

Previsões para os jogos da Copa do Mundo de 2018

Paulo Cezar Pinto Carvalho, Moacyr Alvim Silva, Arthur da Silva Pereira Carneiro
FGV/EMAp - 30/04/2018

Resumo: Um modelo probabilístico foi desenvolvido para fazer previsões sobre os jogos da Copa do Mundo de 2018. Para tal, admite-se que o número de gols que uma equipe A faz sobre uma equipe B tem uma distribuição de Poisson, em que a média depende de parâmetros de força de defesa e ataque, a serem estimados com base em todos os jogos entre seleções realizados nos últimos 4 anos. O modelo permite prever o resultado mais provável de cada partida e, via simulação, as probabilidades de sucesso de cada seleção.

O Modelo

Diversos autores sugeriram que o número de gols assinalados em uma partida de futebol pode ser modelado por uma distribuição de Poisson. Neste trabalho, adotamos uma modificação do método de Mahler (1982), modelando o número de gols marcados por uma equipe A em uma equipe B como tendo distribuição de Poisson com média $\frac{\alpha_A}{\beta_B}$, onde α_A é um parâmetro que mede a força de ataque da equipe A e β_B é um parâmetro que mede a força de defesa da equipe B. Os valores de α e β para as 222 seleções afiliadas à FIFA foram estimados por máxima verossimilhança, usando os resultados de todos os 4010 jogos entre seleções nos últimos 4 anos. Cada partida é afetada de um peso, que é o produto de dois fatores: o primeiro expressa a importância de cada partida e o segundo descreve quão recente ela é. Os pesos adotados para a importância são os mesmos usados pela FIFA na construção de seu ranking: 1 para amistosos, 2,5 para jogos classificatórios para a Copa do Mundo ou Copa Continentais, 3 para jogos de finais Continentais e da Copa das Confederações e 4 para a Copa do Mundo. Para o peso temporal, adotamos um amortecimento exponencial tal que o fator associado a um jogo realizado há 4 anos atrás é 0,25.

Os valores estimados de α e β permitem construir um ranking das equipes, associando a cada uma o valor $f_A = \sqrt{\alpha_A \cdot \beta_A}$, ou seja, a média geométrica das forças de ataque e defesa. Esta média produz um ranking apropriado, no sentido de que se $f_A > f_B$ então A tem maior probabilidade do que B de vencer uma partida entre eles.

A tabela abaixo fornece as forças de ataque e defesa para as 32 seleções classificadas para as finais da Copa do Mundo de 2018, além de sua força combinada f

| Selecao | Força (geral) | ataque | defesa | Força geral | ataque | defesa |
|-----------|---------------|--------|--------|-------------|--------|--------|
| Brazil | 3,370 | 3,561 | 3,190 | 1,351 | 1,428 | 1,279 |
| Spain | 3,089 | 3,432 | 2,780 | 1,238 | 1,376 | 1,114 |
| France | 2,911 | 2,888 | 2,934 | 1,167 | 1,158 | 1,176 |
| Germany | 2,903 | 3,444 | 2,447 | 1,164 | 1,381 | 0,981 |
| Argentina | 2,853 | 2,769 | 2,940 | 1,144 | 1,110 | 1,179 |
| Portugal | 2,631 | 2,547 | 2,717 | 1,055 | 1,021 | 1,089 |
| England | 2,498 | 2,126 | 2,934 | 1,001 | 0,852 | 1,176 |
| Sweden | 2,450 | 2,718 | 2,208 | 0,982 | 1,090 | 0,885 |
| Belgium | 2,444 | 3,308 | 1,806 | 0,980 | 1,326 | 0,724 |
| Croatia | 2,390 | 2,236 | 2,555 | 0,958 | 0,897 | 1,025 |
| Poland | 2,345 | 3,251 | 1,692 | 0,940 | 1,303 | 0,678 |
| Colombia | 2,274 | 2,356 | 2,196 | 0,912 | 0,945 | 0,880 |
| Denmark | 2,264 | 2,279 | 2,250 | 0,908 | 0,914 | 0,902 |
| Morocco | 2,221 | 1,952 | 2,527 | 0,890 | 0,783 | 1,013 |
| Iran | 2,171 | 1,639 | 2,875 | 0,870 | 0,657 | 1,153 |
| Uruguay | 2,128 | 2,494 | 1,815 | 0,853 | 1,000 | 0,728 |
| Peru | 2,098 | 2,080 | 2,116 | 0,841 | 0,834 | 0,849 |

