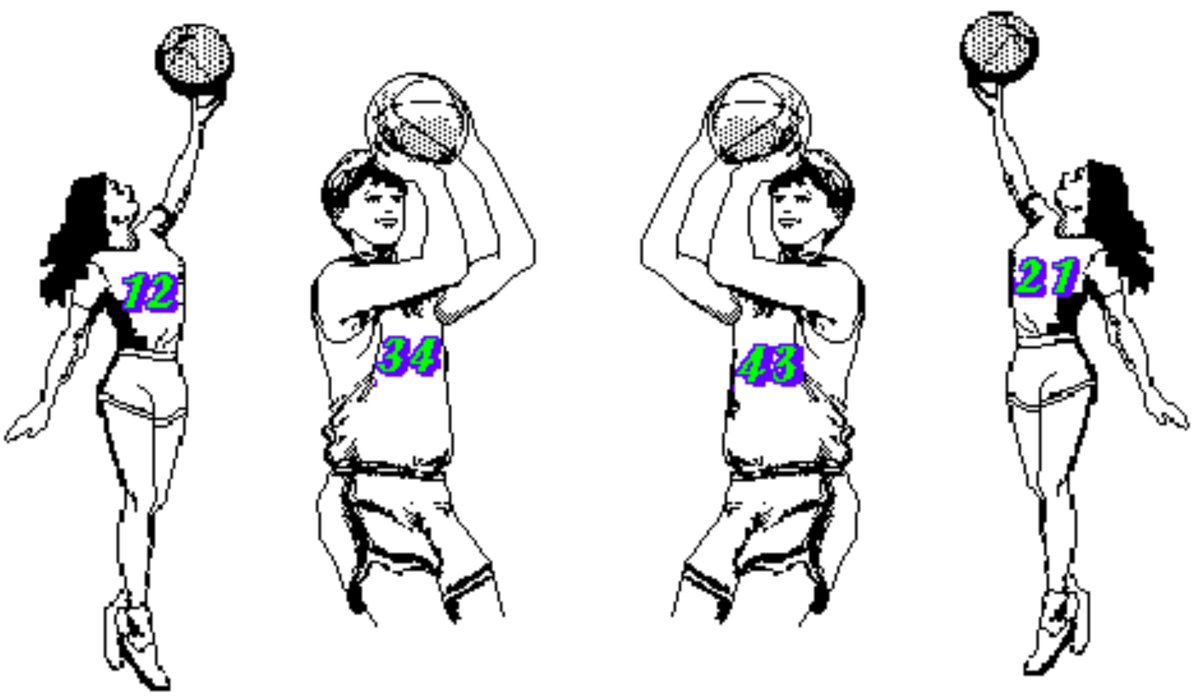
THE "HOT–HAND" THEORY IN BASKETBALL:

A MODEL PROJECT FOR STATISTICS STUDENTS

By the "A–Team:"



Rexanne Bruno, John Reay, Eric Stade, Pat Tarrant

§l. Introduction.

The availability of powerful and inexpensive statistical software, together with the vast information resources provided by the electronic superhighway, afford the modern statistics student an opportunity to go beyond the drudgery of "canned" problems, and to work directly with real data sets relevant to contemporary issues and situations. The resulting trend toward more exploratory data analysis in elementary statistics classes (for example, see [1] and [4] and their references) requires teachers of these classes to provide suggestions of accessible, interesting student projects. We will describe what we believe to be such a project, provide some preliminary data and conclusions, and give references for further student reading.

Namely, we address the "hot–hand" theory in basketball. According to this conjecture, a player has a better chance of hitting a shot after a hit than after a miss. Of course, the phrase "better chance" is subject to interpretation, and already in the literature several different takes on this phrase have been considered: one may study only free throws for an individual player, or field goals for an individual player or for a team, or attempts that are limited to one quarter of continuous playing, or attempts that may span a whole game or different games over several days (see [2], [3], and [5] and their references). We will consider only sequences of field goals of individual players, each sequence within a single game—that is, while a given player may be observed over the course of several games, his sequence of shots in any one of them is not considered as continuing or being continued by his sequence in any other. (Students should be urged to consider carefully the various alternatives, such as those just described, possible for the data–collection component of the project. While the issue of "best possible project design" can be somewhat subjective and nebulous, still such considerations can go a long way to shedding light both on the question posed and the utility of statistics for providing an answer.)

