

# The Rice Honor Code, Athletes, and Random Chance

## A Note to the Rice Board of Trustees and the Rice Community

Marj Corcoran  
April 28, 2004

At the focus group meeting with McKinsey, the question arose as to whether or not athletes are more likely to be involved in honor code violations than other students. Some of us were taken aback by the fact that the McKinsey people seemed to be dismissing the honor code issues as not statistically significant and not important. There are two separate issues as I see it, the first being the statistical significance of honor code convictions among athletes, and the second being the importance of honor code issues.

The first question is straightforward to formulate and to answer. The question is this: "Is the number of athletes involved in honor code violations consistent with random chance, given the percentage of athletes in the student body?" The statistical distribution that governs this sort of process is the binomial distribution, the same distribution that governs coin tosses. I attach a note written by my colleague Steve Baker that gives some details of a calculation for honor code violations from academic year 1992-1993 to 2002-2003. During that period, there were 149 honor code convictions, of which 45 were athletes. During that period, athletes comprised 11% of the student body.

Steve's calculation shows that, for this number of students, the probability that a population comprising 11% of the total will randomly fluctuate to 30% or more is about 5 parts per billion. To visualize what this means, imagine drawing balls from a hopper, as is done in lottery drawings. Imagine a hopper with a large number of balls, 11% of which are green (representing athletes) and 89% of which are blue (representing non-athletes). Then draw 149 balls from the hopper randomly and count the number of green balls (athletes). If you repeat this experiment many times, the number of athletes you draw will vary due to random fluctuations. The question is, how large are those fluctuations?

The plot below shows a computer simulation of exactly this process. Because I have done it on a computer, I can repeat it millions of times in just a few minutes; in fact for this plot I have repeated the experiment 10 million times. We would expect the mean of this distribution to be around 16, since 11% of 149 is 16.4, and we would also expect to see random fluctuations about this mean. In the distribution below we see that the mean is indeed close to 16, and we see that fluctuations do take place. For this set of 10 million random experiments, the largest fluctuation produces one entry at 40.

In many fields, a result which is 2 or 3 standard deviations from the mean is considered statistically significant. In my field (particle physics) we require a significance of 4-5 standard deviations to claim a discovery. In the plot below, the one data point that we have is the actual number of Rice athletes who were convicted of honor code violations over this 10 year period. The arrow on the plot shows that point, which is more than 7 standard deviations from the mean. This one data point is clearly inconsistent with random fluctuations.

I don't want to paint all athletes with the same brush. I do not mean to imply that all athletes, or

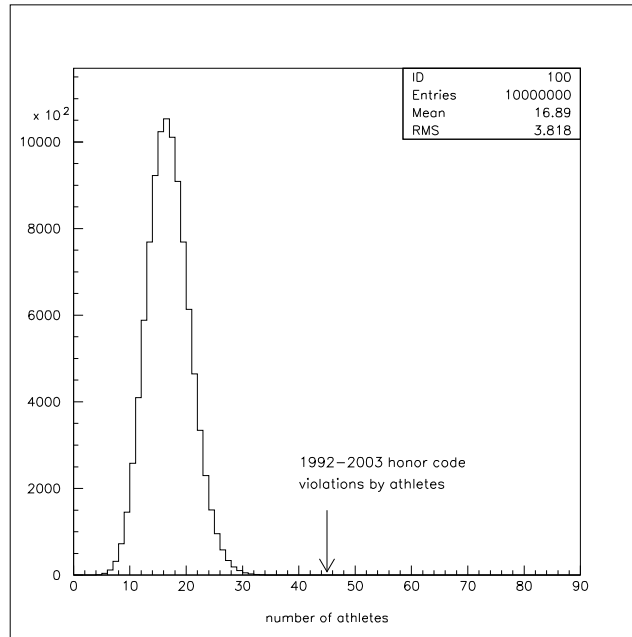


Figure 1: *The number of students who are athletes in a group of 149 students, drawn randomly from a student population which is comprised of 11% athletes. Each entry in the histogram represents one trial. There are a total of 10 million trials. The arrow shows the actual Rice data point for honor code convictions by athletes, from the 1992-1993 academic year to the 2002-2003 academic year.*

even a significant fraction of them, violate the honor code. But among the students who are convicted of honor code violations, athletes comprise a disproportionately large fraction.

On the issue of the importance of the honor code to Rice, I wanted to make two comments. The first is in regard to collaborative cheating, which athletes have been involved in from time-to-time. To my knowledge, all of the collaborative cheating scandals that have occurred over the past 20 or more years have involved athletes. At the focus group I attended, the McKinsey people were minimizing the importance of these occurrences, referring to them as “a single event”. This is completely wrong. Collaborative cheating is not “a single event”. If 45 students are involved, it is 45 people who cheated, 45 “events”. But it is much worse than that. One of the underpinnings of the honor code is the fact that it is set up and enforced entirely by the students. Students are expected to abide by the honor code themselves, but they are also expected to report violations by other students. Collaborative cheating strikes at the very foundation of the honor code. Without student enforcement, the honor code breaks down.

My second comment is in regard to the importance of the honor code to the Rice community. To me as a faculty member, the honor code is central to the Rice culture. It fosters a sense of trust between faculty and students. Our students are our partners in learning, not our adversaries. I don't have to look over the shoulder of my students to see if they are cheating; I know that they are not. Collaborative cheating by tens of students threatens the honor system; its seriousness should not be minimized.