Winning Strategies

Teachers Teaching with Technology™
Professional Development from Texas Instruments

Teacher Notes & Answers

7 8 9 10 **11** **12**

TI-Nspire™ Investigation Student 50 min

Teacher Notes

The idea behind using genuine, AFL data is to motivate students. A short video of an AFL match, highlight reel or club theme songs help establish the theme. If students watch a short section of an actual game, ask them to record the statistics … it is exceptionally challenging!

The main focus of the activity is for students to get an understanding of correlation.

If your school has access to the TI-Navigator system the screen capture tool works extremely well for this activity. Ask all students to do a scatterplot with either position (ladder) or wins on the independent axis and then a selected statistic on the dependent axis. Discuss which one has the strongest correlation. Discuss why some graphs have a positive direction and others have a negative.

What makes a winning football team?

"The most fertile source of insight is hindsight."[1] Last football season may be finished but statistical information can provide insights with regards to successful strategies in subsequent seasons. Commentators often claim that a team is hand-balling too much. Screams of "kick it" from the crowd echo frustrations when a team is not moving the ball forward. Is there any evidence to support a correlation between hand-ball frequency during a game and a team's success or failure? What about marks, contested marks or kicks; which statistics provide the strongest correlation with a team's on field success? Teams employ statisticians to provide evidence rather than anecdotal observations and emotional responses to ensure decisions and strategies are well informed.

The TI-Nspire file: "Winning Strategies" contains data on each AFL team corresponding to the 2016 home and away season. The data has been sourced from the AFL website[2]. In this investigation you will use statistics to identify which game play characteristics have the strongest correlation with a team's final home and away ladder position[3] in the 2016 season. Different ladder placements may be occupied by teams with an equal quantity of wins for the season, so 'win's has also been included in the data. From a strategic perspective, the strength of the correlation is not the only consideration. When analysing the data, consider the evidence carefully and in context. Ideas, suggestions and information relating to the data are included in the Glossary.

---

[1] Morris Kline
[2] http://www.afl.com.au/stats
[3] Ladder Position – Refers to end of 'home and away' season, all games up to but not including finals.

Author: P. Fox

Available Data

Open the TI-nspire file "Football Regression".

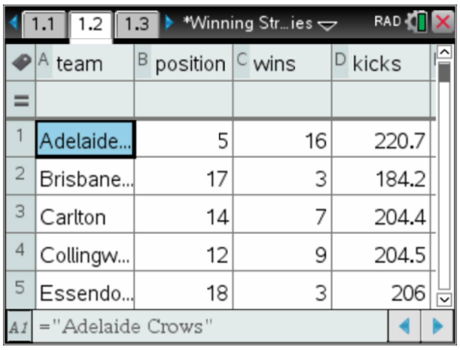
Navigate to page 1.2.

Team Alphabetical listing of the AFL teams.

Ladder Refers to the team's final ladder position (1 to 18) at the conclusion of the home and away season.

Wins Number of games won during the home and away season.

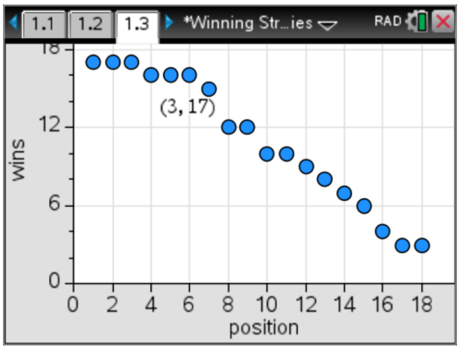
Details on each column are contained in the Glossary section.



	A team	B position	C wins	D kicks
1	Adelaide...	5	16	220.7
2	Brisbane...	17	3	184.2
3	Carlton	14	7	204.4
4	Collingw...	12	9	204.5
5	Essendo...	18	3	206

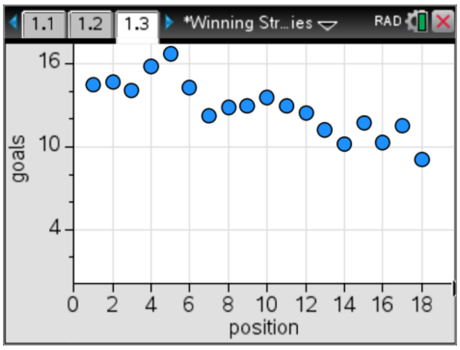
Do not attempt to sort individual columns, this will disassociate the team with their corresponding data. When statistical plots are generated they automatically display in the appropriate order based on the data represented on the x axis.

