A New Agenda
https://sites.google.com/site/gordonburtmathsocsci/home/a-new-agenda

Special Supplement **Commentary, July 2014, No. 6**

The World Cup 2014: probability, prediction and outcome

'The problem with statistics is that while they demonstrate clearly that there will be surprises, it is logically impossible to identify which ones they will be. Otherwise they wouldn't be a surprise.'
The Times (2014) Fink Tank. Daniel Finkelstein. The Game. Saturday 28 June, p. 10.

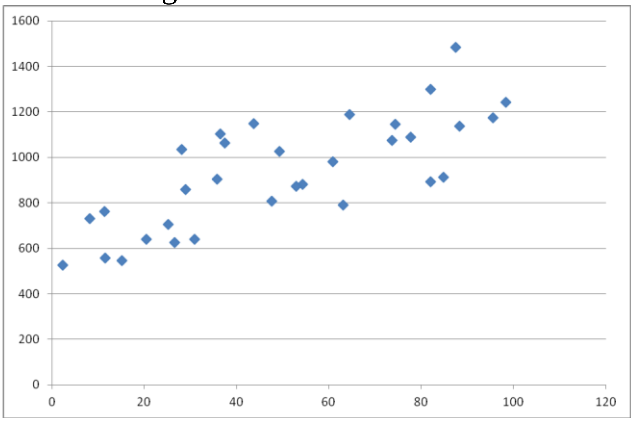
1 Discussion before it started

Before the World Cup started there was extensive discussion about how good the teams were and about their chances of winning. One source of information was the FIFA rankings. Another source of information was the work which is regularly reported on in The Times by Daniel Finkelstein. Just before the World Cup Daniel Finkelstein presented the chances of each country reaching the knockout stage – and their chances of winning the cup.
The Times (2014) The game. World Cup 2014. Wednesday June 11. pp. 12-19.
The Times (2014) The Fink Tank. Daniel Finkelstein. The game. World Cup 2014. Thursday June 12. p. 15.
FIFA rankings: http://www.fifa.com/worldranking/rankingtable/

Looking at the FIFA rankings and Finkelstein's measures, the correlation between the FIFA rankings and the chances of reaching the knockout stage is 0.8, see Figure 1; and the correlation between the FIFA rankings and the chances of winning the cup is 0.46. There is a correlation of 0.56 between the chances of reaching the knockout stage and the chances of winning the cup, see Figure 2. Note that many teams have an almost zero chance of winning the cup.

Figure 1 FIFA rankings June 4 2014 v. chances of reaching the knockout stage

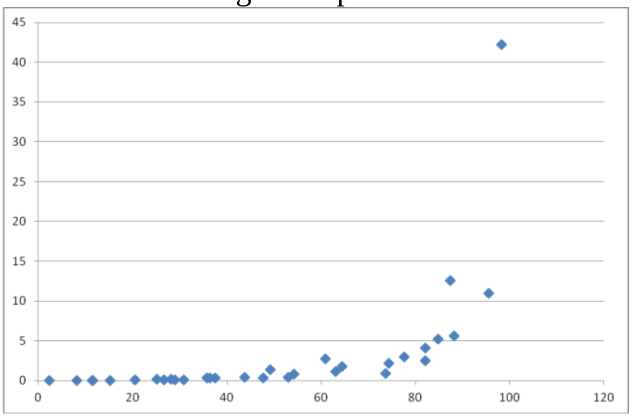
FIFA rankings June 4 2014



. chances of reaching the knockout stage

Figure 2 chances of winning the cup v. chances of reaching the knockout stage

chances of winning the cup



. chances of reaching the knockout stage

Finkelstein made a number of surprising statements which teach us about the concept of probability and about how we express this idea in ordinary language:

The best side in the world probably won't win the World Cup.
'Brazil are the best side in the world ... Brazil probably won't win the World Cup '.
The point is that the probability of Brazil winning the cup is 44.7%.
So the probability of Brazil NOT winning the cup is 54.3%.
So it is more likely that Brazil will not win the cup!

