ON THE QUESTION OF WHETHER THE RATE OF HONOR CODE VIOLATIONS BY ATHLETES IS STATISTICALLY SIGNIFICANTLY HIGHER THAN THAT OF OTHER UNDERGRADUATES

Stephen Baker

April 24, 2004

In the statistics on Honor Code violations violations (Table 9 of the Faculty Council Report on Athletics) we see generally a much higher rate of violations by athletes than by regular students. For instance, let us first consider the eleven-year period 1992-1993 through 2002-2003. Although athletes comprised about 11% of the student body, they accounted for about 30% of the Honor Code violations (45 violations by athletes out of a total of 149 for the entire undergraduate student body). This is a rate of violations by athletes that is about 3.5 times higher than that of non-athletes (45/11%)/(104/89%).

Is this difference in rate statistically significant? One might have hoped that, out of 149 violations, one would have found about 11% of the violations were by athletes, that is, about 16. So, how likely is it that 45, or more, violations by athletes, instead of 16, would occur by mere chance? The answer is that it is not very likely at all, with a cumulative probability of about 2×10^{-10} (or, about 1 chance in 5 billion) of being 45 or more. One can therefore draw the conclusion that, yes, the rate of Honor Code violations among athletes is statistically significantly *much* higher than the rate non-athletes. The calculation of the cumulative probability is described in the Appendix, below.

It should also be pointed out that the period chosen for the analysis above, 1992-2003, excludes the year immediately preceding, when there was a much larger than usual number of Honor Code violations by athletes than in typical years: 45 in that year alone (coincidentally, the same number as in the following 11 years combined), of which 28 were concentrated in one case. Including this year in the sample to be considered would only *reduce* the already vanishingly small probability that the higher rate of Honor Code violations among athletes could be accounted for by chance alone.

APPENDIX: In a population of N students, of which A are athletes, the probability P in a random sample of n students, of finding *a* athletes is given by the expression:

$$P(N,A,n;a) = [n!/((n-a)!a!)][(N-n)!/((N-A-n+a)!(A-a)!)]/[N!/((N-A)!A!)]$$

To obtain the number referred to in the text above, N=29,700 (the number of student years in the period 1992-2003), A=3267 (the number of athlete years in the same period), and n=149 (the total number of Honor Code violations in that period). P is then calculated for values of *a* from 0 to 149. The cumulative probability that *a* is equal to or greater than 45 is obtained by adding the probabilities for values of *a* equal to 45, 46, 47, etc., up to $a = 149$.