Testing Stastical Hypotheses Concerning Test Keys

Bonnie H. Litwiller and David R. Duncan

Contemporary Mathematics teachers are encouraged to incorporate probabilistic concepts into their Mathematics classrooms at all levels. To do this, teachers need examples. We shall discuss one such probabilistic example and an extension of it.

Suppose that a teacher, Lloyd, has a key for a machine-scored 50-question multiple-choice test. Lloyd did not put any label on this key so that a student who might see it would not know which test it was for.

Before Lloyd sends the student tests to be graded, he wishes to check to see that he has the correct key for the test. He doesn't have time to check every question. How many questions does he need to check in order to be "relatively" sure that he does indeed have the correct key? His task is not to double check the key to see if he has solved the problems correctly; this has been done earlier. His task is to verify that he has the correct key for the test that he is grading.

Assume that in constructing a multiple-choice test, Lloyd randomly places the correct response among the five "foils" in each question. Consequently, the probability that the key will have the correct response to a given question by chance alone is 0.2. Since the responses to the questions are independent, the probability that a random sample of r questions are all answered correctly on the key by chance alone is $(0.2)^r$.

Lloyd begins with the pessimistic hypothesis that the key is, in fact, for a different test. He then begins to check answers to random questions. Since he knows that the key is correct for *some* test, a single wrong answer will convince him that this is the wrong key.

We list the consecutive powers of 0.2 :

$$\begin{aligned}
(0.2)^1 &= 0.2\\
(0.2)^2 &= 0.04\\
(0.2)^3 &= 0.008\\
(0.2)^4 &= 0.0016\\
(0.2)^5 &= 0.00032\\
(0.2)^6 &= 0.000064
\end{aligned}$$

Since the probability of six consecutive matches with the wrong key is 0.000064 or 0.0064% (less than of $\frac{1}{100}$ of 1%), Lloyd concludes that the probability that he has the wrong key is remote enough for him to reject that possibility.

Let us extend the problem by supposing that Lloyd's test was true/false rather than a five-foil multiple-choice test. We will retain all other assumptions, including the randomness of correct responses. How many questions must Lloyd check in this situation? Again, Lloyd begins with the pessimistic hypothesis that he has the wrong key. A single incorrect response will confirm this hypothesis. How many consecutive matches are needed for Lloyd to reject this hypothesis? Again, Lloyd will reject the hypothesis if the probability of a certain number of consecutive matches by chance alone is less than of $\frac{1}{100}$ of 1%.

The probabilities of consecutive matches are as follows:

$$\begin{aligned}
(0.5)^1 &= 0.5\\
(0.5)^2 &= 0.25\\
(0.5)^3 &= 0.125\\
(0.5)^4 &= 0.0625\\
(0.5)^5 &= 0.03125\\
(0.5)^6 &= 0.015625\\
(0.5)^7 &= 0.007812\\
(0.5)^8 &= 0.003906\\
(0.5)^9 &= 0.001953\\
(0.5)^{10} &= 0.000977\\
(0.5)^{11} &= 0.000488\\
(0.5)^{12} &= 0.000244\\
(0.5)^{13} &= 0.000122\\
(0.5)^{14} &= 0.000061
\end{aligned}$$

If the key matches the test for 14 consecutive randomly selected questions, Lloyd can conclude that the probability of this occurring by chance alone is less than of $\frac{1}{100}$ of 1%. Would your students think that he made a wise decision?

An important statistical concept is the establishment of a *null hypothesis* that is to be rejected when the probability of certain observed events is less than some specified probability (0.01% in this case). This problem provides students with experience in such situations. Such early familiarity with the notion of hypothesis testing is invaluable to students who will be working in Mathematics or some area of research (e.g., Chemistry, Biology, Environmental Science, Psychology, etc.).

Department of Mathematics
University of Northern Iowa
Cedar Falls, IA 50614-0506