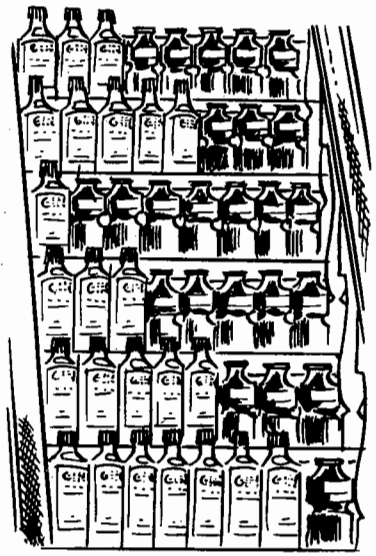

# SETTING DOSAGE LEVELS

**W. J. Dixon** *University of California, Los Angeles*

MANY OF the problems brought to the statistician for solution require pinpointing a level at which an expected response does or does not occur in order to test the strength and efficacy of drugs, pesticides, hormones, explosives, analgesics, stimulants, fuels, and a wide variety of other materials important to man and his environment. A statistician feels particularly successful if he develops a method for solving such problems that has a wide range of applications. One such method is called the *up-and-down,* or *staircase, method.* This is the way it works.

## HOW STRONG SHOULD PUNCH BE?

Mr. and Mrs. Smith, aged 22 and 20, respectively, are planning their first cocktail party, inviting all of the people in Mr. Smith's office. They can't





afford an elaborate party, but they do want to make a good impression, so they have decided to serve a punch made of gin and cranberry juice. They have little experience with alcohol, and so they don't know the proportion of gin to cranberry juice; they decide to try the punch first on some reliable and courageous friends. Four couples agree to help out.

Mr. Smith thinks all the guests will want at least 8 ounces of punch. Mrs. Smith wants her guests to be happy, but doesn't want anyone to become ill at her party. This is where the courageous friends (guinea pigs) come in. Mr. Smith mixes the first drink with 7 ounces of gin and 1 ounce of cranberry juice, and Mr. Big tosses it down. In about half an hour he is green and staggering. Obviously the punch contained too little cranberry juice. The next drink, 5 ounces of gin to 3 ounces of cranberry juice, goes to Mr. Jones. He is still on his feet half an hour later, but he is behaving strangely. Mrs. Big decides she will be safer if she volunteers for the 3-to-5 mixture. This turns out to be a mistake because she soon feels very warm. Mr. Smith refuses to let Mrs. Smith have a drink, and he believes that he himself must stay out of the testing to keep the record straight. The 1-to-7 drink, therefore, goes to Mr. Average who compliments his hostess on the delicious flavor, but says he really doesn't feel a thing. Mr. Small thinks he should have something stronger and asks for 5 ounces of gin to 3 ounces of cranberry juice. Mr. Small suddenly develops a slight speech defect, so the next drink is again made 3 to 5, and Mrs. Average drinks it. She compliments the hostess, saying the drink is excellent. She is happy and relaxed. Mrs. Small asks for a drink with the same proportions, and the Smiths thank their friends for having solved the problem. They will serve punch made from 3 parts gin and 5 parts cranberry juice.

The basic notion illustrated in this simple example is that of moving some important control level up or down each time, depending on the prior level and the outcome of the prior trial. In the example, when the response of the experimenters suggested too much alcohol, the dose was reduced for the next trial. When the response suggested too little alcohol, the dose was increased. In the end, some judgment was made as to final level. What sort of problems can be easily solved by this method?

## APPRAISING STRENGTHS OF OTHER MATERIALS

Some drugs are grown naturally, and they must be tested to determine how strong they are. Penicillin is an example. Most hormones must be similarly assayed for their strength.

(1) Pesticides should be strong enough to destroy insects, but not poison the family cat.

(2) Pain killers should relieve headaches, but not induce palpitations.

(3) Jet propulsion engines must have explosive-type "motors" capable of propelling an airplane, but the explosions must not shake the vehicle to pieces.

How can we design a measurement process that gives us sufficient assurance that we are arriving at a correct dose of a drug, one that will do what is desired? How can we do this in an efficient manner?

In order to test the strength of any given material, we must set up some standard of potency, or performance. With poisons, it is customary to use test animals of similar size and heredity and to inject each with a known concentration of the drug to be studied. A widely used, though arbitrary, choice is that the standard, or threshold, will be that concentration which will kill, on the average, half the animals tested. Obviously, other levels may be used, but if we understand how to handle one level, we are well along toward handling others. Let us examine in detail such a problem and one method for dealing with it.