| | | | | | | |
|--------------|-------|-------|-------|-------|-------|-------|
| Suisse | 2,022 | 2,240 | 1,824 | 0,810 | 0,898 | 0,731 |
| Iceland | 1,974 | 2,178 | 1,789 | 0,791 | 0,873 | 0,717 |
| Serbia | 1,878 | 2,375 | 1,485 | 0,753 | 0,952 | 0,595 |
| Russia | 1,868 | 2,210 | 1,579 | 0,749 | 0,886 | 0,633 |
| Mexico | 1,820 | 2,022 | 1,638 | 0,730 | 0,811 | 0,657 |
| Senegal | 1,714 | 1,696 | 1,731 | 0,687 | 0,680 | 0,694 |
| Japan | 1,712 | 1,947 | 1,505 | 0,686 | 0,781 | 0,603 |
| Egypt | 1,644 | 1,465 | 1,846 | 0,659 | 0,587 | 0,740 |
| Costa Rica | 1,547 | 1,448 | 1,652 | 0,620 | 0,581 | 0,662 |
| Australia | 1,540 | 1,730 | 1,371 | 0,617 | 0,694 | 0,550 |
| Nigeria | 1,509 | 1,388 | 1,641 | 0,605 | 0,557 | 0,658 |
| South Korea | 1,498 | 1,522 | 1,474 | 0,601 | 0,610 | 0,591 |
| Tunisia | 1,494 | 1,513 | 1,475 | 0,599 | 0,607 | 0,591 |
| Panama | 1,313 | 1,298 | 1,328 | 0,526 | 0,520 | 0,533 |
| Saudi Arabia | 1,297 | 1,643 | 1,023 | 0,520 | 0,659 | 0,410 |

As colunas em azul tem as mesmas informações que a coluna verde. A diferença é que na coluna verde usamos como medida padrão a força de ataque do Uruguai. Assim é mais fácil entender o significado destes números. O ataque do Brasil é 1.351 vezes mais eficiente para fazer gols do que o do Uruguai. A defesa do Brasil é 1.279 vezes mais eficiente em evitar gols do que o ataque do Uruguai é para fazer.