Page 1.3 contains a Scatter Plot of each team's ladder position versus the number of wins for the season. Logically a strong correlation must exist between these two variables. Teams with the same number of wins however can occupy different positions on the ladder based on their percentage.



One of the data points: (3, 17) and the two points to the left illustrate that the first three positions on the ladder were occupied by teams with the same quantity of wins. Furthermore, 4th, 5th and 6th position on the ladder were occupied by teams with 16 wins each!

Changing the dependent (y) axis to 'goals' shows that a reasonable correlation exists between ladder position and the number of goals kicked per game. This is a reasonable assumption as a team's score relative to their opponents determines whether they win or lose a game.



The graph shows that Greater Western Sydney (4th) and Adelaide (5th), on average, scored more goals per game than the top three sides.

Kicking Goals

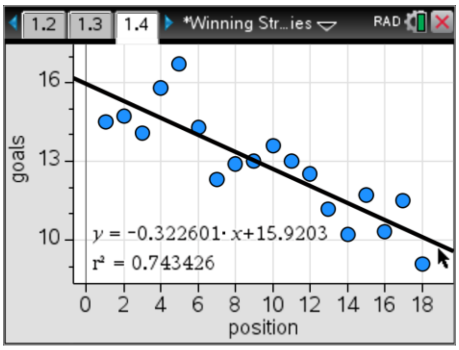
"The formula is simple, if you don't kick enough goals, you don't win!" This statement has been made by many commentators during a game and by coaches in their post-match press conferences. This is essentially true for a single game; however, is it true for the whole season?

Question: 1.

Describe the correlation between the average number of goals and ladder position.

Moderate / Strong + Negative + Linear

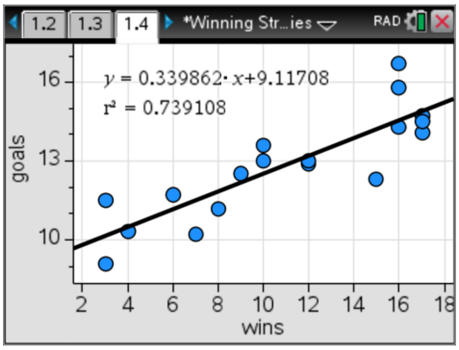
Note that students will not have completed the linear regression line so they may state that the correlation is only moderate.



Author: P. Fox

Question: 2.

Describe the correlation between the average number of goals and the corresponding number of wins.



Moderate / Strong + Positive + Linear

Note that students will not have completed the linear regression line so they may state that the correlation is only moderate.

Question: 3.

Explain the difference between your responses to Questions 1 and 2.

The correlation strengths are almost the same; a small variation exists as teams may have the same quantity of wins but hold a different ladder position. The wins data is positive (wins) as an increase in the number of goals is generally associated with an increase in the number of games won. In contrast, more goals is generally associated with teams that are higher on the ladder which means a 'smaller value' for the ladder position. The ladder has position '1' as the highest, so the larger the number the worse a team's performance. (negative)

Teacher Notes

A brilliant video to help students understand the importance of statistics over anecdotal evidence and opinion is "How not to be ignorant about the world" by Hans[4] and Ola Rosling. In the video Hans poses several questions to the audience and compares their answers to other groups including those in the media. The questions relate to information and perceptions of the world, particularly relating to poverty. When the world's media audience get the answers wrong, what hope have the general public? A quick poll of your class is worth the experience. So when decisions are made based on opinions rather than factual information, data, we can miss the 'mark' entirely. The subtitle "Coaching without Ignorance" is inspired by the late, Hans Rosling and his video "How not to be ignorant".

