



## Football Leagues



In [Football Leagues](#) we have animated bar charts of the league tables of football leagues from across Europe over the past ten years. See [Premier League](#) for a discussion of luck in the distribution of points in the Premier League, and check out our predictions for the final week of the Premier League 2008-2009 in [One game to play!](#)

See [Premier League](#) and [One game to play!](#) Football data extracted from [Football-Data.co.uk](#). The beautiful flag icons were created by Mark James of [famfamfam.com](#).

You need Flash Player 9 or above to view this page.

### What does the animation represent?

#### The main bar chart

The main bar chart describes the number of points that each team has after each match of the season. A single step of the animation corresponds to a match between two teams in which a winning team scores three points, a drawing team scores one point, and a losing team scores no points.

By default, teams are listed alphabetically. Clicking on 'Sort' causes the teams to be listed in order of points gained. If two teams have the same number of points then they are ordered first by goal difference, second by goals scored, and third alphabetically.

#### The point distribution graph

Click on the 'Point distribution' button and a grey column chart appears at the bottom of the screen. The horizontal axis of this column chart coincides with the horizontal axis of the bar chart; both measure the number of points scored. The height of a column in the point distribution graph records the number of teams with the corresponding amount of points. Thus a column with horizontal position 22, and height 3, indicates that, at that moment in time, precisely three teams have 22 points.

#### The chance distribution graph

Click on the 'Chance distribution' button and a white column chart appears at the bottom of the screen. Again, the horizontal axis of this column chart represents the number of points scored. This time, however, the heights of the columns represent the values we would expect if each game were determined by chance alone. We now explain this statement in more detail.

Consider a particular league of  $n$  teams, and  $N$  matches in total, for which there are in all  $W_H$  home wins,  $W_A$  away wins, and  $N - W_H - W_A$  draws. Suppose that the  $n$  teams are equal in ability, and the outcome of each match is independent, and random. This means that, in any particular game, the probability  $p_H$  of a home win, the probability  $p_A$  of an away win, and the probability  $p_D$  of a draw are given by

$$p_H = \frac{W_H}{N}, \quad p_A = \frac{W_A}{N}, \quad p_D = \frac{N - W_H - W_A}{N}.$$

Let the random variable  $X_H$  represent the number of points gained by a particular team playing in a single home match, and let  $m_H$  and  $v_H$  be the mean and variance of  $X_H$ . Then

$$m_H = 3p_H + p_D, \quad v_H = 9p_H + p_D - m_H^2.$$

Likewise if  $X_A$  represents the number of points gained by a team in an away match, and  $m_A$  and  $v_A$  are the mean and variance of  $X_A$ , then

$$m_A = 3p_A + p_D, \quad v_A = 9p_A + p_D - m_A^2.$$

Let the random variable  $Y_k$  denote the number of points held by a particular team after in total  $k$  matches been played by all teams. Since there are  $n$  teams, this particular team has played  $k/n$  home matches and  $k/n$  away matches (provided that the matches are distributed evenly). These  $k/n$  individual matches are identically distributed independent random variables, so the mean  $m_k$  and variance  $v_k$  of  $Y_k$  are given by

$$m_k = \frac{k}{n}(m_H + m_A), \quad v_k = \frac{k}{n}(v_H + v_A).$$

The distribution  $Y_k$  is a *multinomial distribution*, and, for large values of  $k$ , such distributions are well approximated by normal distributions with the same mean and variance.

The previous two paragraphs describe the mathematics behind the chance distribution. We record the values  $n$ ,  $N$ ,  $W_H$ , and  $W_A$  and use these to determine, after  $k$  matches have occurred, the random variable  $Y_k$ . The chance distribution is a plot of the normal approximation to the probability distribution of  $Y_k$ , which has been discretized, and scaled vertically so that the area of the plot is  $n$  (rather than 1). This means that the height of a bar positioned at  $T$  points represents the number of teams we would expect to have  $T$  points if the outcome of each match occurred in the random fashion described above.

The % variance due to chance

The large white number in the bottom right hand corner labelled '% variance due to chance' is a measure of how far the point distribution is from the chance distribution. A small number indicates a large spread in team quality, and a large number indicates that teams are of similar ability.

Formally, the % variance due to chance after  $k$  matches is defined by the formula

$$100 \times \frac{\text{theoretical variance}}{\text{sample variance}}.$$

In this equation, the *theoretical variance* is the variance of the random variable  $Y_k$  described above; it is the variance we would expect if the outcome of each match was determined randomly using the probabilities  $p_H$ ,  $p_A$ , and  $p_D$  defined in the previous section. The *sample variance* is the variance of the set consisting of points per team after  $k$  matches. That is, if after  $k$  matches the teams have points  $T_1, T_2, \dots, T_n$ , then the sample mean  $\mu_k$  and sample variance  $\nu_k$  are given by

$$\mu_k = \frac{T_1 + T_2 + \dots + T_n}{n}, \quad \nu_k = \frac{T_1^2 + T_2^2 + \dots + T_n^2}{n} - \mu_k^2.$$

[Back to Animations Index](#)