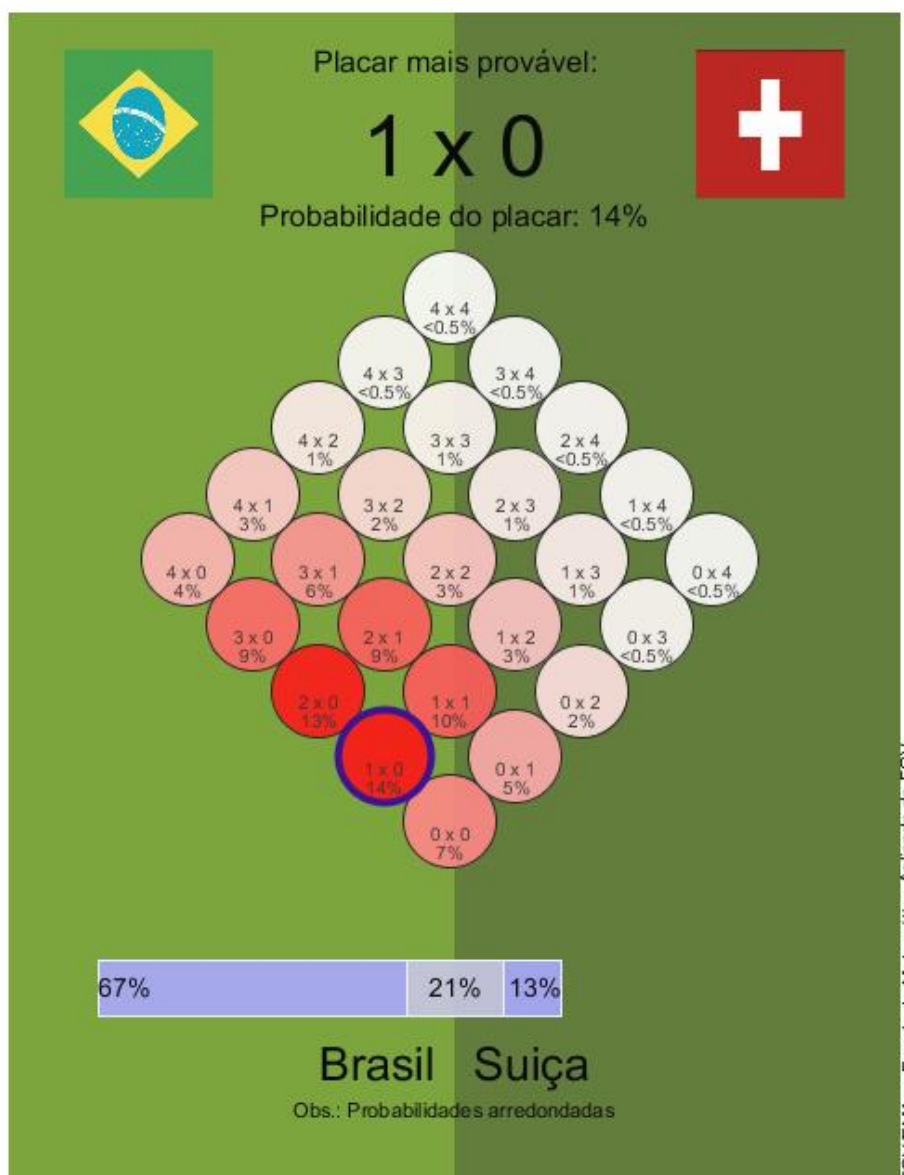
As previsões

Uma vez estimadas as forças de defesa e ataque, produzimos dois tipos de previsão:

- 1) Para os jogos da primeira fase, calculamos as probabilidades associadas aos diversos resultados. Por exemplo, para o jogo de estreia do Brasil, contra a Suíça, os valores das probabilidades dos diversos resultados são dados na tabela abaixo.

| | | Brasil | | | | |
|--------|---|--------|--------------|-------|-------|-------|
| | | 0 | 1 | 2 | 3 | 4 |
| Suisse | 0 | 0,070 | 0,137 | 0,134 | 0,087 | 0,043 |
| | 1 | 0,049 | 0,096 | 0,094 | 0,061 | 0,030 |
| | 2 | 0,017 | 0,034 | 0,033 | 0,021 | 0,010 |
| | 3 | 0,004 | 0,008 | 0,008 | 0,005 | 0,002 |
| | 4 | 0,001 | 0,001 | 0,001 | 0,001 | 0,000 |

Este mapa de probabilidades pode ser melhor visualizado graficamente. Optamos pela representação abaixo, em que a intensidade da cor é proporcional à probabilidade do resultado correspondente. A barra horizontal registra as probabilidades de vitória de cada seleção e de empate.



O resultado mais provável é Brasil 1 x 0 Suíça, com probabilidade aproximadamente igual a 14%. O resultado mais provável na partida A x B é simples de se obter e é dado por $\lfloor \frac{\alpha_A}{\beta_B} \rfloor \times \lfloor \frac{\alpha_B}{\beta_A} \rfloor$, onde $\lfloor x \rfloor$ denota o maior inteiro menor ou igual a x .

Os resultados de todas as previsões estarão disponíveis em breve no site da EMaP:

www.emap.fgv.br/copa-2018

- 2) Usando as forças estimadas podemos simular um grande número de realizações da Copa do Mundo e obter as probabilidades de cada equipe ser campeã. Podemos também estimar as chances de uma equipe atingir cada fase do torneio. Os resultados após um milhão de simulações são dados na tabela abaixo. Vale observar que o obtemos praticamente a mesma tabela com um número bem menor de simulações, como em 50.000 simulações.