The probability of an event can never tell us for certain whether or not an event will occur. So what does a probability tell us about future events? It tells us about their probability. I think that is all I can say – but I'd like to think about it a bit more! And that probabilities sharpen as we look at a larger number of events.

2 The first outcome: reaching the knockout stage

Sixteen teams reached the knockout stage and sixteen teams did not. The outcome is consistent with the chances presented by Daniel Finkelstein. Teams which had high Finkelstein chances of reaching the knockout stage were more likely to actually do so. Looking at the relationship in reverse, teams which qualified had higher chances (a median of 58%); and teams which did not qualify had lower chances (a median of 38%). However the relationship is weak – the correlation between chances and outcomes is only 0.13. See Figure 3 on next page.

Suppose we had predicted the outcome by selecting the two teams in each group which had the two highest chances. How successful would our prediction have been? Of the sixteen teams we predicted, only nine would qualify for the knock out stage and seven would not.

Five of the teams with the highest chances in their group – Brazil, Colombia, France, Argentina and Germany – were group winners.
Three of the teams with the highest chances in their group - Spain, England, Russia - did not even qualify.

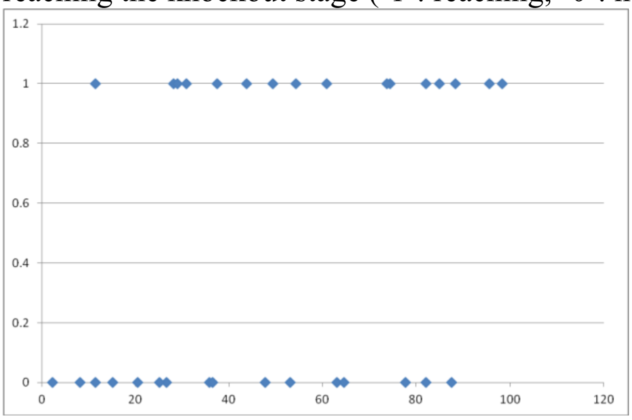
Four of the teams with the second highest chances in their group – Mexico, Holland, Uruguay, Belgium – qualified, Holland and Belgium being group winners.
Four of the teams with the second highest chances in their group – Ivory Coast, Ecuador, Bosnia, Portugal – did not.

The Appendix gives the detailed results.

In summary the Finkelstein chances give only a very weak prediction of the outcomes. Figure 4 makes this point by presenting the cumulative distribution of successful outcomes (up to 16) as a function of predictor on an ordinal scale for (i) a random predictor; (ii) the actual predictor; and (iii) a perfect predictor. The actual predictor is closer to the random predictor than to the perfect predictor. Note that this does not mean that the Finkelstein's chances are wrong – it just means that chances are not certainties. For example it is true to say that the chance of a coin falling heads is ½ - but our predictions will be correct only 50% of the time.

Figure 3 Reaching the knockout stage v. chances of reaching the knockout stage

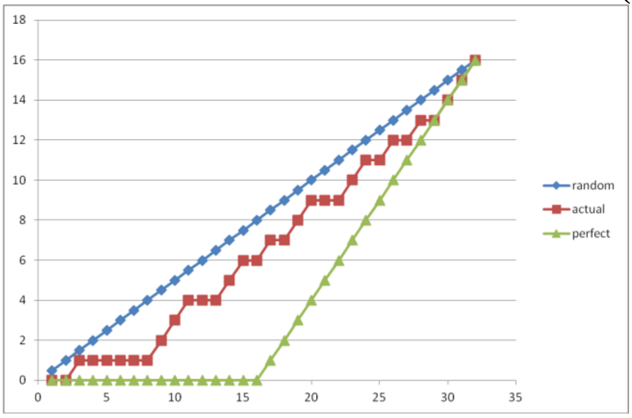
reaching the knockout stage ('1': reaching; '0': not reaching)



. chances of reaching the knockout stage

Figure 4 Cumulative distribution of successful outcomes (up to 16) as a function of predictor on an ordinal scale for (i) a random predictor; (ii) the actual predictor; and (iii) a perfect predictor.

