

# Brasileiro\_2021\_EFA

Eduardo Cecconi

6/4/2022

## Análise Fatorial Exploratória

### Dados do Campeonato Brasileiro 2021 extraídos do InStat

Assim como na Análise Fatorial Confirmatória que publiquei recentemente (neste link, , o processo inicial foi o mesmo para iniciar a Análise Fatorial Exploratória (AFE).

O banco de dados é o mesmo, extraído do provedor InStat com 115 variáveis de 379 jogos do Campeonato Brasileiro 2021; o primeiro passo é o data wrangling, que neste caso se baseia na criação da variável “Local”, definindo os mandantes (H) e visitantes (A) de cada partida, para servir de referência à construção do código de outras quatro variáveis:

- Pontos (Points)
- Gols Concedidos (Goals\_Conceded)
- Chutes Concedidos (Shots\_Conceded)
- Finalizações Concedidas em Bolas Paradas (Set\_Pieces\_Conceded)

Das 120 variáveis resultantes, 39 foram selecionadas para a elaboração das métricas, formando um novo banco de dados com 24 variáveis:

Abaixo resumo os conceitos utilizados na elaboração das métricas. Vale destacar que eles foram definidos a partir da análise das variáveis e das consequentes possibilidades de manipulação delas; um banco de dados diferente, com outras variáveis que não estas providas pelo InStat, provavelmente exigiria adaptações, incluindo exclusão de alguns e criação de novos. São eles:

- **Conversion:** razão entre os gols marcados e a expectativa de gol (xG), para privilegiar a eficiência;
- **Attacks\_Shot:** soma dos chutes realizados em 3 origens distintas de ataques (Posicionais, Transições e Bolas Paradas), o que na execução em si da AFC tem pouca importância, mas pode subsidiar paralelamente gráficos para identificar diferentes estilos de jogo ofensivo;
- **Shot\_Assist:** soma dos passes-decisivos e dos cruzamentos que geraram finalização;

- **Conversion\_Conceded:** razão entre gols sofridos e expectativa concedida;
- **Imposition:** soma das vitórias pessoais em três tipos de duelos listados (ofensivos, defensivos e aéreos), subtraída a soma de três tipos de infrações (faltas cometidas, cartões amarelos e vermelhos recebidos);
- **Contention:** soma de quatro ações defensivas listadas: tackles (um item de scout comum nas plataformas dominantes mas que, na prática, não existe no Brasil, onde o conceito disseminado é uma espécie de fusão entre tackles e recoveries - a **Roubada**), recuperações, interceptações e pressão baixa eficiente;
- **Pressing:** soma de três ações de pressão listadas: recuperações no campo de ataque, pressão alta eficiente e o que o InStat chama de “team pressing”, e que embora vago quando analisado nos vídeos, é uma espécie de “pressão na bola”;
- **Ball\_Care:** relação entre passes certos e dribles certos com as perdas de bola;
- **Passing\_Speed:** este indicador é uma adaptação do conceito de VEL (velocidade de circulação da bola); sem a cronologia das ações nos dados extraídos, estimei o tempo individual de retenção da bola por jogada através do cruzamento de três variáveis originais: passes certos, quantidade de posses de bola e tempo médio (em segundos) de cada posse;
- **Build-up:** subtração entre a variável “construção”, disponibilizada nos dados originais, e perdas de bola no próprio campo;
- **Progression:** relação entre as quantidades de ingressos da equipe do campo de defesa para o de ataque, do campo de ataque para o terço final, e do terço final para a área adversária;
- **PPDA:** Passes permitidos por ação defensiva, em inglês, variável que já foi entregue originalmente pelo InStat, sem necessitar de manipulação. Conta a quantidade de passes que o adversário troca na construção até que a equipe realize alguma ação defensiva.

Na AFE as matrizes de variância, covariância e correlação são muito importantes, pois amparam as análises dos testes estatísticos. Nesta técnica ainda não sabemos exatamente como vamos manipular os dados, logo os testes serão os guias do analista para a construção do modelo (quais métricas descartar, quais vão se relacionar, etc).

