SCI199Y: November 21, 1995

Required for next week

Reading "Statistical evidence of cheating on multiple choice tests." S.P. Klein (1992), *Chance* **5** (3-4), 23–27. and " The use of statistical evidence in allegations of exam cheating." P.J. Boland and M. Proschan (1990), *Chance* **3** (3), 10–14.

Your task Come prepared to discuss these articles, and to ask questions about the parts you didn't understand.

Technical note: testing hypotheses

An extremely common form of statistical inference is an "hypothesis test", or a "significance test".[1] It is used to answer the question "Are these data consistent with ——— ?", where ——— is a formulation of an hypothesis. Usually, the formulation will be a mathematical one, relating to some chance mechansism, i.e. probability model, that may have generated the data.

In the context of the streak shooting article *TG*, the hypothesis that the authors considered was that the sequence of hits and misses for a given player was just like a sequence of independent coin tosses, with the probability of a hit equal to the player's overall hit rate. In other words, their hypothesis was "there is no streak shooting".

Their hypothesis testing procedure is outlined in the last paragraph on p.18 and the first two columns on p.19. For each player in Table 2, they took strings of four shots, and counted for each string whether it had 3 or 4 hits (high), 2 hits (moderate), or 0 or 1 hit (low). *If successive shots are like coin tosses*, then the probabilities for high, moderate, and low can be easily calculated for each player. For Clint Richardson, who has an overall hit rate of 0.50, these probabilities are 5/16, 6/16, 5/16, respectively. They don't give us the data to reproduce their calculations, but what they did next was to compare the observed counts of high, medium and low to these predictions. Their conclusion was, yes, the data is consistent with their hypothesis that successive shots are like coin tosses. They carried out this hypothesis test for each of the 9 players in Table 2, and for all the players combined. Here's what they found for all the players combined:[2]

	high	medium	low
observed	33.5%	39.4%	27.1%
expected	34.4%	36.8%	28.8%

[1] There is in fact a technical difference between hypothesis tests and significance tests, but we needn't worry about that here.

[2] There's more work to be done once this table is constructed; the actual test needs to be carried out. They used something called a 'chi-squared test'. More later.

It might seem odd to use as a hypothesis that there is no streak shooting, rather than, for example, to hypothesize that there is streak shooting. It is very common to test such 'non-event' hypotheses; so common that the hypothesis being tested is called the null hypothesis. There are two reasons for this. The first reason is pragmatic: it's pretty simple to construct a probability model for the no streak shooting assumption (coin tossing). It's quite a bit more difficult to construct a probability model for streak shooting. The second reason is caution. Until the data provide strong evidence that an alternative explanation is needed, it is usually wiser to stay with the simpler and better understood explanation.

If researchers have a particular alternative model in mind, that is undoubtedly the correct model if the date are not consistent with the null hypothesis, they are in a much stronger position. They can then see whether the data provide stronger evidence for the null hypothesis model or the alternative hypothesis. What is more often the case is that a very well specified alternative model is not available, but a general idea about some characteristics of possible alternative models are available. In this case, statisticians will usually try to test the null hypothesis in a way that is particularly sensitive to these characteristics. The technical term for the sensitivity of an hypothesis test is "the power of the test".

Reference

TG The cold facts about the "hot hand" in basketball. A. Tversky and T. Gilovich (1989) *Chance* **2** (1) 16–21. (handout on November 7 for November 14)