YARDS, POINTS AND WINS

By Pete Palmer

In order to relate points scored and allowed to wins and losses, data for all teams from 1970 through 1980 was accumulated. The technique used, regression analysis, is a way to determine the relationship between one or more independent variables (such as points for or points against) and a dependent variable (wins).

For example, if one wanted to assign a value to points scored and another to points allowed in order to predict wins over .500 for two teams, this would be a situation with two equations and two unknowns, and the answer could be found exactly. However, this answer might not work very well when applied to other teams. The regression method mathematically minimizes the overall error in the prediction of the result for a large number of teams. This prediction is obtained by summing the products of anywhere from one to many variables, each multiplied by a factor determined by the analysis. The error is defined as the sum of the squared differences between predicted and actual results for each member of the sample.

One very important thing to remember with regression is that variables used are often not really independent, and that incorrect conclusions can be reached from the results. In football, points scored and allowed are really not independent because of the interaction between offense and defense. Also, mathematical correlation does not necessarily mean a cause-and-effect relationship.

Taking all 296 team-seasons in one group, the regression analysis was used to solve this equation:

(Points For)/A + (Points Against)/B = Wins + (Ties)/2 - (Games)/2

The values obtained were 36.9 for A and 37.3 for B. This means that on the average, increasing a team's points by thirty-seven over the season (or taking the same amount away from the points allowed) would result in one more win. There was a slightly negative correlation between points scored and allowed, meaning that teams with more points than average tended to allow fewer points than average.

To check this relationship for various subsets of the teams, two tests were run with the clubs divided into two subgroups. In the first test, all clubs averaging above thirty-eight points per game for both teams were compared against those below that figure. The values for A and B for the high-scoring teams were 37.6 and 37.0, while for the low-scoring ones, they were 35.3 and 38.4.

The teams were again divided into subgroups, this time by average point differential per game. The break point was five points per contest. The results were 39.5 and 36.3 for the high difference teams and 29.2 and 31.7 for the others. This indicates that overall points are relatively unimportant in determining the relationship between points and wins, but that a large spread in points does not produce quite as big a swing in wins and losses as might be expected from the smaller differences. This could be because a team might have been involved in a few very one-sided games which make the overall point differential look more extreme than it should. The conclusion is that it takes about thirty points to produce an extra win for average teams, but that this figure could be as much as forty for particularly strong or weak teams.

Since the figures for A and B in the equation were relatively close for each of the tests, further runs were made in which only one independent variable was used, points scored minus points allowed. In each case, the value obtained was midway between the two found in the first study, and the overall error was virtually the same. Tests were also run in which the independent variable was chosen as the point differential divided by the total number of points scored by both teams, and the differential divided by the square root of total points. In each case, the prediction error was slightly larger than in the first two cases. Dividing by total points squared gave prediction errors a fair amount larger than other methods.

Taken independently, points allowed correlated slightly better with wins than points scored, but the difference was negligible.

* * *

In order to relate team performance to points, several tests were made on offensive and defensive data for all clubs from 1975 through 1980.

The basic unit chosen was the number of points scored or allowed per possession, or drive. The drive unit was selected because throughout most of the game the team's main objective is to score on each possession, regardless of the number of plays required. Since the exact figures were not available, the total number of drives was estimated to be equal to the sum of the scoring plays by the opponents plus their missed field goals, punts and turnovers. No records were available for the number of times the ball was lost on downs. One drive per game per team was added to cover the opening kickoff for each half and one was taken away for each touchdown on a return.

The average number of drives per team per game was twelve and a half, with a range of eleven to fourteen. Forty-two per cent of these ended in punts, thirteen per cent in field goal attempts (of which sixty-two per cent were successful), twenty per cent in turnovers (fifty-five per cent being intercepted passes), and seventeen per cent in touchdowns. Eight per cent were stopped by the clock.

