How to choose an appropriate experimental design type (Part 2)

Liang-ping Hu, Xiao-lei Bao, Chen-long Lü
Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China

ABSTRACT: How to choose an appropriate design type to arrange research factors and their levels is an important issue in scientific research. Choosing an appropriate design type is directly related to the accuracy, scientificness and credibility of a research result. When facing a practical issue, how can researchers choose the most appropriate experimental design type to arrange an experiment based on the research objective and the practical situation? This article mainly introduces the related contents of the design of one factor with two levels and the design of one factor with \( k (k \geq 3) \) levels by analyzing some examples.

KEYWORDS: statistics, medical; research design; matched-pair analysis; factor analysis, statistical

Scientific research design includes investigational design, experimental design and clinical experimental design, whose main contents are covered in the experimental design. The kernel of the experimental design includes the three elements, the four principles and the design type. Before conducting an experiment, researchers should fully understand all the important issues that are related to the research in order to work out a feasible plan. The value of a well-designed research plan is that it can achieve a reliable result with less human power, money and time. It can also accurately control and estimate the error, and examine more experimental factors by running as fewer experiments as possible. In this article, we will continue to discuss the design of one factor with two levels and the design of one factor with \( k (k \geq 3) \) levels which belong to the nonregression design by analyzing practical examples.

1 The design of one factor with two levels

1.1 Example 1 A researcher expected to evaluate the clinical therapeutic effects of cervical fixed-point traction manipulation in treating patients with cervical spondylotic radiculopathy. The temperature of the skin of the upper limbs (the normal limb and the abnormal limb) detected by infrared thermal imaging system was supposed to be contrastively analyzed. Which design type should be adopt[1]?

In example 1, the research subjects are patients
with cervical spondylotic radiculopathy. The experimental factor is “using cervical fixed-point traction manipulation or not”. The observational index is “the temperature of skin of the upper limbs”. The research objective is to analyze whether the difference of the temperature of skin of the upper limbs between the experimental group and the control group is statistically significant. The research factors include experimental factors and nonexperimental factors. Statistically, the external factors which may influence the result are called experimental factors while the internal characteristics of the subjects themselves are called nonexperimental factors. Some of the nonexperimental factors have a slight influence on the result, while others may have a strong influence. The latter is called important nonexperimental factors, or block factors. In a scientific research design, researchers usually pay more attention to the experimental factors. The purpose of considering the block factors is to control and eliminate their influence on the result as much as possible so that researchers can make a true evaluation of the influence of the experimental factors. If researchers only need to examine the block factors, they can view them as the experimental factors. The experimental factors and the block factors together are called research factors.

As to example 1, the research objective is to analyze whether the difference of the temperature of skin of the upper limbs between the experimental group and the control group is statistically significant. Therefore, first, the researcher needs to randomly select some patients suffering from cervical spondylotic radiculopathy as the research subjects and divide them randomly into two groups. One group uses the cervical fixed-point traction manipulation; the other does not. The sample sizes of the two groups should be close, or equal if possible, and should be calculated based on corresponding formulas. Second, measure the temperature of skin of the upper limbs of the two groups and make a comparison. Thus, example 1 involves the design of one factor with two levels. Since the levels of the experimental factor are not related to the grouping of subjects, example 1 is called the design of one factor with two levels by complete randomized sampling and grouping. If the levels of the experimental factor are related to the grouping of the subjects and the subjects should be selected from two specified populations, it is called the design of one factor with two levels by group randomized sampling.

### 1.2 Example 2
A researcher wanted to investigate the effects of chemotherapy combined with Kangliu Zengxiao Decoction (KLZX), a compound Chinese herbal drug, on tumor markers carbohydrate antigen 50 (CA50), cytokeratin 19 fragment (CYFRA21-1) and carcinoembryonic antigen (CEA) in patients with advanced non-small-cell lung cancer (NSCLC). Which design type should he choose?

In example 2, the research subjects are patients suffering from advanced NSCLC; the experimental factor is “using chemotherapy combined with KLZX or not”; the quantitative observational indexes are CYFRA21-1, CA50 and CEA; the research objective is to analyze whether the differences of the three quantitative indexes of the two groups are statistically significant. Therefore, the researcher needs to select a certain number of patients suffering from advanced NSCLC for the experiment and randomly divide them into two groups. One group uses chemotherapy combined with KLZX while the other does not. Respectively measure the values of CYFRA21-1, CA50 and CEA of the two groups and make a comparison. Therefore, example 2 is three-variante quantitative data with the design of one factor with two levels. Since the levels of the experimental factor are not related to the grouping of subjects, example 2 belongs to the design of one factor with two levels by complete randomized sampling and grouping.