Cumulative distribution of successful outcomes (up to 16)



.predictor (chances of reaching the knockout stage) on an ordinal scale ...
... for (i) a random predictor; (ii) the actual predictor; and (iii) a perfect predictor.

Moreover qualifying is a rather impoverished measure of how well the teams performed. If we look at more sophisticated measures of performance then Finkelstein's measure does moderately well. The hypothesis is that Finkelstein's chances of success predict the following variables: place in the group table; points in the group; goal difference; and the percentage of goals. Indeed recalibrating the variables to equal the percentage of the maximum score for each variable we can postulate an ideal equation:

. chance of success = place = points = goal difference = percentage of goals

Notes:
Recalibrated place = (4-place)/3;
Recalibrated points = points/9
Recalibrate goal difference = (goal difference+3)/6
Percentage of goals = (goals for) / (goals for + goals against)
Percentages 0% to 100% are equivalent to decimals 0.0 to 1.0.

Thus 100% chance corresponds to: first place, maximum points (9), a goal difference of +3 per game and 100% of the goals. Colombia comes closest to this ideal with 88.3% chance, first place, maximum points (9), goal difference of +2.3 per game and 82% of the goals.

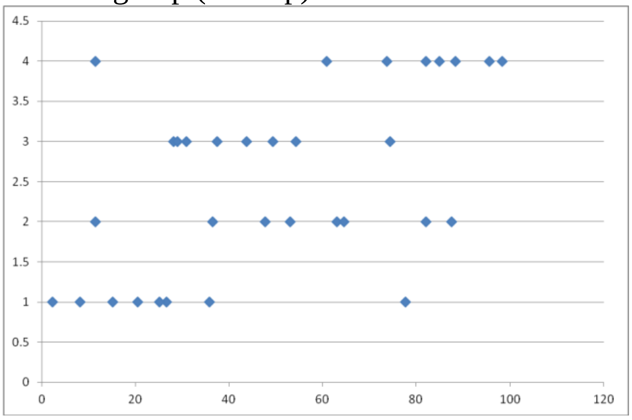
Likewise 0% chance corresponds to last place, zero points, goal difference of -3 per game and 0% of the goals. Australia comes closest to this ideal with 2.3% chance, last place, zero points, goal difference of -2.0 per game and 25% of the goals.

The empirical relationship between these variables is not perfect. Chances correlates with place (r=0.52), see Figure 5; with points (r=0.60), see Figure 6; with goal difference (r=0.56) and with percentage goals (r=0.51). The variables of place, points, goal difference and percentage of goals all correlate about 0.9 with one another.

Notes:
.(i) Finkelstein's chances can be used to predict places and this correlates 0.50 with actual places.
.(ii) Whereas Finkelstein's chances correlate 0.60 with points, FIFA rankings do almost as well, correlating 0.52 with points.
.(iii) Fixed-sum points (2 for a win, 1 for a draw and 0 for a loss) correlate 0.966 with actual points.

Figure 5 Place in group is related to predicted chances

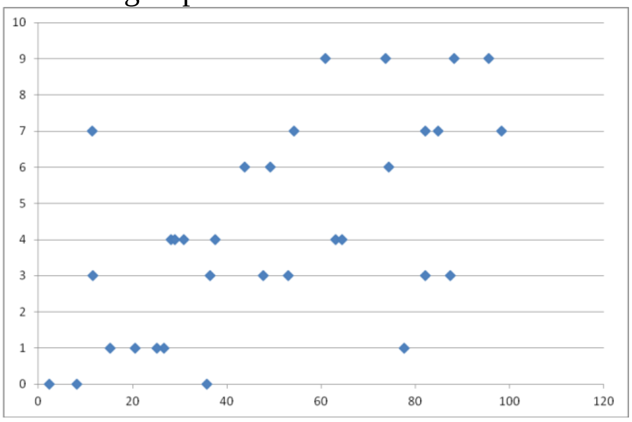
Place in group (4 is top)