Curare, a poison that paralyzes the heart and motor-nerve endings in striated muscle, was used for thousands of years by primitive people to destroy their enemies. Doctors now use this lethal substance for the benefit of mankind in certain surgical and medical procedures. Another poison with similar properties is the venom of the scorpion fish. To use this venom properly, we must have a precise measure of the strength of any batch we plan to use. Scorpion fish venom, in fact, has been assayed by the up-and-down method. This is how we went about setting up the trials.

## APPRAISING SCORPION FISH VENOM USING FIXED SAMPLE SIZES

An amount of the venom (the stimulus) will be injected into test animals (in this case, mice), and we will record whether a response (death) occurs in a given time period, say, 30 minutes. If we have chosen a dose that is too large, all (or almost all) the animals will die. If the dose is too small, possibly none of the animals will die. How can we find the dose that corresponds to the amount of poison that, if increased, causes more than half of the animals, on the average, to die and, if decreased, causes fewer than half to die? Even if he lives, the same animal cannot be used in a second test because he now may be less able or more able to stand the venom. The task would still not be very difficult if all animals behaved in the same way. Even in carefully selected animals, however, the amount of a drug required to bring about a response differs greatly from animal to animal. The amount of a drug just sufficient to cause a response in a particular animal is called his *threshold level.* We want to estimate some sort of average threshold for the population of animals.

Our measurement for a particular animal given a certain dose will be a response (in this case, death), which we shall record as "X," or a nonresponse, which will be recorded as "O."



How do we decide on the dose (stimulus) to give each animal? If an individual's threshold is unaffected by the test, we could merely increase a single animal's stimulus gradually until the threshold is reached. Unfortunately, this is not the case; estimating a mean threshold requires careful experimental design.

How can we attack this problem in an efficient manner? One natural design gives an experiment in which the same number of animals are tested at each of a variety of dose levels. To introduce some order into this situation we choose four different dose levels (say, 1, 2, 4, 8 mg) of venom. (It turns out that dosages of a wide variety of chemicals show approximately uniform increments in effectiveness if each dose is a certain percent larger than the preceding dose, i.e., if doses are chosen so that each is a multiple of the preceding one. In our example the doses increase by factors of 2.) We test five mice at each level. If we are so unlucky as to have picked a set of dosages that are all clearly below the threshold of all animals, we learn nothing about the location of the threshold for these animals except that it is greater than the largest dose given, 8 mg. We don't know how much greater. Or, if all animals at all levels respond, we discover only that the threshold is below the smallest dose, 1 mg. Even if the set of dosage levels chosen covers the general threshold level, we may test at a number of doses to which all or none of the animals respond.

Table 1 and Figure 1 show a set of outcomes from such a procedure. Five animals were tested at each of four dosage levels, with the outcomes as shown. Twenty animals were required.

## SEQUENTIAL TESTING: THE UP-AND-DOWN METHOD

It is only common sense to seek a testing strategy that leads the experimenter quickly to the proper levels for the tests. We wish to destroy as few mice

TABLE 1. Outcomes for Five Animals at Each Dosage Level in the Order in Which They Are Treated (X Means Death; O Means Survival)

| DOSAGE LEVELS | OUTCOMES | | | | |
|---|---|---|---|---|---|
| 8 mg | X | X | X | X | X |
| 4 mg | X | X | O | X | X |
| 2 mg | O | O | X | X | O |
| 1 mg | O | O | O | O | O |

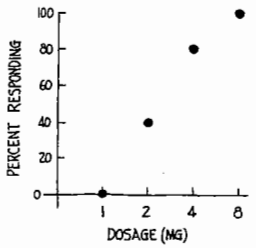


FIGURE 1

*Data of Table 1 plotted to show relation between response rate (percent dying) and dosage level*

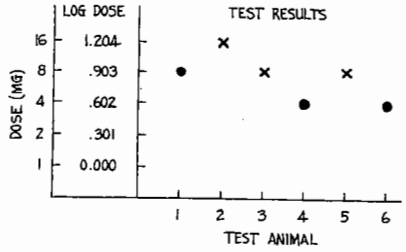


FIGURE 2

*Results of a sequence of six tests using the up-and-down method*

as possible, and since our supply of scorpion fish venom is limited we must try to get a good estimate from a small number of tests.

One design that has good properties is conducted one test at a time and consists of merely increasing the dose for the next animal if the last one tested did not respond to the dose administered and of decreasing the dose if the last animal did respond.