The data that we use here was gathered during the (June) 1995 NBA playoffs, while the authors were attending a STATS (Statistical Thinking And Teaching Statistics) workshop at Colorado State University. The various analyses presented were performed using the Minitab (for Windows) statistical software package. (Some of the actual output from Minitab has been reformatted here, for clarity of presentation.)

§2. Methods and Results.

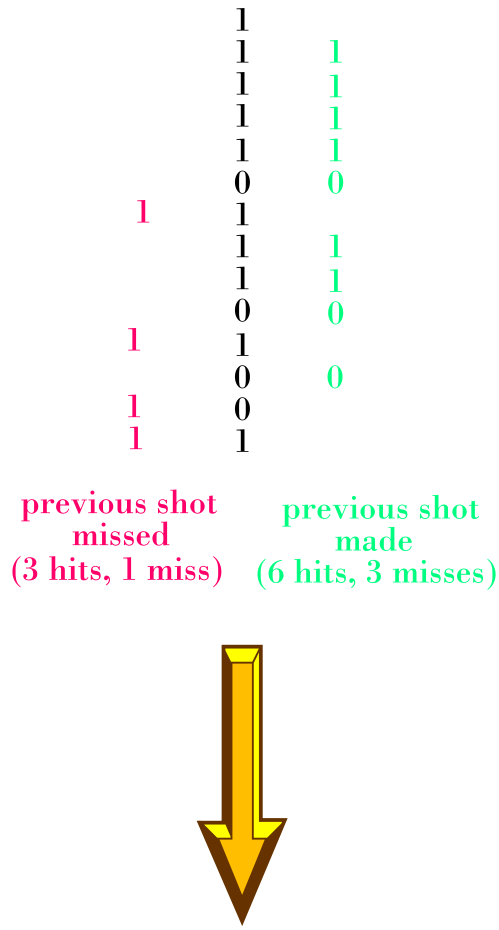
When this project was undertaken, the NBA championship games were into the semifinals: the four teams left competing were the Houston Rockets, the Orlando Magic, the Indiana Pacers, and the San Antonio Spurs. Our project team—hereafter referred to as the "A–team"—gathered data from two of these semifinal games (due to time conflicts and constraints, we were unable to observe more. The real world obviously places limitations on experimental design and execution; a first–hand encounter with such limitations would undoubtedly enhance the education of any statistics student).

A data collection scheme was devised, as follows: First, a list was made of players on the above teams who had taken at least 75 shots in recent playoff games. (This was in an attempt to ensure that shot sequences of statistically significant lengths could be observed.) Next, members of this list were randomly assigned to members of the A–team, who were to track shot sequences of their designated players. If a player assigned to an A–team member went out of the game, or did not play in the first place, the A–team member was to start tracking the shots of the next available player on the list. In this way, we ultimately ended up with data for fourteen players.

Entire sequences of each selected player's field goal attempts (two and three point shots) were recorded, noting whether each shot was made or missed. The raw data for each player thus consisted of a string of 0's (missed shots) and 1's (made shots). An asterisk ("*") in a string was used to indicate that a player left the game, and upon his return a new sequence, not considered contiguous with old one, was started for the player. This was done to avoid the possibility that an A–team member might miss the actual point at which a player returned, and thus might create spurious relationships between shots.

Table 1 below illustrates how the above raw data was converted into previous/current hit/miss data for a given player's sequence. The data for all such sequences was then combined to create a 2×2 contingency table—see Table 2 below.

Shot sequence for Shaquille O’Neal (jersey #32, Orlando Magic):



CURRENT SHOT

PREVIOUS SHOT	missed	made	totals
missed	1	3	4
made	3	6	9
totals	4	9	13

Table 1. Conversion of raw data to contingency data

PREVIOUS SHOT \ CURRENT SHOT	missed	made	totals
missed	18	25	43
made	23	27	50
totals	41	52	93

Table 2. 2×2 contingency table of combined shot–sequence data

A Chi–squared test of the null hypothesis—that making the current shot is independent of making the previous shot—was conducted. There was insufficient evidence of a relationship between the current and previous shots ($\chi^2 = 0.161$, p–value $= 0.689$, $df = 1$). An approximate randomization test of the equivalent hypothesis

$$H_0 : p_1 = p_2,$$

where p_1 is the proportion of successful shots in the population of all shots in which the previous one was made, and p_2 is the proportion of successful shots in the population of all shots in which the previous one was missed, was also conducted using the same data. (The alternative hypothesis here is the two–tailed one, $p_1 \neq p_2$.) For this test, no assumptions were necessary regarding the distribution of the population nor the randomness of the sample. Similar results were obtained ($z = -0.005$, p–value $= 1.00$).