. % chances of reaching knock-out stage

Figure 6 Points in group is related to predicted chances

Points in group



. % chances of reaching knock-out stage

3 The second outcome: winning the cup

'What now? A huge amount of randomness is about to be injected into the tournament.'
The Times (2014) Fink Tank. The Game. Daniel Finkelstein. Saturday 28 June, p. 10.

Today, Saturday 28 June, will be the first game in the knockout stage. Let us use the Finkelstein's statistics to predict what will happen now. Display 1 below presents the remaining four stages of the tournament. In each game the team on top is the one predicted to win.

First the eight games involving the last 16 are considered. The first game is Brazil v. Chile. The favoured team is Brazil and is placed in the top row; it has an 84% chance of winning – see row A. Its chances of winning the cup were 44.7% before the start and are now 42.9% - see rows B and C. In contrast Chile has only a 3.7% chance of winning the cup – see row D.

Similarly for the other games involving the Last 16. Note that Brazil, France, Germany and Argentina have good chances of winning their first games. In contrast Columbia, Holland, Greece and Belgium are only slightly more likely to win than their opponents – their chances of winning are only 63.4, 62.3, 56 and 58.2%.

Similarly for the later stages. So overall we predict:
Brazil, Holland, France, Argentina, Columbia, Greece, Germany, Belgium in the quarter-finals;
Brazil, Holland, France and Argentina in the semifinals;
Brazil and Argentina in the finals;
Brazil as winner;
France in third place.

Is this what will actually happen? Unlikely! See Finkelstein's warning comments before the World Cup started. See also the poor predictions of which teams would reach the knockout stage. The predictions for this knockout stage are likely to be worse since knockout depends on just one game.

4 Now that it is all over

Display 1 presents in bold **the winning teams**.

Last 16. Seven of the eight teams predicted to win did so, the exception being Greece who lost to Costa Rica in a penalty shoot-out. However Greece only had a 56% chance of winning anyway.

Quarter-final. The predictions for the winners of the quarter-finals remain as before: Brazil, Holland, France and Argentina (Holland would be predicted to win over Costa Rica just as they were predicted to win over Greece). Three of the four teams predicted to win did so, the exception being France who lost to Germany.

Semi-final. One of the two teams predicted to win did so, the exception being Brazil who lost to Germany.

Third place. The team predicted to win did not do so. Brazil lost to Holland.

Final. The team predicted to win did not do so. Argentina lost to Germany.

Note that a Brazil-Argentina final was predicted.

Note that all the predictions are based on the teams' chances of winning the world cup as stated by Finkelstein.

Overall then 11 out of 16 of the knock out stage games were correctly predicted. Not Greece at the last 16 stage; not France in the quarter-finals; not Brazil in the semi-finals; not Brazil in the play-ff; and not Argentina in the final.

Display 1 The knockout stage: predicted winners on top; actual winners in bold

Last 16

Brazil	**Columbia**	**Holland**	Greece	**France**	**Germany**	**Argentina**	**Belgium**
A 84%	63.4%	62.3%	56%	77.5%	80.8%	75.8%	58.2%
B 44.7	7.0	6.0	0.3	9.7	5.1	13.5	1.9
C 42.9	6.7	9.3	-	10.1	5.8	14.3	-
Chile	Uruguay	Mexico	**Costa Rica**	Nigeria	Algeria	Switzerland	USA
D 3.7	0.8	1.8	0.3	0.3	0.1	0.4	0.5

Quarter-finals

Brazil	**Holland**	France	**Argentina**
44.7, 42.9	6.0, 9.3	9.7, 10.1	13.5, 14.3
Columbia	Greece/Costa Rica	**Germany**	Belgium
7.0, 6.7	0.3, -	5.1, 5.8	1.9, -