## Teste de Esfericidade de Bartlett

Com p-valor inferior ao nível de significância de 5%, é possível rejeitar a hipótese nula e afirmar que as variáveis têm correlação suficiente para aplicar as técnicas de análise fatorial, de análise de componentes principais ou de regressão. Testadas as 758 observações da matriz de correlações, o p-valor foi igual a zero, o que justifica o procedimento.

```
## $chisq
## [1] 11610.49
##
## $p.value
## [1] 0
##
```

```
## $df
## [1] 253
```

## #KMO (KAISER-MEYER-OLKIN MEASURE OF SAMPLING ADEQUACY)

Outro teste para validar a escolha pela Análise Fatorial Exploratória é o KMO, que tem parâmetro a partir de 0.5. Como o valor geral do teste aplicado chegou a 0.68, o KMO reforça que há correlação entre as variáveis.

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = Metrics_EFA[, 2:24])
## Overall MSA = 0.68
## MSA for each item =
```

	Points	Expected_Points	Conversion_Points	Goals
	0.75	0.75	0.67	0.59
	xG	Attacks_Shot	Conversion	Shot_Assist
	0.66	0.88	0.63	0.91
	Goals_Conceded	xG_Conceded	Conversion_Conceded	Shots_Conceded
	0.62	0.62	0.55	0.90
	Contention	Pressing	Imposition	PPDA
	0.53	0.70	0.75	0.23
	Possession_Time	Passing_Speed	Ball_Care	Build_Up
	0.61	0.31	0.61	0.88
	Progression	Set_Piece_shot	Set_Pieces_Conceded	
	0.81	0.83	0.91	

Dois aspectos chamam a atenção na tabela de valores do teste KMO: os resultados dos indicadores Passing\_Speed (uma estimativa de VEL, como já foi explicado) e PPDA. Ambos são conceitos atrelados ao imaginário do futebol moderno, sendo a velocidade de circulação da bola um princípio levado em alta consideração por treinadores simpáticos ao jogo posicional, por exemplo. Estatisticamente, entretanto, a variável Passing\_Speed apresentou baixo índice de correlação (apenas 0.31), enquanto o PPDA (um indicador de pressão alta defensiva) não passou de 0.23.

A baixa correlação do Passing\_Speed com este banco de dados do Brasileiro 2021 vai ficar muito evidente no artigo com uma regressão múltipla que vou publicar a seguir, mas já pode ser observada na tabela abaixo. Naturalmente, por ser um indicador com valor inversamente proporcional ao desempenho (quanto menor, melhor), esperava-se que Passing\_Speed tivesse correlação negativa com os pontos, porém o valor é de -0.02 na tabela demonstrada abaixo (ou seja, muito próximo de zero). O que permite concluir que **a velocidade de circulação da bola não foi determinante para conquistar pontos no Brasileiro.**

Ainda falando em posse de bola, a variável Possession\_Time tem correlação negativa com os pontos (-0.10), também contrariando o conceito vigente de dominação pela posse de bola para obter vitórias - pelo menos, não foi isso que se verificou no Campeonato Brasileiro do ano passado. Na verdade, **ter menor posse de bola que o adversário foi mais influente para conquistar pontos.**

