Correspondence

When is an hypothesis null?

DEAR SIR,

May I attempt an answer to Mr. Backhouse's query (*Mathl Gaz.*, June 1973)?

A convenient and authoritative source of definitions of statistical terms is Kendall and Buckland's *Dictionary* [1]. The entry for "null hypothesis" in the 1970 edition explains that in general the null hypothesis is the hypothesis under test as distinct from the alternative hypotheses which are under consideration: it is the hypothesis which determines the probability of the Type I error. In some contexts the term is restricted to hypotheses of "no difference".

Since the null hypothesis is simply the hypothesis to be tested, it is not essential that the word "null" should appear in elementary texts, and maybe it would be better if it were not used at this stage. Sampling is a matter of common sense: most of us are willing to judge the whole from our knowledge of a part, and it is certainly a long-established practice in commerce. The basic ideas of sampling can be expressed in every-day language. We should encourage beginners to explain their calculations and assumptions in their own words rather than use a jargon which they only half understand. Technical terms can be introduced later when students have a firm grasp of principles.

Like Mr. Backhouse, I have checked my own textbook [2]. I certainly included the phrase "null-hypothesis", though without giving a formal definition, and its use adds nothing to my explanation of sampling theory. I included it for two reasons. I assumed that in elementary courses, students should become familiar with the terms they will certainly meet in other texts. Also I find that when faced with the problem of testing the significance of the difference of two means, many students are not sure whether they should be testing the hypothesis of some difference or the hypothesis of no difference. A special name for the latter can help them to make the right choice. On the other hand it might be better if students were encouraged to state both the hypothesis to be tested and the alternative hypothesis to be accepted, if the first is rejected, before making any calculations.

Though the name "skittle hypothesis" might appeal to beginners, its use could be more misleading than helpful.

It is unfortunately true that some social scientists regard significant results as more important and more worthy of publication than non-significant results, but such an attitude can lead to a bias in the publication of research results [3]. This is not an attitude that should be encouraged.

Statistical inference is part of scientific method: in principle, then, significant and non-significant results are equally useful.

Also the analogy with skittles suggests a conclusiveness about the result, which is not a characteristic of statistical inference. At the end of the round, the skittles are either up or down, but hypotheses can only be accepted or rejected at some level of probability. It is possible to accept a wrong hypothesis, and reject a correct one, though the chance of doing so may be difficult to calculate in practice.

Yours faithfully,

FREDA CONWAY

Department of Sociology, Government and Administration, University of Salford, M5 4WT

References

1. M. G. Kendall and W. R. Buckland, *Dictionary of statistical terms.* Oliver and Boyd (1971).
2. Freda Conway, *Sampling: an introduction for social scientists.* Allen and Unwin (1967).
3. G. W. Webster and T. A. Cleary, A proposal for a new editorial policy in the social sciences, *American Statistician* 24, No. 2 (April 1970).

DEAR MR. BACKHOUSE,

I was interested to see your letter in *Mathematical Gazette* about the null hypothesis. The first part of your quotation from Siegel makes for a reasonable definition. For example, experiments were made by the third Lord Rayleigh to determine the density of nitrogen prepared in various ways. In one series of determinations, he weighed the nitrogen in a standard bulb after preparation by passing air over hot copper. In another series, air was passed over hot iron. The null hypothesis is that the method of preparation makes no difference to the mean weight of nitrogen in the bulb.

However, not all hypotheses are null hypotheses. In general, a hypothesis is an assertion about the probability distribution of a random variable. Your example falls into this category. The hypothesis is that the probabilities are

   (i) $\frac{1}{4}$ for two heads,
  (ii) $\frac{1}{2}$ for a head and a tail,
 (iii) $\frac{1}{4}$ for two tails.

The conclusion of Kendall and Stuart, *The advanced theory of statistics*, Vol. 2, ch. 22, is sensible. They have a footnote as follows: "The hypothesis under test is often called 'the null hypothesis', and the size of the test 'the level of significance'. We shall not use these terms, since the words 'null' and 'significance' can be misleading."

<div align="right">
Yours sincerely,
R. L. PLACKETT
</div>

Department of Statistics, University of Newcastle upon Tyne, NE1 7RU

DEAR SIR,

The traditional Neyman–Pearson hypothesis-testing procedure is as follows. Two statements (*hypotheses*) are made which are exclusive and exhaustive. On the basis of some statistical evidence E, one is to be *accepted* and the other *rejected*. Initially, one hypothesis is provisionally accepted and called the null-hypothesis (H_0): on this basis a probability model is constructed which, after choice of a significance level α which limits the probability of wrongly rejecting H_0, specifies a critical region R and the test "Reject H_0 if the evidence E is in R, otherwise accept H_0". Hence a simple definition of null-hypothesis is "the hypothesis which is initially accepted in constructing the test".

This leaves open the question of "how?"; how *does* one choose one's hypothesis, and how does one decide which to accept provisionally as H_0? The only short answer is "It all depends on circumstances". Choosing hypotheses is usually easy. Translate a nonmathematical request, e.g. "Does this machine need an overhaul?" into a mathematical one "Is this machine producing nails with mean length 2 cm? at least 2 cm? too many headless? (how many?)" and the answers "Yes" and "No" provide the hypotheses. Deciding which is null is harder and may be influenced by prior evidence, by which is the 'status quo', or even by the consequences of wrongly rejecting each hypothesis. For example, faced with a claim "This treatment produces significant weight loss", a civil servant advising a minister whether to prosecute under the Trade Descriptions Act would take the claim as his H_0, but the same man deciding whether to pay for the treatment for himself would take the opposite as H_0.

Since it depends so much on the individual case, any brief definition of 'null-hypothesis' is bound to be unsatisfactory, either in its vagueness or in being positively misleading in some circumstances (e.g. the 'skittle' view).

<div align="right">
Yours faithfully,
R. M. WHITEHOUSE
G. T. FIELDING
</div>

University of Keele, Staffs. ST5 5BG