Coaching without Ignorance

Modern coaches are provided with streams of data on game day and during post-game analysis. On game day it is important to know where the team's strengths and weaknesses lie. If the team is being beaten at the centre bounce, it may be time to swap out the ruck, but how important is this statistic? Are there more influential plays that determine a team's success?

Question: 4.

Determine which statistic you will use for your independent (x axis): wins or ladder position. Select five different game statistics (handballs / bounces / clearances etc…) and generate a scatter plot for each.

a. Based on estimation rank your selected game statistics in order of their correlation strength. (Highest to Lowest)

Answers will vary depending on which statistics have been selected. The gradient of the least squares regression line is included in parenthesis.

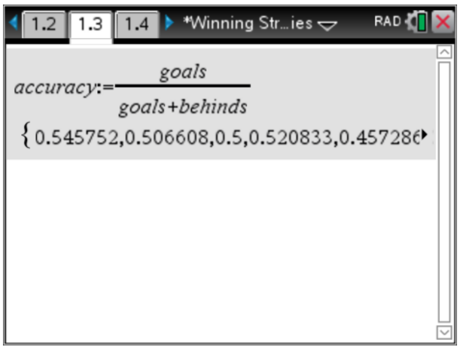
Data List	Position r^2	Wins r^2	Data List	Position r^2	Wins r^2
Behinds	0.533 (-0.146)	0.501 (0.150)	Contested Posses.	0.399 (-1.075)	0.341 (1.05)
Bounces	0.023 (-0.068)	0.029 (0.081)	Disposals	0.210 (-1.567)	0.235 (1.751)
Centre Clearances	0.269 (-0.095)	0.257 (0.257)	Frees Against	0.039 (-0.058)	0.009 (0.030)
Clangers	0.000 (-0.004)	0.011 (-0.056)	Frees For	0.046 (-0.048)	0.054 (0.055)
Clearances	0.477 (-0.286)	0.432 (0.287)	Goals	0.743 (-0.323)	0.739 (0.340)

[4] Hans Rosling – [1948 – 2017] Swedish physician, academic, statistician and public speaker. Co-founder and chairman of Gapminder http://www.gapminder.org/

Data List	Position r²	Wins r²	Data List	Position r²	Wins r²
Handballs	-0.031 (-0.425)	0.040 (0.513)	Marks	0.025 (-0.235)	0.048 (0.343)
Hit Outs	0.086 (-0.271)	0.080 (0.278)	Marks inside 50m	0.354 (-0.169)	0.425 (0.196)
Interchange	0.021 (0.038)	0.033 (-0.051)	Frees For	0.046 (-0.048)	0.054 (0.055)
Kicks	0.461 (-1.143)	0.485 (1.239)	Points Against (PA)	0.318 (1.612)	0.369 (-1.836)
Marks Contested	0.395 (-0.170)	0.399 (0.180)			

b. Describe the direction, form and correlation strength for each scatter plot.

Direction and correlation strength can all be determined from the previous table. The residual plots for each of the scatterplots do not indicate anything other than linear correlations.



Statistics can be combined to form a more inclusive view of what is happening during a match. A very simple example is to consider whether a team is accurate in front of goals.

The percentage of scoring shots that result in goals can be calculated by:

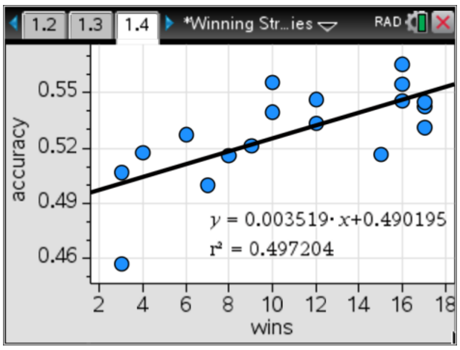
$$\frac{goals}{goals + behinds}$$

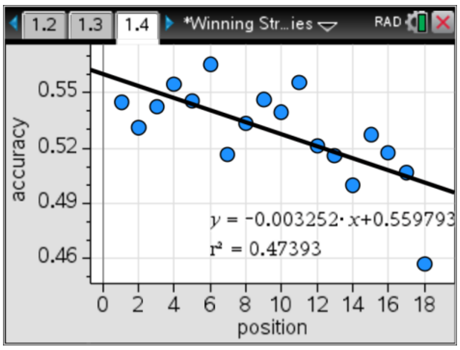
This can be computed either in a calculator application or in the spreadsheet. The instructions shown opposite are for a calculator application. A new list called "accuracy" is now available.