```
##          Ponts Exp_P Cnv_P Goals xG    Att_S Cnvrs Sht_A Gls_C xG_Cn
## Points          1.00
## Expected_Points 0.38  1.00
## Conversion_Points 0.47 -0.27 1.00
```

## Goals	0.64	0.30	0.26	1.00						
## xG	0.31	0.75	-0.13	0.43	1.00					
## Attacks_Shot	0.18	0.57	-0.16	0.23	0.66	1.00				
## Conversion	0.44	-0.12	0.42	0.69	-0.10	-0.10	1.00			
## Shot_Assist	0.10	0.48	-0.13	0.15	0.57	0.56	-0.12	1.00		
## Goals_Conceded	-0.58	-0.28	-0.30	0.02	-0.05	-0.04	0.06	-0.04	1.00	
## xG_Conceded	-0.29	-0.70	0.29	-0.05	-0.18	-0.24	0.07	-0.21	0.43	1.00
## Conversion_Conceded	-0.40	0.12	-0.31	0.06	0.07	0.11	0.03	0.14	0.69	-0.10
## Shots_Conceded	-0.28	-0.42	0.06	0.01	-0.13	-0.19	0.10	-0.18	0.48	0.60
## Contention	0.18	-0.03	0.19	0.08	-0.01	-0.03	0.10	-0.05	-0.20	0.02
## Pressing	0.10	0.32	-0.13	0.07	0.23	0.33	-0.07	0.29	-0.11	-0.29
## Imposition	0.02	0.16	0.00	-0.01	0.16	0.21	-0.09	0.18	-0.10	-0.10
## PPDA	-0.03	0.03	-0.09	0.03	-0.02	-0.04	0.00	-0.04	0.14	-0.07
## Possession_Time	-0.10	0.33	-0.27	-0.03	0.27	0.45	-0.18	0.39	0.03	-0.27
## Passing_Speed	-0.02	0.06	-0.09	0.01	0.01	0.03	-0.01	0.10	0.04	-0.10
## Ball_Care	-0.06	0.29	-0.26	0.00	0.20	0.39	-0.12	0.34	0.05	-0.27
## Build_Up	-0.17	0.14	-0.25	-0.09	0.09	0.19	-0.17	0.16	0.11	-0.17
## Progression	-0.17	0.23	-0.25	-0.19	0.13	0.36	-0.24	0.27	0.01	-0.24
## Set_Piece_shot	0.05	0.40	-0.16	0.11	0.48	0.57	-0.10	0.32	0.04	-0.16
## Set_Pieces_Conceded	-0.04	-0.39	0.27	0.04	-0.16	-0.20	0.14	-0.16	0.11	0.48
##	Cnv_C									
## Points										
## Expected_Points										
## Conversion_Points										
## Goals										
## xG										
## Attacks_Shot										
## Conversion										
## Shot_Assist										
## Goals_Conceded										
## xG_Conceded										
## Conversion_Conceded	1.00									
## Shots_Conceded	0.15									
## Contention	-0.18									
## Pressing	0.06									
## Imposition	-0.01									
## PPDA	0.16									
## Possession_Time	0.18									
## Passing_Speed	0.07									
## Ball_Care	0.18									
## Build_Up	0.18									
## Progression	0.17									
## Set_Piece_shot	0.14									
## Set_Pieces_Conceded	-0.10									
##	Sht_C	Cntnt	Prssn	Impst	PPDA	Pss_T	Pss_S	Bll_C	Bld_U	Prgrs
## Shots_Conceded	1.00									
## Contention	0.04	1.00								
## Pressing	-0.26	0.22	1.00							
## Imposition	-0.10	0.32	0.13	1.00						
## PPDA	0.00	-0.26	0.16	-0.36	1.00					
## Possession_Time	-0.25	-0.25	0.33	0.14	-0.26	1.00				
## Passing_Speed	-0.11	-0.16	0.10	0.02	-0.02	0.22	1.00			
## Ball_Care	-0.23	-0.30	0.28	0.04	0.01	0.85	0.48	1.00		
## Build_Up	-0.17	-0.41	0.12	-0.05	0.01	0.67	0.31	0.75	1.00	