### 1.3 Definition of the design of one factor with two levels
The design of one factor with two levels is applied to research that only involves one experimental factor with two levels. If the experimental factor is independent of the subjects, they can be

<table>
<thead>
<tr>
<th>Related Articles</th>
<th>推荐阅读</th>
</tr>
</thead>
</table>
divided into two groups by complete randomization; otherwise, the subjects should be selected from two specified sub-populations and divided into two groups by complete randomization. Observe the quantitative observational index(es) of the subjects after they take different treatments. Above is the main idea of the design of one factor with two levels.

1.4 Forms of the design of one factor with two levels

The design of one factor with two levels has two common forms: one is the design of one factor with two levels by complete randomized sampling and grouping; the other is the design of one factor with two levels by group randomized sampling. The former means that the levels of the experimental factor are unrelated to the grouping of the subjects. Pick $n$ subjects from a specified population and divide them into two groups by complete randomization. The two groups of subjects receive different treatments. For example, examine whether the efficacy of two drugs treating the same disease has a statistically significant difference. The latter means that the levels of the experimental factor are related to the grouping of the subjects so that the researcher can only select the subjects from two specified sub-populations. For example, analyze whether the average height of ordinary male adults and female adults has a statistically significant difference. As to the observational index, if only one quantitative index is involved, the data are called univariate quantitative data with the design of one factor with two levels; if there are $m$ ($m > 1$) related quantitative indexes in the data which need to be analyzed together, the data are called $m$-variate quantitative data with the design of one factor with two levels.

1.5 Characteristics of the design of one factor with two levels

The design of one factor with two levels only deals with one experimental factor with two levels. No important nonexperimental factor is taken into consideration. The researcher hopes to balance the influence of nonexperimental factors on the result by randomized sampling and grouping.

1.6 Scope of application

The researcher only concerns one experimental factor with two levels and no other experimental factors or block factors are involved. The research does not meet the requirements of the paired design. In that case, the design of one factor with two levels can be used.

1.7 Implementation

The following aspects need to be taken into consideration when the design of one factor with two levels is implemented:

First, estimate the minimum sample size based on the research objective, the requirements of the experiment and the sample size estimation formulas when running hypothesis testing to quantitative data with the design of one factor with two levels.

Second, correctly choose the populations which the two levels of the experimental factor are corresponding to. If the design of one factor with two levels by complete randomization is adopted, the $n$ subjects should be randomly selected from the same population and randomly divided into two groups corresponding to the two levels of the experimental factor; if the design of one factor with two levels by group randomization is adopted, the subjects should be randomly selected from two sub-populations corresponding to the two levels of the experimental factor and take different treatments.

Third, the two groups of research subjects should be the same on other nonexperimental factors (for example, individual difference, state of illness, etc.) in order to avoid the influence of nonexperimental factors on the result.

Fourth, choose observational indexes that well reflect the difference of the two levels of the experimental factor based on specialty and measure them scientifically.

Fifth, pay attention to the difference between the design of one factor with two levels and the paired design. As to the paired design, every two data that correspond to each quantitative index are paired, which may be measured from the same subject, from two subjects with the same parental generation or from two subjects with similar conditions. In terms of the design of one factor with two levels, only one experimental factor with two levels is observed, and the subjects are not paired.

2 The design of one factor with $k$ ($k \geq 3$) levels

2.1 Example 3

A researcher expected to observe the effects of cinobufacini injection on serum levels of thyroid-stimulating hormone (TSH) and adrenaline (ADR) in rats, and to speculate the property (cold or heat) of the drug, which design type should he use?°

In example 3, the research subjects are known as rats; the research factor is “treatment”; the quantitative observational index is “TSH and ADR levels”; the research objective is to analyze whether TSH and ADR levels among the four groups (blank control group, heat syndrome group, cold syndrome group and cinobufacini injection group) have a statistically significant difference. Therefore, the researcher needs to randomly pick a certain number of rats and randomly divide them into four groups. Each group receives one treatment. Measure TSH and ADR levels of the four groups and make a comparison. The sample sizes of the four groups should be close. Thus, example 3 deals with the design of one factor with four levels. Since the levels of the experimental factor are not related to the grouping of subjects, example 3 is called two-variate quantitative data with the design of one factor with four levels by complete randomized sampling. If the levels of the experimental factor are related to the grouping of subjects and the subjects should be selected from $k$ specified populations, it is called the design of one
factor with \( k (k \geq 3) \) levels by group randomized sampling.