Let us examine the up-and-down version of our scorpion fish venom experiment in detail, limiting ourselves, for brevity, to animals of the same weight and a fixed venom concentration. For a particular concentration and animal weight, we proceed to test several animals at different dosages following our planned strategy.

We prepare to administer doses of sizes 1, 2, 4, 8, and 16 mg, and we begin by testing the first animal at any one of these levels. We choose, say, 8 mg for the first dose; the animal survives, and we record an O toward the left in Figure 2. The next animal is given the next higher dose of 16 mg. It does not survive; nor does the third animal (tested at 8 mg). In all, six animals are tested following this rule. From the data of Figure 2, because none of the 4 mg and two-thirds of the 8 mg doses produced a response, we might guess that the average threshold is somewhere between 4 mg and 8 mg. We would be uncertain, and we would not have a systematic way of making an estimate. A fairly precise estimate can be obtained, however, by averaging the levels at which the tests were done (in logarithmic units). Greater precision can be obtained by using a special table worked out mathematically to obtain the best estimate possible. By using this strategy, we can obtain about as much information from six animals as we could obtain by testing 20 animals in the design of Table 1 and Figure 1. The sequential character of the new approach tends to concentrate the doses where they are needed to get a good estimate.



The technique is also one of extreme simplicity to use. The sequence of trials in Figure 2 is completely described by indicating the sequence of O's and X's as they are observed and stating the dosage interval and the dosage administered on the last trial. For the example in Figure 2, the series is OXXOXO; the spacing of doses is log 2 = .301, and the final test dose was 4, which, in logarithmic units, is log 4 = .602. The average threshold dosage is estimated to be

$$.602 + k(.301)$$

where a value for $k$ may be obtained from a table in Dixon and Massey (1969) for the configuration OXXOXO. The value of $k$ is .831, so the estimate is

$$.602 + .831(.301) = .852$$

Because .852 is the logarithm of 7.11 we obtain 7.11 mg for our estimate of the average threshold dose.

## SUMMARY

We have illustrated the solution to a measurement or assay problem in which a special design of the sequence to be followed in the collection of the data allows us to get the resulting estimate with high efficiency. Other results of the statistical theory forming the basis of this procedure also show that this is about as good as we can do when we are able to observe only whether we exceed or fall short of the desired level. The theory shows that it requires only twice as many observations to obtain a threshold estimate of the same accuracy as would be obtained if we could measure precisely the exact dosage corresponding to each animal's own threshold. This, of course, is an impossibility for poisons, although we can imagine coming close to it in some problem where repeated measurements of the same animal are possible.

## PROBLEMS

1. Give three examples (not mentioned in the article) where appraising strength of material is needed.

2. In the 7th sentence of the example of appraising scorpion fish venom, explain the meaning of the words "on the average."

3. What is the threshold level of an animal?

4. How does this article define the average threshold level for the scorpion fish venom experiment? Would you use the same criterion to

define the threshold level when deciding the dosage level of curare to be used in medical procedures with human beings?

5. What are the advantages of a sequential testing procedure?

6. In Figure 2 we notice that the 1st animal survived when administered a dose of 8 mg. and the 2nd animal died when administered 16 mg. Why don't we conclude at this point that the threshold is between 8 and 16 mg.?

7. In the sequential up-and-down method of finding a dosage level, when does one stop taking samples?

8. Suppose we test 7 animals at the 5 dosage levels of venom described in the text, and obtain the series of responses XXOXOOX, with the final test dose being 8. Draw the results of the sequence of 7 tests similar to Figure 2. Estimate the average threshold, using the value $k = -1.237$ obtained from Dixon and Massey (1969).

9. Draw a graph similar to Figure 2 for the punch example. (Hint: use "parts of gin" as dosage level.) Can you estimate the average desired strength of punch by the up-and-down method described later in the text? Why or why not?

REFERENCES

J. J. Blum, R. Crease, D. J. Jenden, and N. W. Scholes. 1957. "The Mechanism of Action of Ryanodine on Skeletal Muscle." *Journal of Pharmacology and Experimental Therapeutics* 111:477–486.

Lincoln P. Brower, William N. Ryerson, Lorna L. Coppinger, and Susan C. Glazier. 1968. "Ecological Chemistry and the Palatability Spectrum." *Science* 161:1349–1351.

W. J. Dixon and F. J. Massey, Jr. 1969. *Introduction to Statistical Analysis,* Third Edition. New York: McGraw-Hill.