```

## Progression          -0.21  0.02  0.43  0.27 -0.12  0.62  0.18  0.49  0.37  1.00
## Set_Piece_shot       -0.12 -0.11  0.16  0.16 -0.04  0.26  0.03  0.16  0.08  0.26
## Set_Pieces_Conceded  0.37  0.07 -0.24 -0.09 -0.15 -0.26 -0.09 -0.31 -0.21 -0.27
##                      St_P_
## Shots_Conceded
## Contention
## Pressing
## Imposition
## PPDA
## Possession_Time
## Passing_Speed
## Ball_Care
## Build_Up
## Progression
## Set_Piece_shot       1.00
## Set_Pieces_Conceded -0.16
## [1] 1.00

```

##Velocidade de circulação (tempo de retenção) x Tempo de posse

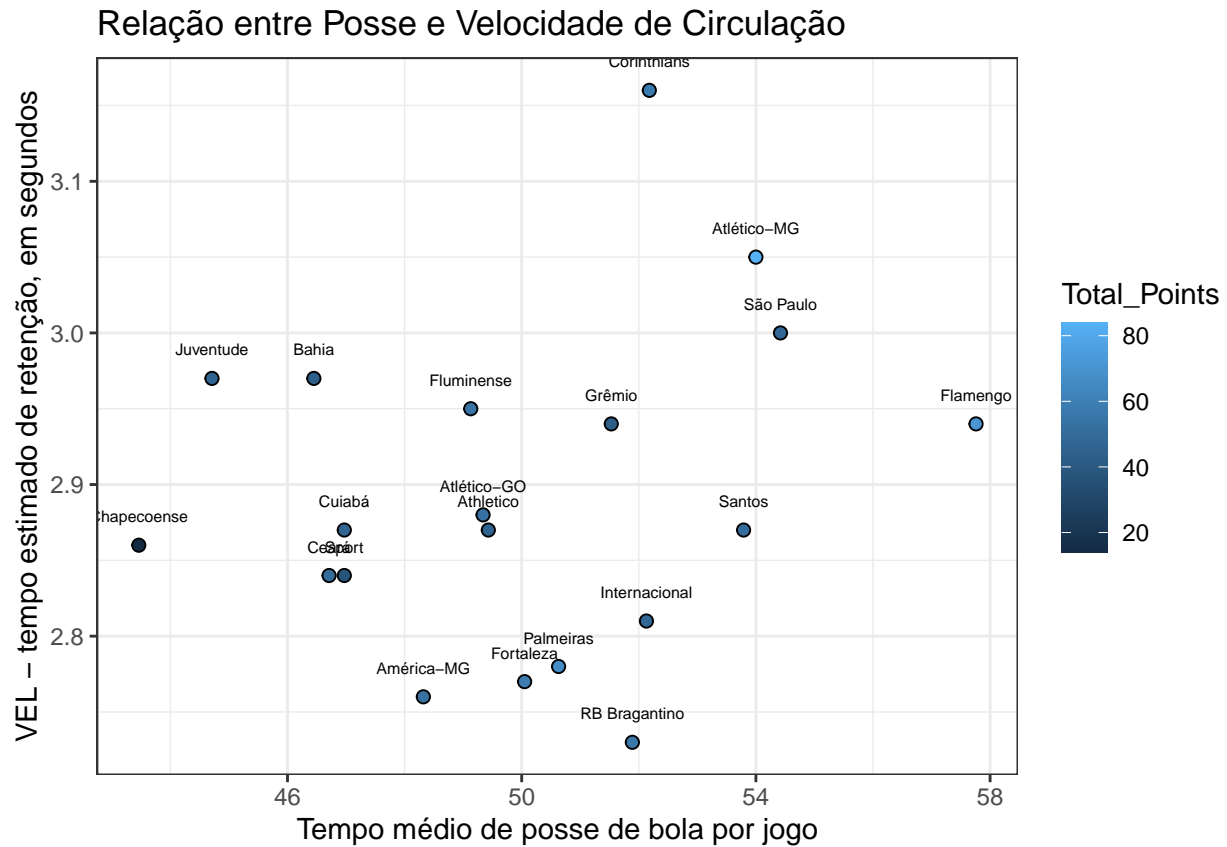
Abriendo um breve parênteses no processo de AFE, apresento um gráfico de dispersão comparando as variáveis numéricas Passing\_Speed (VEL estimada, tempo de retenção, velocidade de circulação, enfim, há várias nomenclaturas possíveis) e Possession\_Time.

Em relação ao código, primeiro extraí apenas as variáveis de interesse e agrupei pelos times, para então executar as funções de estatística descritiva (média de VEL e de posse, e total de pontos).

O campeão Atlético-MG, por exemplo, teve a 2ª retenção mais alta, superado apenas pelo Corinthians, ambos com tempo médio de posse superior a 50% (ou seja, ambos foram dominantes em posse, porém com velocidade baixa de circulação). O São Paulo se aproximou do Atlético-MG nas duas variáveis, porém a performance em pontos foi muito inferior.

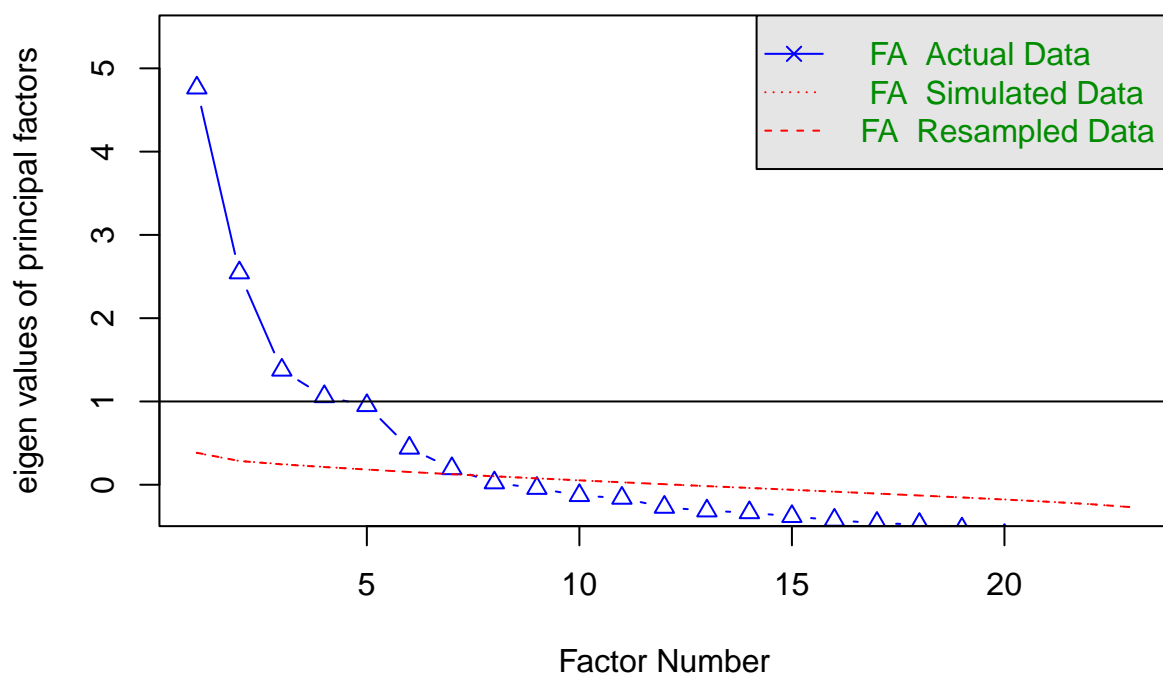
Já a Chapecoense, pior equipe do Brasileiro, teve VEL média estimada em menos de 2.9 segundos (8ª circulação mais ágil), o que não teve nenhum impacto em sua performance prática.

Palmeiras e Bragantino foram as equipes que conseguiram associar agilidade na circulação da bola, tempo de posse superior e boa performance em pontos, enquanto o Fortaleza se aproximou de ambos mas dividiu o controle da bola com os adversários.



Voltando à AFE, após realizar os testes e analisar a matriz de correlações, o passo seguinte é estimar o número de fatores necessários para melhor sintetizar as informações contidas no banco de dados, com a função de análise paralela (**Parallel Analysis**):

