3. L. Harrison, J. J. Gallagher, J. H. Kasell, et al., Circulation, 55, 463 (1977).
4. H. B. Nell, A. S. Ketchan, and W. C. Hammond, Arch. Surg., 102, 45 (1971).

PLANNING OF AN INVESTIGATION BASED ON STATISTICAL COMPARISON OF MEANS

S. G. Kornilov

UDC 612-087+616-07]:31

KEY WORDS: statistical comparison of means; relative group size; optimization.

A considerable, if not the major, part of all medico-biological research consists essentially of the comparison of mean values of some features in experimental and control groups of objects. However, there is as yet no general agreement regarding the relative sizes of the groups (samples) for comparison of means (I am not discussing comparison of pairs), although most workers prefer them to be of the same size, on the grounds that this enables some simplification (in the case of equal groups) of the mathematical equations used to analyze the data [5].

Nowadays the soundness of this argument is questionable: Computer techniques have made everything much easier and, in any event, in all experimental research the work of mathematical calculation is only a small fraction of the work involved in the investigation as a whole.

Nevertheless, if dispersions in groups are not known beforehand and if the number of experimental and control groups is equal, it is best to distribute the obejcts studied equally among all groups: The symmetry of the scheme of such an experiment gives no grounds for any other distribution. However, in real investigations, the number of experimental groups as a rule will be greater than the number of control groups (most often of all there is only one control group), and in that case the control groups should preferably be larger than the experimental groups.

The reasons for this will be clear from the following example. Let us assume that an experimenter has available 100 animals and that he needs to have one control and nine experimental groups. In that case, with the traditional equal division of the animals each group would contain 10, and in each comparison of means 10 experimental and 10 control animals, a total of 20 animals, would be considered. But if, breaking with tradition, we remove one animal from all the experimental groups and transfer it to the control group, in each comparison 9 + 19 = 28 animals will be considered, so that the attainable level of significance of the difference between means will be increased. Yet this is achieved without any change in the total number of animals used in the experiment!

The control group also can be enlarged at the expense of the experimental groups. The answer to the question of what is the optimal number of animals in the groups is given by the equations presented below.

The mean level of significance $\bar{\gamma}$ obtained by a certain comparison will, of course, be increased if the corresponding difference between the general mean values exists objectively. If, however, the general means are equal (this, of course, happens extremely rarely in biology or medicine), no redistributions of the animals can help to prove the opposite.

The rational choice of group size is usually effective in medicine when clinical material is in short supply. Examples (see Fig. 2) when the use of an optimal scheme of investigation instead of the traditional scheme enables the number of patients in experimental groups to be reduced by almost half, and with the same number of patients six more experimental groups can be organized, are given in [3]. Clearly this represents both economy of work and a significant reduction in the time of the investigation. The situation when the switch from an optimal to a traditional experiment involves lowering of the mean level of significance of the results from $\bar{\alpha} = 0.05$ to 0.11, i.e., it actually reduces their value, is illustrated in Fig. 1 (it will be recalled that the significance of the results of the investigation $\gamma = 1 - \alpha$, so that a maximum of γ corresponds to a minimum of α).

Let us turn to the mathematics. We shall consider that each of μ experimental groups of animals will be compared with each of ν controls. When such investigations are planned, one of the folllowing problems will have to be solved.

1. The total number L of animals taking part in the experiment is known. How should this number be distributed among groups so as to obtain the highest level of significance $\bar{\gamma}$ of the experimental results (Fig. 1)?

Saratov Medical Institute. (Presented by Academician of the Academy of Medical Sciences of the USSR M. L. Studenikin.) Translated from Byulleten' Eksperimental'noi Biologii i Meditsiny, Vol. 92, No. 8, pp. 112-114, August, 1981. Original article submitted February 16, 1981.

 0007-4888/81/9208-1130$07.50

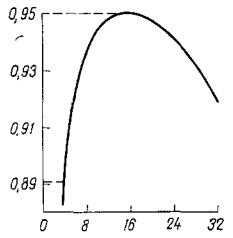

Fig. 1

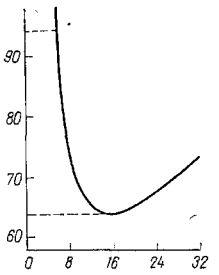

Fig. 2

Fig. 1. Example of relationship between mean significance $\bar{\gamma}$ of difference between means obtained as a result of experiment (ordinate) and size of control group n (abscissa) while total number of animals remains unchanged L = 62 (one control group, 16 experimental groups). Lower broken line shows level of $\bar{\gamma}$ obtained with traditional equality of size of all groups; top broken line shows level of γ corresponding to optimal value of n.

Fig. 2. Example of relationship between number of animals L needed for an experiment (ordinate) and size of control group n (abscissa) to give unchanged experimental results (one control group, 16 experimental groups). Top broken lines show level of L needed with traditional equality of size of all groups; bottom broken line shows level of L corresponding to optimal n. In the second case the number of experimental animals L – n is reduced by almost half compared with the first.

