Report on findings relating to the prediction of AFL game results. 🏈

Intro

The prediction of AFL game results is conducted by both professionals and football fans in Australian culture, and it is a large part of Australian history. This poster analyses how game results can be statistically analysed to predict game results as games are being played.

Aim

The aim of this research was to find how well the winner of a game could be predicted as the game was being played. This prompted the research into whether the winning team of past games was winning at each quarter, and on average, by how much. This was extended by researching whether this had changed over football's history, and whethor it varied between grand finals and typical games.

Hypothesis

The leading team at any given quarter is more than 50% likely to win the game, and this probability will increase as the game progresses.
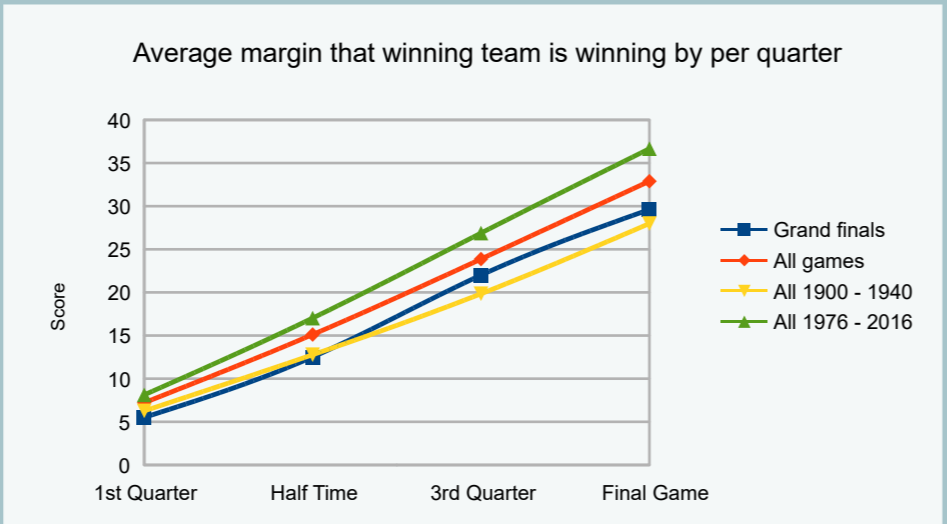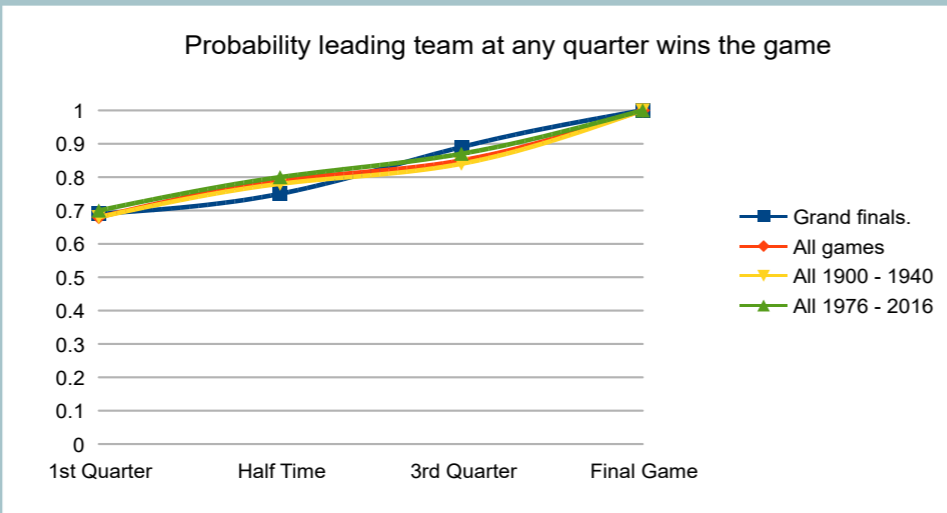
Process

Due to the scale of the data (every AFL game ever recorded) a script was created by a team member to scrape the data avaliable online into a database that could be further analysed by a secondary script. This script was written by the team member in Node JS and is avaliable online. See final section.

Reliability of data

The data avaliable online was verified from various websites. The sample size was counted by the script to be over 16000 games analysed, giving highly accurate results. The scripts results were manually checked for a number of seperate years and tested correct. With computers making no mistakes, we believe this data is highly precise.

Results

Probability leading team at any quarter wins the game

Average margin that winning team is winning by per quarter

Probability leading team at any quarter wins the game

	1st Quarter	Half Time	3rd Quarter	Final Game
Grand finals.	69.00%	75.00%	89.00%	100.00%
All games	68.00%	79.00%	85.00%	100.00%
All 1900 - 1940	68.00%	78.00%	84.00%	100.00%
All 1976 - 2016	70.00%	80.00%	87.00%	100.00%

Average margin that winning team is winning by per quarter

	1st Quarter	Half Time	3rd Quarter	Final Game
Grand finals	5.51	12.51	22	29.61
All games	7.21	15.11	23.87	32.89
All 1900 - 1940	6.26	12.77	19.84	27.98
All 1976 - 2016	8.12	17.02	26.88	36.66

Per team data was also gathered, however analysation showed that differences between teams were minimal and no major observations were made.

Observations

When the probability of the leading team at each quarter winning the game is ploted, the hypothesis that probability will increase over time is verified as being true. The leader by the games end is the winner 100% of the time, which makes sense, as the winner is by definition the leader at the fourth quarter.

This graph depicts the average margin by which the winning team of a game is winning by per quarter. This data includes quarters in which the winning team was behind the leading team, pulling averages down. As it is clear, by the end of the game, the leading team is typically beating the losing team by a significant score.

The data above is a representation of this table, which shows the exact analysis. This table can be used to form several conclusions, such as that grand finals are won by 29.61 points on average, the predictability of games has increased over time, and by the third quarter, grand finals are the most predictable games.

Discussion / Conclusion

When these graphs are viewed together, it can be said that as a football match progresses, the winning team typically excels in front of the losing team, and the losing team has a lower chance of making a comeback. This could be because losing by a lot is harder to catch up on than losing by a small margin.

When the years 1900 - 1970 are compared to the years 1976 - 2016 in football, the average probability the third quarter leader won the game increased 3% and the average margin the winning team was leading by in the third quarter increased 8.68 points. This increase in both likelyhood of third quarter leader winning and third quarter margin could be due to rule and player development, or could simply be a random difference in scores.

Analysis could have been improved by extraploting trends in data, and placing games individually on a graph. Better visualisation, such as a year by year graph may have also been insightful.

Overall, the hypothesis proved true and a variety of insights were made.

Notes

All scores are calculated as (Goals x 6) + Behinds

'All Games' is all games in rounds or grandfinals in the years 1900-2016 that were not draws or other special exceptions.

Sources

Database scraped from http://afltables.com

Code written to analyse database: https://github.com/5Mixer/FootyStats