2.2 Example 4 In order to assess the therapeutic effects of Sisheng Decoction, a compound traditional Chinese herbal medicine, on a mouse model of yin deficiency syndrome induced by thyroid hormone, and to make the preliminary study on its mechanisms, a researcher wanted to study the effects of Sisheng Decoction with different doses (low, medium and high) and use serum concentration of malondialdehyde as a measuring index. Which design type should he use?4

In example 4, the subjects are known as rats with yin deficiency syndrome induced by thyroid hormone; the research factor is “dose of Sisheng Decoction”: the quantitative observational index is serum concentration of malondialdehyde; the research objective is to analyze whether the serum concentrations of malondialdehyde of the four groups (blank control group, low-dose Sisheng Decoction group, medium-dose Sisheng Decoction group and high-dose Sisheng Decoction group) have a statistically significant difference. Therefore, the researcher needs to randomly select a certain number of rats, induce them into yin deficiency syndrome by thyroid hormone, randomly divide them into four groups, respectively measure the serum concentration of malondialdehyde of the four groups and make a comparison. Thus, example 4 involves the design of one factor with four levels. Since the levels of the experimental factor are not related to the grouping of the research subjects, example 4 belongs to the design of one factor with four levels by complete randomized sampling and grouping.

2.3 Definition of the design of one factor with \( k (k \geq 3) \) levels The design of one factor with \( k (k \geq 3) \) levels involves an experimental factor with \( k (k \geq 3) \) levels. If the experimental factor is independent of the research subjects, the subjects can be divided into \( k \) groups by complete randomization; otherwise, the subjects should be randomly chosen from \( k \) sub-populations. After the subjects take different treatments, measure the observational index(es). Above is the main idea of the design of one factor with \( k (k \geq 3) \) levels.

2.4 Forms of the design of one factor with \( k (k \geq 3) \) levels The design of one factor with \( k (k \geq 3) \) levels has two common forms: one is the design of one factor with \( k (k \geq 3) \) levels by complete randomization, meaning that when the levels of the experimental factor are not related to the grouping of research subjects, the subjects can be divided into \( k \) groups by complete randomization and take \( k \) different treatments. For instance, examine whether the efficacy of \( k \) drugs for the same disease has a statistically significant difference. The other is the design of one factor with \( k (k \geq 3) \) levels by group randomized sampling, meaning that when the levels of the experimental factor are related to the grouping of research subjects, the subjects can only be randomly selected from \( k \) specified populations. For example, examine whether the average height of common adults with different blood types (A, B, AB and O) has a statistically significant difference.

With regard to the design of one factor with \( k (k \geq 3) \) levels, if there is only one quantitative observational index, the data are called univariate quantitative data with the design of one factor with \( k (k \geq 3) \) levels; if there are \( m (m > 1) \) indexes that are related in specialty and are required to be analyzed together, the data are called \( m \)-variate quantitative data with the design of one factor with \( k (k \geq 3) \) levels.

2.5 Characteristics of the design of one factor with \( k (k \geq 3) \) levels The design of one factor with \( k (k \geq 3) \) levels only involves one experimental factor with \( k (k \geq 3) \) levels. The researcher does not make any arrangement for important non-experimental factors; instead, he wants to balance the influence of all the nonexperimental factors on the result by complete randomized grouping.

2.6 Scope of application When the researcher only cares one experimental factor with \( k (k \geq 3) \) levels and hopes to eliminate or decrease the influence of all the nonexperimental factors by randomization, the design of one factor with \( k (k \geq 3) \) levels can be used.

2.7 Implementation The first four steps that have been mentioned in the implementation of the design of one factor with two levels are also suitable for the design of one factor with \( k (k \geq 3) \) levels. Besides, the researcher should pay special attention to the experimental requirements and research objectives in order to fully understand the real meaning of the word “groups” and avoid false judgment.

There are three types of common false judgment involving the design of one factor with \( k (k \geq 3) \) levels: the first one is using the word “groups” to represent the experimental conditions that are formed by the complete combination of the levels of multiple factors. In that case, it is very likely that the data are from a standard multifactor design (for example, the factorial design is the most commonly used multifactor design); the second one is to use the word “groups” to represent the experimental conditions that are scientifically selected from the complete combination of multiple factors. In that case, it is very likely that the data are from a standard multifactor design that aims to save the sample size (for example, the multifactor orthogonal design or the uniform design); the third type is to use the word “groups” to represent the experimental conditions that are arbitrarily chosen from the complete combination of multiple factors. In that case, it is very likely that the data are not from a standard multifactor design, which is named as multifactor nonbalanced-combination design5,6.
REFERENCES


