1, 0 and + 1 with the frequen-

cies c, (a + d) and b respectively, as illustrated in Fig. 1

Fig. 1. Distribution of $z = x^A - x^B$



We notice that the average of the variable z is

$$\bar{z} = [(-1)\cdot c + (0) \cdot (a + d) + (+1) \cdot b] /n = (b - c)/n = \bar{D}$$

and it is easily verified that the standard deviation s_z is obtained from

$$s_z^2 = [n(b + c) - (b - c)^2]/(n - 1)\, n$$

If n is large, $\bar{z}$ can be assumed to be a normally distributed variable - irrespective of the shape of the distribution of the variable z. Therefore, we can give a confidence interval for the true difference D by means of

$$\bar{z} \pm k\ SE(\bar{z}) \qquad \text{where} \quad SE(\bar{z}) = s_z/\sqrt{n}$$

and the value of k is obtained from a table of the standard normal distribution. We denote the upper and lower confidence limits $\hat{D}_U$ and $\hat{D}_L$ and apply 95% confidence level, when calculating them for the six different data sets presented in Table 2.

Instead of using McNemar's test, it is possible to form the critical ratio

$$C.R. = \sqrt{n}\ (\bar{z} - 0)/s_z$$

This will give a very similar result, since the two test statistics are strongly related. In fact, it can easily be shown that

$$(C.R.)^2 = \chi^2\ (n - 1)/(n - \chi^2)$$

Table 2. Data from six fictitious trials.

Frequencies	Data set I	II	III	IV	V	VI
a	10	120	320	10	120	320
b	15	15	15	12	12	12
c	5	5	5	8	8	8
d	20	260	660	20	260	660
n	50	400	1000	50	400	1000
χ^2	5.0*	5.0*	5.0*	.8	.8	.8
C.R.	2.33*	2.25*	2.24*	.89	.89	.89
$\hat{D}_L 100$	3.2%	.3%	.1%	-9.6%	-1.2%	-.5%
$\hat{D}_U 100$	36.8%	4.7%	1.9%	25.6%	3.2%	1.3%

*) Significant. $P < .05$

DISCUSSION

When comparing data sets I, II and III we notice that the χ^2-values are exactly the same, indicating a statistically significant difference. However, the confidence interval for this difference is very wide in data set I, where n=50, while it is quite narrow and not far from the point zero in data set III, where n = 1000. For this latter material, it can be argued that, in spite of the fact that the result is statistically significant, the difference might be without any practical importance. Indeed, there is no contradiction in this: The results from the statistical test procedure just tells us whether a difference is larger than what could be due to chance, if the so called null hypothesis is true. When the number of observations is large, even a very unimportant difference can be statistically significant

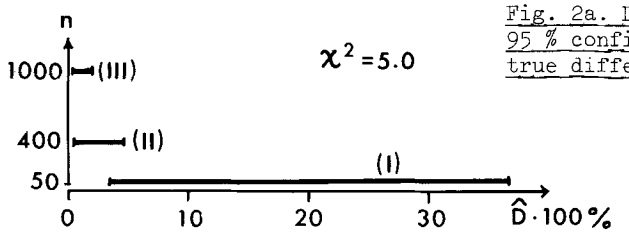


Fig. 2a. Data sets I, II and III. 95 % confidence intervals for the true difference $D \cdot 100\%$.

It is essential to stress that the judgement whether a difference should be considered as being important or not, is not possible to do merely by means of statistical techniques. The statistician can present a confidence interval, but it must be up to the subject matter specialist to decide the magnitude of what should be considered as an important difference.

The data sets IV, V and VI all give small χ^2-values. This means that no statistically significant differences have been found. The corresponding confidence

intervals always contain the point zero, but this is certainly not enough in order to demonstrate that the two drugs have equal effects. We notice that in data set IV, where n = 50, the confidence interval is quite wide.

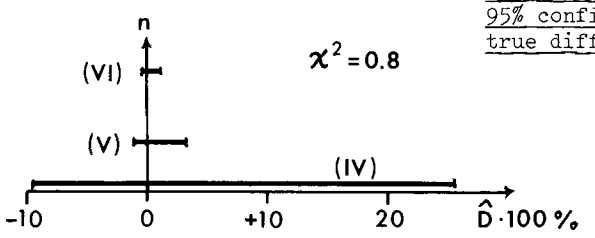


Fig. 2b. Data sets IV, V and VI. 95% confidence intervals for the true difference D·100%.

An interesting interpretation of a confidence interval in a situation like this is that all possible values of the true difference D within the calculated limits are not in contradiction with the data. This means that even if data set IV gives quite a low χ^2-value, there is still a possibility that the true difference could be of the magnitude 25%. However, when we consider data set VI, we find a very narrow confidence interval, indicating that even as small values of the difference D as, say 1.5%, would be in contradiction with the data.

Until now, we have just used the confidence level 95%. For data set VI, the 99% confidence limits for D would be $\hat{D}_L \cdot 100 = -.8\%$ and $\hat{D}_U \cdot 100 = +1.6\%$ respectively, thus giving an interval which includes the above mentioned value of 1.5%. Apparently, the conclusion about which values of D that are possible or not, is not only depending upon the number of observations, n, but also upon the confidence level chosen.

If it is desired to establish whether there is equality between the two treatments, it must be recognized that it is not possible to prove that D = 0. For a finite number of observations, the confidence interval for D will always have a certain length, unless b = c = 0 , which is the uninteresting case, when there is no random variation involved. Therefore, the equality statement must be substituted with the condition that the true difference is smaller than a certain value, i.e. $|D| < D'$, where D' is the smallest non-zero value of the difference which is of any interest to discover. If D' is specified by the subject matter specialist and also the confidence level and the desired power of the test is decided upon, it is possible to calculate the necessary number of observations in order to establish "equality" in accordance with this modified meaning of the word. One gets the impression that this procedure has seldom been applied, when it is stated in medical articles that "there is no difference" or that "the treatments are equal".

Indeed, everything in this discussion is of a very basic, elementary character. As the references (1, 3, 4, 5, 6) to this little note demonstrate, the problems touched upon are mentioned now and then in various journals. It is the

the author's impression that they are not sufficiently stressed in statistical text books, and it is easy to find applied scientific articles where the results from tests of significance have been interpreted incorrectly.

REFERENCES

1. Lutz, W. & Nimmo, I.A.: The inadequacy of statistical significance. Europ J of Clin Investigation 7:77-78, 1977.
2. McNemar, I.: Note on the Sampling Error of the Difference between Correlated Proportions or Percentages. Psychometrika 12:153-157, 1947.
3. Rennie, D.: Vive la différence (P < 0.05). New Eng J Med 299:828-829, 1978.
4. Spriet, A. & Bieler, D.: When can "non significantly different" treatments be considered as "equivalent"? Br J Clin Pharmacol 7:623-624, 1979.
5. Taube, A.: Blev det inte signifikant? Slutsatser vid statistisk hypotesprövning. Läkartidningen 69:65-68, 1972. (Swe)
6. Wade, C.L. & Waterhouse, J.A.N.: Significant or Important? Br J Clin Pharmacol 4:411-412, 1977.

Accepted January 15, 1980

Address for reprints:

Adam Taube

University of Uppsala

Department of Statistics

P.O. Box 513

S-751 20 Uppsala

Sweden