## Parallel Analysis Scree Plots



`## Parallel analysis suggests that the number of factors = 7 and the number of components = NA`

A linha preta que delimita o auto-valor igual a 1 é o ponto de corte. O parâmetro é estar acima, logo podemos concluir que **4 fatores** são suficientes para analisar este banco de dados, embora um deles esteja no limite, enquanto um 5º fator ficou abaixo do eixo de corte.

No comando acima, o método de extração escolhido (fm - factor method) foi Maximum Likelihood (ML), depois que uma tentativa com Principal Axis Factor (PA) gerou a mensagem de erro *“The estimated weights for the factor scores are probably incorrect. Try a different factor score estimation method”*.

A rotação escolhida foi oblíqua (Promax) porque assumimos que há correlação entre os fatores, o que é lógico de se imaginar tratando-se de uma análise esportiva (os processos se encadeiam). Caso fosse uma rotação ortogonal, assumiríamos que os fatores não têm relação.

O fator **ML4** explica 32% dos dados; o **ML2** explica 28% dos dados; e os fatores **ML3** e **ML1** explicam 20% dos dados cada, somando a proporção cumulativa de 100%.

O índice **TLI** (Tucker-Lewis Index) chegou a **0.564** (ideal é  $> 0.9$ ) e o **RMSEA** foi de **0.16** (ideal  $< 0.05$ ). Como já foi dito no artigo com a Análise Fatorial Confirmatória, os modelos de indicadores para futebol aparentam se inclinar para resultados baixos nos índices de ajuste, o que suponho esteja relacionado com a imprevisibilidade de um esporte onde nem sempre o melhor vence, e onde temos inúmeros registros de equipes derrotadas com maior posse, mais finalizações, enfim, melhores indicadores que o adversário. O teste estatístico mede as correlações e não consegue, obviamente, enxergar este contexto próprio do futebol.

Inspirado em outras pesquisas publicadas, estabeleci o corte de variáveis com cargas de correlação inferiores a 0.30, e criei uma nova matriz com os quatro fatores sugeridos pela EFA, organizados de forma decrescente (das maiores cargas para as menores):