Semifinals

Brazil	**Argentina**
44.7, 42.9	13.5, 14.3
France/ **Germany**	Holland
9.7, 10.1	6.0, 9.3

Final	Losing semifinalists
Brazil/ **Germany**	France / Brazil
44.7, 42.9	9.7, 10.1
Argentina	**Holland**
13.5, 14.3	6.0, 9.3

Winner	Winner of losing semifinalists
Brazil, **Germany**	France, **Holland**
44.7, 14.3	9.7, 10.1

How else might we judge the worth of the predictions? A useful boundary is the 4% chance of winning the cup. Seven teams had more than 4% chance of winning the cup: Spain, Brazil, France, Argentina, Columbia, Holland and Germany. Of these, six made it to stage 2 (quarterfinals) – Spain did not even qualify; four made it to stage 3 (semifinals) – Colombia and France did not; two made it to the finals – Brazil and Holland did not; and one was the winner of the final. So out of 7, (7,6,6,4,2,1) reached stages 0 to 5. In contrast, of the 25 teams which had less than 4% chance, (25,10,2,0,0,0) reached stages 0 to 5.

These figures are presented in Table 1. At each stage 50% of the teams go on to the next stage. Table 1 shows that the percentage of the high-chance teams proceeding to the next stage is greater than the percentage of the low-chance teams proceeding to the next stage. It also shows that the proportion of the high-chance teams remaining increases as the tournament progresses.

Table 1 Progress of high-chance and low-chance teams through the tournament

Reach:	start 0	last 16 1	quarter 2	semi 3	play-off	final 4	winner 5
Teams with >4% chance	7	6	6	4	2	2	1
Progress to next stage	86%	100%	67%	50%	50%	50%	-
Teams with <4% chance	25	10	2	0	0	0	0
Progress to next stage	40%	20%	0%	-	-	-	-
% of >4% teams	22%	38%	75%	100%	100%	100%	100%

5 A simple probability model ... complex simulation

How should we conceptualise the probability of winning the world cup? One way is to break the winning of the cup into its component parts. In what follows 'p(event E)' is used to denote the probability of event E; '(A/B)' is used to denoted event A, conditional on event B.

The probability of winning the cup (*) equals the probability of reaching the knockout stage (K) multiplied by the conditional probability of winning the cup, given that one has reached the knockout stage (*/K).

$$p(*) = p(*/K)\, p(K)$$

The probability of reaching the knockout stage is complicated. So I make a simplification and say it can come about either by defeating all the other three teams in one's group or by defeating two of them and losing to a third. Denoting the other three teams A, B and C; and denoting defeating A by p(A), we have:

$$p(K) = p(A)\, p(B)\, p(C) + (1-p(A))\, p(B)\, p(C) + p(A)\, (1-p(B))\, p(C) + p(A)\, p(B)\, (1-p(C))$$

The probability of winning the cup, given that one has reached the knockout stage is the successive conditional probabilities of defeating the other team in the next four rounds – teams W, X, Y, Z.

$$p(*/K) = p(Z/Y)\, p(Y/X)\, p(X/W)\, p(W/K)$$

Thus the overall probability of winning the cup is dependent on the probability of winning individual games against specific opponents. A final point is that the probability of winning an individual game depends on the relative strengths of the two sides – as say reflected in the FIFA rankings. This is discussed in Social Modelling Note 11, 'Outcome value and strength in a series of tournaments' and Note 17, 'Half-way through the season'

'The expected value E(v) of each game depends on: the pair strength difference d of the two players; the constant sum c of the game; the group strength range, r; and

E*(v) is the expected value of the game for the top player against the bottom player. The formula is: E(v) = c + (d/r) (E*(v)-c).' [Note 11]

'A measure of the superiority s of team A over team B is provided by the difference in points scored. ... The results are also presented in the form of a graph – see Figure 2. Crude straight line approximations for percentage wins, draws and losses are given by the three equations below. Note that: the intercepts are the percentages for the case of two teams of equal strength; the gradients are the ration of x and the maximum difference of strength; and x equals either minus the intercept (for draws and losses) or a hundred minus the intercept (for wins).