| | 1o. do grupo | 2o. do grupo | chegar oitavas | chegar quartas | chegar semi | Terceiro | Vice | Campeão |
|--------------|--------------|--------------|----------------|----------------|-------------|----------|-------|--------------|
| Brazil | 76,5% | 17,0% | 93,5% | 66,8% | 49,2% | 11,5% | 11,3% | 21,3% |
| Spain | 47,7% | 26,2% | 73,9% | 59,5% | 37,5% | 9,2% | 9,6% | 13,5% |
| Germany | 58,2% | 28,1% | 86,3% | 55,6% | 36,9% | 9,4% | 9,7% | 11,4% |
| France | 54,9% | 26,9% | 81,8% | 50,9% | 31,0% | 8,4% | 7,7% | 9,5% |
| Argentina | 51,8% | 28,3% | 80,1% | 48,8% | 28,7% | 7,4% | 7,5% | 8,5% |
| Portugal | 26,2% | 29,4% | 55,6% | 39,0% | 20,8% | 5,4% | 5,4% | 5,0% |
| England | 40,7% | 39,6% | 80,3% | 48,0% | 21,8% | 5,6% | 5,7% | 4,8% |
| Belgium | 50,7% | 33,9% | 84,6% | 51,5% | 21,7% | 5,7% | 5,9% | 4,3% |
| Sweden | 30,7% | 40,1% | 70,7% | 33,9% | 18,6% | 4,6% | 5,0% | 3,8% |
| Croatia | 28,9% | 34,6% | 63,5% | 31,3% | 15,3% | 4,0% | 3,9% | 2,9% |
| Poland | 44,7% | 28,8% | 73,5% | 38,9% | 16,3% | 4,1% | 4,3% | 2,8% |
| Colombia | 34,4% | 32,0% | 66,5% | 33,3% | 13,3% | 3,3% | 3,5% | 2,2% |
| Denmark | 23,5% | 33,2% | 56,7% | 26,7% | 12,3% | 3,1% | 3,1% | 2,0% |
| Morocco | 13,6% | 22,4% | 36,0% | 21,5% | 9,3% | 2,4% | 2,3% | 1,4% |
| Uruguay | 48,8% | 28,5% | 77,2% | 28,8% | 11,4% | 3,0% | 2,6% | 1,4% |
| Iran | 12,4% | 22,1% | 34,5% | 19,9% | 8,7% | 2,3% | 2,1% | 1,3% |
| Peru | 17,4% | 28,9% | 46,3% | 19,6% | 8,1% | 2,1% | 1,9% | 1,0% |
| Suisse | 11,9% | 37,0% | 48,9% | 17,2% | 7,2% | 1,7% | 1,7% | 0,8% |
| Iceland | 14,8% | 25,2% | 40,0% | 15,0% | 5,8% | 1,4% | 1,2% | 0,6% |
| Russia | 29,9% | 32,6% | 62,5% | 18,3% | 5,9% | 1,3% | 1,3% | 0,5% |
| Serbia | 8,3% | 29,8% | 38,1% | 11,3% | 4,2% | 1,0% | 0,9% | 0,3% |
| Mexico | 8,2% | 21,1% | 29,3% | 8,2% | 2,9% | 0,6% | 0,7% | 0,2% |
| Egypt | 16,3% | 27,0% | 43,3% | 11,0% | 3,2% | 0,6% | 0,6% | 0,2% |
| Senegal | 10,5% | 20,0% | 30,6% | 10,1% | 2,4% | 0,5% | 0,5% | 0,2% |
| Japan | 10,3% | 19,2% | 29,5% | 9,3% | 2,1% | 0,4% | 0,4% | 0,1% |
| Costa Rica | 3,2% | 16,3% | 19,5% | 4,4% | 1,3% | 0,2% | 0,3% | 0,1% |
| Tunisia | 5,6% | 16,2% | 21,8% | 6,1% | 1,2% | 0,2% | 0,2% | 0,0% |
| Australia | 4,2% | 11,0% | 15,2% | 3,6% | 0,8% | 0,1% | 0,1% | 0,0% |
| Nigeria | 4,4% | 11,9% | 16,4% | 4,0% | 1,1% | 0,2% | 0,2% | 0,0% |
| South Korea | 2,9% | 10,7% | 13,6% | 2,6% | 0,6% | 0,1% | 0,1% | 0,0% |
| Panama | 3,0% | 10,3% | 13,2% | 2,8% | 0,4% | 0,1% | 0,1% | 0,0% |
| Saudi Arabia | 5,1% | 11,9% | 17,0% | 2,0% | 0,3% | 0,0% | 0,0% | 0,0% |

Atualização das previsões

Todas as estimativas acima dependem das estimativas das forças de cada seleção. Tais valores podem ser estimados novamente à medida que novos amistosos e, principalmente, os jogos iniciais da Copa sejam realizados. Todos os resultados serão atualizados diariamente durante a Copa e publicados no site da EMaP.

Referência:

Maher M.J. (1982), *Modelling Association Football scores*. *Statistica Neerlandica*, 36, 109-118.

Diniz, M. A., Izbicki, R., Lopes, D., & Salasar, L. E. (2018). Comparing probabilistic predictive models applied to football. *Journal of the Operational Research Society*, 1-13.