```
## Factor Analysis using method = ml
## Call: fa(r = Metrics_EFA[, 2:24], nfactors = 4, rotate = "Promax",
##       scores = TRUE, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      item      ML4      ML2      ML3      ML1      h2      u2      com
## Expected_Points      2  1.023                      0.9375 0.0625 1.05
## xG                    5  0.833                      0.6472 0.3528 1.32
## xG_Conceded          10 -0.647                      0.321 0.5517 0.4483 1.54
## Attacks_Shot         6  0.568                      0.4441 0.5559 1.36
## Set_Piece_shot       22  0.488                      0.2312 0.7688 1.20
## Shot_Assist          8  0.482                      0.3208 0.6792 1.26
## Set_Pieces_Conceded  23 -0.351                      0.2238 0.7762 1.66
## Pressing             14                      0.1672 0.8328 1.91
## Imposition           15                      0.0426 0.9574 1.61
## Ball_Care            19          1.041          0.9442 0.0558 1.03
## Possession_Time      17          0.875          0.7822 0.2178 1.00
## Build_Up             20          0.836          0.6037 0.3963 1.08
## Passing_Speed        18          0.539          0.2115 0.7885 1.22
## Progression          21          0.448          0.3535 0.6465 1.52
## Contention           13         -0.316          0.1368 0.8632 1.72
## Goals                4                      0.977          0.9247 0.0753 1.26
## Conversion           7                      0.821          0.6503 0.3497 1.20
## Points              1  0.307          0.627 -0.457 0.8081 0.1919 2.34
## Conversion_Points    3 -0.344          0.391 -0.334 0.4549 0.5451 2.94
## Goals_Conceded       9                      0.996 0.9950 0.0050 1.09
## Conversion_Conceded  11                      0.759 0.5977 0.4023 1.17
## Shots_Conceded      12 -0.339          0.442 0.3678 0.6322 2.22
## PPDA                16                      0.0293 0.9707 1.97
##
##      ML4      ML2      ML3      ML1
## SS loadings      3.645 3.158 2.283 2.340
## Proportion Var    0.158 0.137 0.099 0.102
## Cumulative Var    0.158 0.296 0.395 0.497
## Proportion Explained 0.319 0.276 0.200 0.205
## Cumulative Proportion 0.319 0.595 0.795 1.000
##
## With factor correlations of
##      ML4      ML2      ML3      ML1
## ML4  1.000  0.508 -0.089 -0.027
## ML2  0.508  1.000 -0.290  0.153
## ML3 -0.089 -0.290  1.000 -0.236
## ML1 -0.027  0.153 -0.236  1.000
##
## Mean item complexity = 1.5
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are 253 and the objective function was 15.512 with Chi
## The degrees of freedom for the model are 167 and the objective function was 4.587
##
## The root mean square of the residuals (RMSR) is 0.077
```