% wins = 30 + s (70/33)
% draws = 40 – s (40/33)
% losses = 30 – s (30/33)

These three equations have the general form:

. P(s) = P(0) + (s/s-max) (P(s-max)-P(0))' [Note 17]

In my reasoning I have avoided complications and chosen to look at a simple model. One approach to addressing the complications is to use a computer simulation. Rather than solve equations to find a probability the simulation is run a huge number of times to give a relative frequency which is then taken as the probability.

6 Gambling

http://en.wikipedia.org/wiki/Odds

'Funny old game
Sir, It is fortunate I didn't place a bet on the World Cup semi-final between Brazil and Germany following Daniel Finkelstein's Fink Tank predictions (July 8). He gave Brazil a 79.8 per cent chance of prevailing. Did he not factor in the suspension of Brazil's Silva and the absence of Neymar? Perhaps like many managers who get it so wrong, he will be considering his position.
John Bretherton
West Wickham, Kent'
The Times (2014) Letters to the Editor. Thursday July 10, p. 25.

Mr Bretherton may or may not be right about the effect of losing Silva and Neymar. He may or may not be right that Daniel Finkelstein's figure of 79.8 per cent is wrong. But he is wrong to think that he can deduce that Daniel Finkelstein's figure of 79.8 per cent is wrong simply from the fact that Brazil lost – or from the fact that they lost so badly. The probability of an event can never be established simply by looking at whether or not the event occurred.

Nor do probabilities of events necessarily relate in a straightforward fashion to betting decisions ...

... in mathematics the odds is a ratio of probabilities – in gambling the odds is a ratio of amounts of money. A fair bet is when the gambling odds equals the mathematical odds and all parties have an expected gain of zero.

In mathematics, if the probability of an event happening is p then the odds in favour of the event is p/(1-p); and the odds against the event is (1-p)/p.

‘In the modern era, most fixed odds betting takes place between a betting organisation, such as a bookmaker, and an individual, rather than between individuals. Different traditions have grown up in how to express odds to customers.
Favoured by bookmakers in the United Kingdom and Ireland, and also common in horse racing, fractional odds quote the net total that will be paid out to the bettor, should he win, relative to his stake.[8] Odds of 4/1 ("four-to-one" or less commonly "four-to-one *against*") would imply that the bettor stands to make a £400 profit on a £100 stake. If the odds are 1/4 (read "one-to-four", or "four-to-one *on*"), the bettor will make £25 on a £100 stake. In either case, *against* or *on*, should he win, the bettor always receives his original stake back, so if the odds are 4/1 the bettor receives a total of £500 (£400 plus the original £100). Odds of 1/1 are known as *evens* or *even money*.’

In gambling, if a bettor bets £x that an event will happen and the (fractional*) odds are y:1 then ...
if the event occurs then the bettor wins £yx together with the original stake of £x;
and if the event does not occur then the bettor wins nothing and does not get his £x back.
* The decimal odds would be y.

Suppose the event has probability p.
Then the expected amount of money received back is p(yx+x)+(1-p)0=xp(y+1).
The expected gain is E(G)=xp(y+1)-x=x(p(y+1)-1).

‘The gambling and statistical uses of odds are closely interlinked. If a bet is a fair one, then the odds offered to the gamblers will perfectly reflect relative probabilities. If the odds being offered to the gamblers do not correspond to probability in this way then one of the parties to the bet has an advantage over the other.’

If it is a ‘fair bet’ then the expected gain is zero, E(G)=0. So:
either x=0 or p(y+1)-1=0
p(y+1)=1
p=1/(y+1)
p/(1-p)=1/y
So the mathematical odds equal the (reciprocal of the?) betting odds.