Note: This accuracy does not include shots on goal that may have completely missed their target and gone out of bounds or simply did not reach their target; nor does it take into account 'rushed behinds'.

Question: 5.

Accuracy or lack thereof is often linked with team success. Comments such as "that miss may cost them later in the game", particularly in close matches places increased pressure on a team to ensure they do not waste scoring opportunities. Using the 'accuracy' data, determine the strength of this correlation with your selected independent variable and comment on the result.





The correlations for accuracy versus wins or position are both moderate but are noticeable better correlated than most of the other data sets. Of particular note is the bottom team with the lowest accuracy, however the team with the second highest accuracy finished 11th. Excluding the bottom team there is only a 6% variation in accuracy across the remaining 17 teams. This variation accounts for less than 1 goal in an average game. So despite the strong correlation the cause – effect is not significant despite the frequent reference to accuracy.

Author: P. Fox

Teacher Notes

Simply adding combinations of statistics may not produce an accurate representation. For example the average number of contested marks per game is 11. The average number of contested possessions per game is 142. Simply adding the two sets of data reduces the impact of 'marks per game'. A better option may be to re-write each as a percentage referenced against the average.

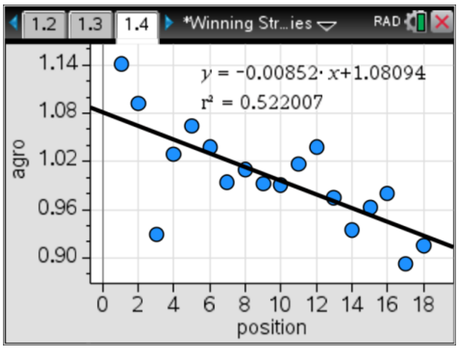
Suppose 'agro' (aggression) consists of contested marks, contested possessions and tackles.

Contested marks: {12.9,8.5,10.5,11.2,10.1,10.6 … }

These figures can be converted into percentages measured against the average:

Percentage Contested Marks, (PCM) calculated by markcontest/mean(markcontest): {1.134, 0.747, 0.923…}

Applying this approach to contested marks, contested possessions and tackles and averaging these results produces an 'agro' list: {1.065, 0.893, 0935, 1.039 …}



One of the stand-out teams for 'agro' is the team that finished the home and away season in 3rd position. Interestingly this team was Hawthorn. If Hawthorn's result was removed from the data the correlation is much stronger. ($r^2 = 0.80$). Also interesting to note that Hawthorn did not win any of their finals matches and their performance in the 2017 season may also be reflective of this 'agro' statistic.

Question: 6.

Construct at least three other sets of logically concatenated data to form alternative data sets. Use the same selection for the independent axis and generate scatter plots for each of the new data sets.

a. Justify each of your concatenated data sets.

Answers will vary enormously. The aim is to find data sets that may combine logically to produce a stronger correlation. A good combination might be contested marks, contested possessions and tackles. These three statistics can all be related via a team's determination at the ball, however to ensure each statistic is reasonable weighted students should consider the statistics as percentage of a total. Other logical combinations may combine marks, kicks, handballs and bounces giving a more complete picture of how a team possesses the ball.

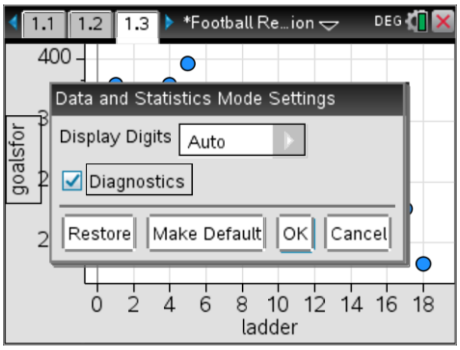
b. Rank the new data sets in order of their correlation strength. (Highest to Lowest)

Again, answers will vary enormously based on the data sets combined. Recommend that students include a copy of their 'formula' used to combine teams and a copy of the corresponding scatter plots.

c. Describe the direction, form and correlation strength for each scatter plot.