Each drive averaged twenty-four yards from scrimmage, with a range of about eighteen to thirty, and resulted in one to two points. Thus a difference of twelve yards led to one extra point. Correlation of yards per drive to points was quite high, and adding kick and return yardage to the calculation improved the correlation only slightly.

Since offense and defense are dependent on each other, yards per drive on offense and yards allowed per drive on defense were correlated against points scored. The number of yards per point for offense was again about twelve, but defense had very little effect. It appeared that about sixty yards on defense were needed to change the offense by one point. Correlating offense and defense to defense produced similar figures. The explanation here is that as extra yards are gained, the ability to push the other team back is reduced by the closeness of the goal line, meaning that the drive might end in a missed field goal or touchback.

Frequencies of individual plays, both from scrimmage and on returns, produced little correlation with points. Average yards per play, which showed virtually no correlation for kicks or returns, was quite high for yards per pass attempt, a fair amount lower for yards per rush. Yards per play from scrimmage was slightly better than yards per pass attempt, but not as high as the yards per drive measure already developed.

An attempt was made to find time of possession from the data available. All plays were divided into two types, long and short. Long plays were rushes, sacks, and completed passes, less touchdowns and fumbles lost. Short plays were incomplete and intercepted passes, fumbles lost, field goals made or missed, punts and their returns, touchdowns from scrimmage and their resulting kickoffs and returns, and penalties. The average time derived per long play was thirty-one seconds, while nine seconds was calculated for each short play. Overall team minutes for the season were found to be within about fifteen minutes of the expected value using this method. Dividing the above categories into their components did not improve the correlation.

* * *

The relationships between rushing and passing and of fumbles and interceptions to winning and losing were studied using play-by-play data.

It has often been claimed that the way to win a football game is to establish a running attack. Figures have been quoted showing how often a team that rushes at least forty times ends up on top. My theory has always been that rushing and passing are equally important. It is true that the club that runs most is most likely to win, but that is because the team builds a lead from a balanced attack and then protects it by emphasizing the ground game. Meanwhile, the opposition is forced to pass in order to catch up in the decreasing time remaining.

To illustrate this point, data was taken from New England Patriot games from 1978 through 1980, forty-nine contests in all. In the first half, the eventual winner rushed fifty-seven per cent of the time, compared to fifty-three per cent for the loser – not much difference. But in the second half, the victor's ground percentage increased to sixty-five per cent, while the opponent's shrunk to just thirty-eight per cent. There were eighteen times in which the winner had built at least a ten-point lead at the midway break. In these, the second half percentages were seventy-one and thirty-two, respectively. The average lead in these games was eighteen points at the half and only twenty-one overall. There was one additional game in

which the club ahead by ten or more at halftime did not win, so the conservative strategy worked most of the time.

In the remaining thirty-one games, the eventual winner was equally likely to be ahead or behind at the half, but had an average lead of eight by the final gun.

In thirty-two of the games, a team rushed at least forty times, and twenty-nine were victories. However, looking at the first half, there were thirty-one cases in which one team rushed twenty or more times, and only seventeen were by the winner. In the second half, there were forty-four twenty-rush situations, thirty-nine by the victors. So the conclusion is that winning leads to rushing, not the other way around.

* * *

Another figure often quoted is how interceptions lead to defeat, but in reality defeat is more likely to lead to interceptions. In the games analyzed the losers outfumbled the winners fifty-four to forty-six, but they were outintercepted ninety-four to fifty-one. On the surface, this looks as though fumbles are relatively unimportant and interceptions are disastrous. However, if the number of running and passing plays are taken into account, the fumble and interception frequencies are each about fifty per cent higher for the losers than for the winners.

Approximately two-thirds of the losers' interceptions occurred in the second half, often when the team was already behind and had little chance of winning. Data from a play-by-play study of games in 1970 showed that the team was one point behind on the average when a fumble occurred, but for interceptions the average deficit was five points. The conclusion is that fumbles and interceptions are equally damaging to a team's chances of victory.