Betting organisations aim to make a profit A with the bettor making a loss of A. So they choose y so that the E(G) for the bettor is negative equal to -A.
E(G)=x(p(y+1)-1)=-A
y=[(1-A/x)/p]-1

A case study: the quarter-finals and the final

Table 2 below presents the odds offered by Ladbrokes on 2 July 2014 for the final and at 17.32 on 13 July 2014 for the final.

For the quarter-finals, it was odds on that Germany, Argentina, Brazil and Holland would each win. This agrees with the probability-based prediction in three of the cases but in the case of France-Germany match, Finkelstein probabilities predict a France win whereas Ladbrokes odds predict a Germany win.

Note that bookmakers want to make money. So they promise betters more money if the betters bet on low likelihood events. Holland are very likely to defeat Costa Rica and so Ladbrokes offer what I have called a loss/win ratio of 12 and this corresponds to a high ratio of Finkelstein probabilities: 31. At the other extreme the Germany-France game is likely to be close and so Ladbrokes offer what I have called a loss/win ratio of 1.38 and this corresponds to a low ratio of Finkelstein probabilities: 0.57.

For the final it is odds on that Germany would defeat Argentina. It's likely to be close with Ladbrokes offering a loss/win ratio of 1.60 and a low ratio of Finkelstein probabilities: 0.43.

Table 2 Ladbrokes betting odds and Finkelstein probabilities quarter-finals

.	Ladbroke betting odds				Finkelstein		
.	win	draw	loss	loss/win	% probability......		team 1/
.	for first-named team				team 1	team 2	team 2
Quarter-finals							
Germany-France	1.45/1	2.25/1	2.00/1	1.38	5.8	10.1	0.57
Argentina-Belgium	1.10/1	2.25/1	2.80/1	2.55	13.5	1.9	7.11
Brazil-Colombia	0.83/1	2.60/1	3.33/1	4.01	42.9	6.7	6.40
Holland-Costa Rica	0.50/1	3.20/1	6.00/1	12	9.3	0.3	31.00
Final							
Germany-Argentina	2.25/1	3.20/1	3.60/1	1.60	5.8	13.5	0.43

http://sports.ladbrokes.com/en-gb/Football/World-Cup-2014Football/World-Cup-2014-t210005575

7 Notes on football and sport

Social Modelling Notes
Note 11 Outcome value and strength in a series of tournaments
Note 14 Modelling sport
Note 17 Half-way through the season
Note 22 Football: a season of two halves 17.8.13
(link via: https://sites.google.com/site/gordonburtmathsocsci/home/mathematical-social-science)

Commentary, May 2014, No. 4, Section 6 Manchester United fail – blame David Moyes?
https://sites.google.com/site/gordonburtmathsocsci/home/a-new-agenda

Appendix Detailed results for the eight groups
http://www.fifa.com/worldcup/groups/

	Upsets:
Brazil 98.4	
Mexico 54.3	
Cameroon 11.5	Cameroon above Croatia 1
Croatia 35.8	
Holland 60.9	
Chile 49.3	
Spain 87.5	Spain should have been first 2
Australia 2.3	
Colombia 88.3	
Greece 37.5	Greece above Ivory coast 1
Ivory Coast 47.7	
Japan 26.6	
Costa Rica 11.4	last shall be first 3
Uruguay 74.4	
Italy 36.5	
England 77.7	first shall be last
France 84.9	
Switzerland 43.8	Switzerland above Ecuador 1
Ecuador 63.1	
Honduras 8.2	
Argentina 95.6	
Nigeria 30.9	Nigeria above Bosnia 1
Bosnia 53	
Iran 20.5	
Germany 82.1	
USA 28.1	USA above Portugal 1
Portugal 64.5	
Ghana 25.2	
Belgium 73.7	
Algeria 29	
Russia 82.1	Russia should have been first 2
South Korea 15.2	