2. Let us assume that the significance of the experimental results is $\bar{\gamma}$. What should be the ratio between the numbers of animals in the groups so as to obtain this value of γ with the smallest number L of animals used (Fig. 2)?

A condition of the other tasks may be maximization of the number μ of experimental or the number $\mu\nu$ of control groups or of the resultativeness of the experiment – the probability of achieving the assigned value of γ by averaging for all $\mu\nu$ comparisons. This value, introduced in [3], depends on the type of statistical criterion used to compare the means and it is a complex function of γ, L, μ, ν and the other experimental parameters. It has been simplified and the solutions to the problems corresponding to the extremum of each of the parameters taken separately have been found. The solutions, which were found to be the same for all tasks, are as follows.

If the mathematical procedure of comparison of means is carried out with the aid of Wilcoxon's criterion for independent samples [6] (in some books on biometrics this is incorrectly described as White's criterion [1]), the total number of individuals in the control groups N must be:

$$N = \frac{\nu}{2} \cdot \frac{2L + \mu - \sqrt{\mu\nu}}{\nu + \sqrt{\mu\nu}}.$$

The number N is distributed as far as possible equally among the control groups. The rest of the L – N individuals are also divided equally among the experimental groups.

If the curves of probability density distribution of the values of the test features in the groups are definitely symmetrical in form, the generalized Student's t test is used for their comparison [2, 4]. In that case the planning has two possible variants.

A. Dispersions of the feature in the groups are not known beforehand. In that case

$$N = \frac{\nu L}{\nu + \sqrt{\mu\nu}} \cdot \left(1 + \frac{\tau_\alpha^2}{4} \frac{\mu - \sqrt{\mu\nu}}{L - 4\sqrt{\mu\nu}}\right).$$

Here τ_α is the critical value of the normal distribution. Some of the values of $\tau_\alpha^2/4$ are given in Table 1. The term α represents the level of significance regarded as adequate for the conduct of the planned investigation.

B. Dispersions in the groups are known approximately from previous experiments. Let the dispersion in the experimental group with serial number i be Δi and the dispersion in the l-th control group be D_l. In that case the number of individuals in each experimental group m_i and in each control group n_l must be:

$$m_i = \frac{L\Delta_i}{a + \sqrt{ab}}, \quad n_l = \frac{LD_l}{b + \sqrt{ab}},$$

TABLE 1

Criterion	α	
	0,05	0,01
One-way	0,68	1,35
Two-way	0,96	1,66

where

$$a=\sum_{i=1}^{\mu}\Delta_i,\quad b=\sum_{l=1}^{\nu}D_l.$$

It will be noted that there is no suggestion of making any comparisons of means between experimental and experimental groups, or equally between control and control groups, or that is not the main purpose of the investigation. If, however, all possible comparisons between all groups are planned to be equally important, if dispersions are unknown the size of all groups must be the same, but if dispersions are known:

$$m_i=\frac{L\sqrt{\Delta_i}}{\sum_{q=1}^{\mu}\sqrt{\Delta_q}+\sum_{r=1}^{\nu}\sqrt{D_r}},$$

$$n_l=\frac{L\sqrt{D_l}}{\sum_{q=1}^{\mu}\sqrt{\Delta_q}+\sum_{r=1}^{\nu}\sqrt{D_r}}.$$

If in the course of the experiment each group is isolated spatially (a cell, or an aquarium for animals, a room for people, and so on), each large (and also each average, if present) group is best divided into isolated subgroups, equal in size to the smaller groups, in order to achieve identity of the conditions of existence and equalization of possible collective effects. For this purpose, the sizes of the groups obtained from the equations given above must be modified a little in order to ensure that they are multiples of one another. Division into subgroups also has the advantage that it enables the subgroups to be arranged (from the spatial point of view) uniformly among the small groups, so that identity of physical (temperature, lighting, and so on) conditions in them can be achieved.

In this way the purity of the experiment is increased. However, disturbance of the optimal group sizes requires (if the resultativeness is to remain unchanged) some increase in the total number of individuals.

Everything stated in this paper is applicable, of course, not only to biology and medicine, but also to other sciences – in all cases when statistical comparison of means has to be undertaken.

LITERATURE CITED

1. J. Hajek and Z. Sidak, Theory of Rank Tests, Academic Press (1967).
2. L. Sachs, Statistischen Methoden, Berlin, Springer (1972).
3. V. N. Pagurova, A Criterion for Comparison of Mean Values [in Russian], Moscow (1968).
4. G. W. Snedecor, Statistical Methods Applied to Research in Agriculture and Biology [Russian translation], Moscow (1961), p. 100.
5. V. Yu. Urbakh, Statistical Analysis in Biological and Medical Research [in Russian], Moscow (1971), p. 111.