Again, answers will vary enormously based on the data sets combined, remembering the aim is to identify aggregated data that may provide a stronger correlation and therefore better strategic planning.

Quantifying Correlation

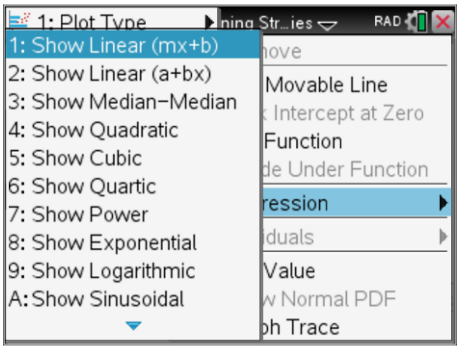


Pearson's Product-moment Correlation Coefficient provides a quantitative measure of the strength of the linear relationship between two variables. This value can be displayed on the screen with the corresponding equation by switching on the diagnostics option.

While in the **Data and Statistics** application (Page 1.3), use the menu key to access:

> Settings

Check the box for **Diagnostics** then select OK (or Make Default)



Linear regression can be used to determine a 'line of best fit' for the data and the r^2 value will quantify the strength of the correlation.

When the axis data selection is changed the equation and corresponding correlation coefficient are automatically updated.

Question: 7.

Question 4 required you to estimate the strength of each correlation. Next to each of your graphs in Question 4 include the Least Squares Regression equation (line of best fit) and the corresponding correlation coefficient.

Correlation coefficient and gradient all included in previous table.

Question: 8.

Question 6 required you to estimate the strength of each correlation. Next to each of your graphs in Question 6 include the Least Squares Regression equation (line of best fit) and the corresponding correlation coefficient.

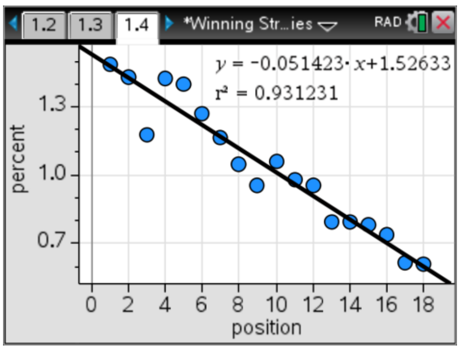
Answers will vary depending on the data that students aggregated.

Question: 9.

The AFL determines a team's percentage by calculating the total score for the team in question and dividing it by the total score of their opposition. While the figures supplied in this spreadsheet reflect the average for each game, the percentage can still be computed by using the formula below: (PA = average Points Against)

$$percent := \frac{6 \times Goals + Behinds}{PA}$$

a. Graph 'percent' vs 'position' and include a Least Squares Regression line.



Author: P. Fox

b. Describe the strength of the correlation and identify any values (teams) that may be outliers.

Correlation strength is strong. One relatively distinct outlier, the team that finished 3rd. (Hawthorn)

c. Use the regression equation to determine an appropriate 'percentage target' for any team or teams that were well below those predicted by the least squares regression line.

The regression equation can be retrieved automatically from the Variable menu. The 3rd team considerably below the regression line, this team should aim for a percentage of approximately 137%. Of course if they achieved such a percentage the regression line would move … therefore demanding a slightly higher percentage again.

Investigation – The mysterious 19th Team

Imagine an additional team existed in 2016 called the Tasmanian Devils. Unfortunately only some of the team's data was recorded. The available data includes:

Accuracy: 54%	($r^2 = 0.474$)
Percentage: 105%	($r^2 = 0.931$)
Average Disposals per game: 360	($r^2 = 0.210$)
Centre Clearances per game: 13	($r^2 = 0.477$)
Frees Against per game: 21	($r^2 = 0.039$)
Frees For per game: 20	($r^2 = 0.046$)
Marks inside 50 per game: 12.2	($r^2 = 0.354$)
Contested Marks per game: 11.4	($r^2 = 0.395$)

Based on the information above; estimate the number of wins and corresponding position on the ladder for the Tasmanian Devils team. (Justify your estimate.)

