

## The Three-Year-Ago Correlation

by Paul Conlin

In response to your question about baseball statistics:

If you look at victories by teams from 1901 to the present, and calculate the correlation to the number of wins that same team had 1 year ago, 2 years ago, 3 years ago, 4 years ago... 10 years ago, to test which is the most predictive of current year wins; I believe you will find (if I did it correctly) that 3 years ago has the highest correlation. I was surprised that 3 years ago was higher correlated than previous year or 2 years ago.

---

### *Bill James' response:*

This would be a very interesting phenomenon if it were true. I studied the issue, and was unable to substantiate your conclusion.

In response to your query, I figured the average year-to-year change in team wins, and in team winning percentage, for all major league teams, 1901-2001. I found that the average change from one year to the next was 10.19 wins, and .059 in winning percentage. In my study, this figure increased as the distance in years increased:

<i>Time</i>	<i>Wins</i>	<i>Percentage</i>
1 year	10.19	.059
2 years	11.82	.069
3 years	12.37	.073
4 years	12.92	.076
5 years	13.70	.081

6 years	14.24	.084
7 years	14.62	.087
8 years	14.69	.088
9 years	14.78	.089
10 years	15.00	.091

You used the expressions “highest correlation” and “to test which is most predictive.” I haven’t actually figured a correlation here, exactly, and the term “predictive” also does not really apply to what I have done. Maybe I’m not understanding what you are saying. But . . . as nearly as I can see, it seems to me that a team’s won-lost record is more closely tied to last year’s won-lost record than it is to any previous year’s.