Finally, a runs test was conducted on each of the three longest observed sequences: one of 17 consecutive shots by David Robinson of the Spurs; one of 14 consecutive shots by Shaquille O'Neal of the Magic; and one of 12 consecutive shots by Hakeem Olajuwon of the Rockets. The objective of such a test is to detect an unusually high, or low, number of "runs" in the data, a run being (in our case) a sequence of one or more consecutive 0's or consecutive 1's. Again the conclusion was the same: there was no statistically significant dependence of one shot on the previous one. The output from one of these runs tests is presented in Table 3 below.

Runs Test: David Robinson (jersey #50, San Antonio Spurs)

Raw data:

0 0 1 0 0 0 0 1 1 1 0 1 1 0 0 0 0

$K = 0.3529$

The observed number of runs = 7

The expected number of runs = 8.7647

6 observations above K; 11 below

The test is significant at 0.3303

Cannot reject at $\alpha = 0.05$

Table 3. A runs test

§3. Conclusions.

Even to quite permissive levels of statistical significance, our data do not indicate a dependence of the success of a given shot on the outcome of the previous one. This result is consistent with much of the literature, although some authors have argued that there *is* evidence of such a dependence (see the references). (Perhaps the lesson to be learned by this apparent discrepancy is that, before using statistics to answer a question, one should make sure that one has *asked* the question precisely. Again, we found that this particular project led to many discussions that, while open–ended, were still quite educational.)

We also had an awful lot of fun with this project. We need not elaborate further on the important role of the fun factor in the modern undergraduate experience.

§4. Teaching Research Habits.

Beyond the gathering and analyzing of data, and the drawing of conclusions from the same, there are many ideas to be gleaned from the process of completing the above (or a similar) project. First of all, students should be aware of the references [2], [3], and [5], where others have analyzed the hot–hand problem. We think it best that students be asked to look up and retrieve the articles themselves, rather than receiving photocopies or the like. We hope that, by the time they are nearing completion of the project, they would be eager to find what others have written; moreover, beginning (and other) students will learn from this experience of looking up related research articles.

Further, the students might be asked to find other articles on the hot–hand problem. Learning which library tools might be available for researching a particular topic, and how to access and use these tools, is always a useful endeavor. The students might find and report on articles of which the professor is not aware. (No claim is made, for example, to the completeness of the references given in this article!) The time to teach useful skills is when there is a genuine interest on the part of the students.

Finally, the students might be asked to explain why and how different researchers can come up with opposite conclusions, when looking at (for example) the hot–hand problem. The students should clearly report on whether these apparently conflicting conclusions are due to statistical flukes in the various observed data sets, or to the measuring of different quantities, etc. Questions of this type should lead the students to reconsider their own analyses, and to determine exactly what questions they tried to answer, and what alternative, related questions could have been asked.

In sum, it's never too early to teach good research habits.

References

1. George Cobb, Teaching Statistics, MAA Notes #22, Heeding the Call For Change: Suggestions for Curricular Action, L. Steen Editor (1992) 3-34

2. Ronald Forthofer, Streak Shooter—the Sequel, Chance, vol. 4, no. 2 (1991) 46-48

3. Amos Tversky and Thomas Golovich, The Cold Facts About the "Hot Hand" in Basketball, Chance, vol. 2, no. 2 (1989) 16-21

4. Paul Velleman and David Heaglin, Data Analysis, MAA Notes #21, Chapter 2, Perspectives on Contemporary Statistics, D. Hoaglin and D. Moore, Editors (1992) 19-37

5. Robert Wardrop, Simpson's Paradox and the Hot Hand in Basketball, Amer. Statistician, vol. 49, no.1 (1995) 24-28