Students must consider correlation strength and cause / effect. Based on the data the strongest predictor of a team's position on the ladder (selecting from the options above) is: percentage. The next strongest is the group: accuracy, centre clearances, marks inside 50 and contested marks per game. Using the regression equation for each the predictions are as follows:

Percentage Predication: 9th (9.26)

Accuracy 6th (6.09), Centre Clearances: 5th (4.69), Marks inside 50: 8th (8.35) and Contested Marks: 9th (9.34).

Based on the above calculations it is reasonable to assume that the team finished either 8th or 9th. The strongest predictor (percentage) places the team in 9th, however most other predictions put the team higher than this, whilst they all have a weaker correlation, the collective impact would tend to move the team higher rather than lower.

Author: P. Fox

Glossary of Statistical Information

Additional statistical measures can be generated from this data. For example, the number of disposals is simply the sum of kicks and handballs, these are referred to as concatenated data.

Data (List)	Description
Position	Position on the ladder at the end of the home and away season. (1 = Top)
Wins	Teams holding different ladder positions may have the same number of wins, differing only by percentage.
Kicks	Average number of kicks per game.
Handballs	Average number of handballs per game.
	The average number of handballs and kicks can be combined to produce the average number of disposals per game: disposals = handballs + kicks
Bounces	The average number of bounces can give an indication of the amount of 'running' a team does during a game. Teams also use GPS tracking to determine the distance a player moves during a game.
Marks	This statistic can be contreversial. If teams pass the ball around a lot in what is often referred to as 'transitioning play' or to 'maintain possesion', can become a 'cheap' statistic.
Marks in 50	Marks inside a teams forward 50 are rarely about maintaining possession, they generally represent either a penetrating kick from outside the 50m arc or an attempt to gain a better shooting angle.
Hit outs	There is no doubt a team cannot play without a ruck, but just how effective is a good ruck? Their job is to ensure the ball is tapped directly to their team mates at each bounce of the ball.
Free For	This data represents free kicks awarded to a team and often generate an enormous amount of angst amongst supporters, 'we were robbed' is often the claim, but does it really effect a team's success?
Free Against	Free's against are equally as controversial as Free's for when it comes to a supporter's perspective. So the question remains, does the data support a correlation between free kicks and a teams success?
	A new measure called 'favour' can be created using the tallies of Free For and Free Against. Dividing Free For by Free Against provides an indication of how much advantage may have been generated by umpiring decisions.
Tackles	This is a very interesting statistic! In order to rate high on this statistic a team cannot be in possession, however there is no doubt that a team that tackles fiercly wins possesion. So, does a high tackle rate count for anything?
Clearances	Perhaps a measure of the effectiveness of a team's hit-outs. This statistic refers to when a team manages to clear the ball away from a bounce anywhere on the ground.
Centre Clearance	A subset of the clearance statistic, this focuses specifically on the centre ball bounce after a goal.
Goals	Goals (6 points) scored per game.
Behinds	Behinds (1 point) scored per game.

Author: P. Fox

The average number of points ‘for’ can be obtained by adding up the behinds and six times the average number of goals.

A measure of a teams accuracy for goal can be determined by dividing the average number of goals per game by the average number of scoring shots. Note that this does not account for the number of times the ball was missed completely or whether the opponents rushed a behind (point).

- **Points Against** — Points against is the combination of goals (6 points each) and points a teams opponents scored.
- **Interchange** — Number of times per game that a player or players are taken on/off the bench.
- **Marks Contested** — A contested mark is where an opposition player has the opportunity to interfer with the person marking the ball.
- **Clangers** — Fundamental player errors. These may be very frustrating for the coach and supporters, however they often relate to the amount of pressure the opposition side is applying.
- **Contested Possession** — When neither team has clear possession of the ball it is deamed to be ‘in-dispute’. When a team comes away with the possession after a ball has been in-dispute it is referred to as a contested possession.

Game statistics can be concatenated to produce a single measure. Contested marks, contested possessions and tackles may be combined to give an overall picture of how aggressive a team may have been.

Author: P. Fox