```

## The df corrected root mean square of the residuals is  0.095
##
## The harmonic number of observations is  758 with the empirical chi square  2300.254  with prob <  0
## The total number of observations was  758  with Likelihood Chi Square =  3420.997  with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.5644
## RMSEA index =  0.1603  and the 90 % confidence intervals are  0.1558 0.1651
## BIC =  2313.673
## Fit based upon off diagonal values = 0.914
## Measures of factor score adequacy
##
##                                     ML4   ML2   ML3   ML1
## Correlation of (regression) scores with factors  0.977 0.980 0.969 0.996
## Multiple R square of scores with factors         0.955 0.961 0.940 0.992
## Minimum correlation of possible factor scores     0.909 0.923 0.879 0.985

##                                     ML4           ML2           ML3           ML1
## Points                0.25416113 -0.11344867  0.7126391570 -0.61591835
## Expected_Points       0.95106494  0.35655708  0.0385735779 -0.13176342
## Conversion_Points    -0.36406033 -0.32823570  0.4975350223 -0.41522734
## Goals                 0.21798197 -0.06166355  0.8936003286 -0.04257895
## xG                   0.74384452  0.24873047  0.1895269733  0.04902036
## Attacks_Shot         0.63898630  0.42945146  0.0227803707  0.06229070
## Conversion           -0.21313355 -0.20282613  0.7782974194 -0.06095155
## Shot_Assist          0.54865349  0.37914228 -0.0267006341  0.04861725
## Goals_Conceded      -0.20076291  0.04569426 -0.1107286214  0.97626610
## xG_Conceded         -0.67376112 -0.32789130  0.0988276977  0.30971960
## Conversion_Conceded  0.19165188  0.21334472 -0.1480595169  0.74191186
## Shots_Conceded      -0.42765786 -0.26846204  0.0709142770  0.40609861
## Contention          -0.07902432 -0.31149223  0.1589691043 -0.22953369
## Pressing             0.37537897  0.31723736 -0.0494055860 -0.04808751
## Imposition           0.18609037  0.07450117 -0.0626795222 -0.06245391
## PPDA                0.02628921 -0.02726429 -0.0004086895  0.14812582
## Possession_Time     0.47097959  0.88393710 -0.2406912898  0.12567933
## Passing_Speed       0.11185001  0.43339614 -0.0498987266  0.05691614
## Ball_Care           0.42732275  0.96445414 -0.1955713443  0.13013306
## Build_Up            0.26235982  0.76213162 -0.2460919119  0.17505719
## Progression         0.34726171  0.55712838 -0.3413620741  0.10554053
## Set_Piece_shot       0.45886267  0.22013401 -0.0500048744  0.12414132
## Set_Pieces_Conceded -0.43392305 -0.34380362  0.1821869259  0.01927034

```

Os nomes dos fatores são determinados pelo pesquisador a partir da análise dos resultados da análise exploratória, ou seja, da identificação das variáveis que têm as maiores cargas em cada um deles. Caso não veja necessidade de nomeá-los, o pesquisador pode apenas manter o padrão da EFA - neste exemplo, os fatores estão batizados como ML, de Maximum Likelihood.

Se esta pesquisa fosse levada adiante e implementada como um método de trabalho, o fator **ML4** poderia ser chamado de **Expectativa**, pois nele as maiores cargas estão relacionadas com **Expected\_Points**, **xG** e **xG\_Conceded**.

O fator **ML2** encontrou a correlação entre as variáveis de **Posse**, e assim poderia ser chamado. As cargas são altas para **Ball\_Care**, **Possession\_Time** e **Build\_Up**. Mais uma vez nota-se que a variável **Passing\_Speed** não alcançou correlação próxima às demais (0.43, contra 0.96 de **Ball\_Care**, por exemplo).

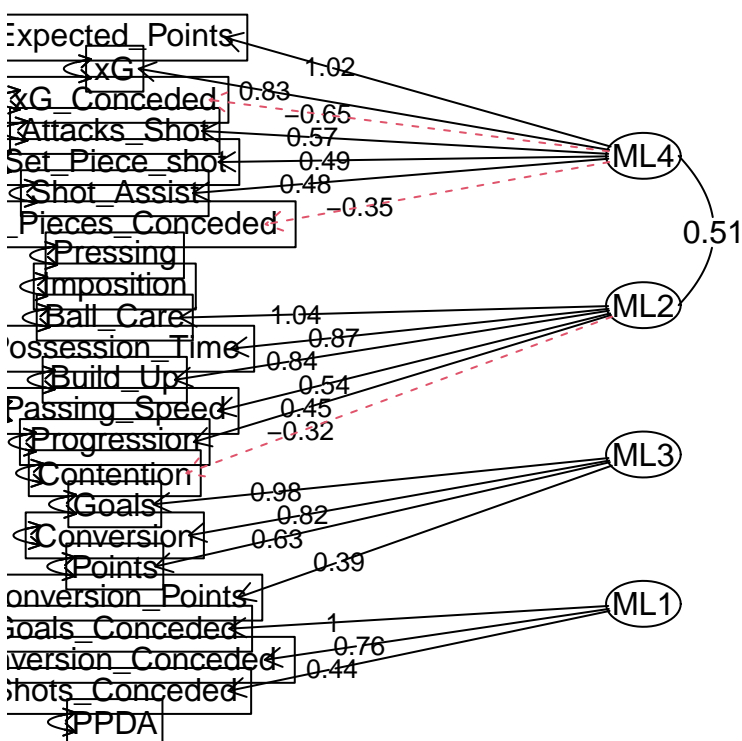
O fator **ML3**, ao contrário do ML1, não trata de expectativa, e sim de **Eficiência**, ao correlacionar as variáveis com maior influência para os pontos conquistados, sendo que apenas **Goals e Conversion** apresentam cargas altas, o que valida o dito popular “futebol é bola na rede”. Muitas variáveis tiveram cargas próximas a zero, ou seja, mostraram-se quase sem influência para os pontos conquistados no Campeonato Brasileiro.

E o fator **ML1** é o inverso do ML3, ou seja, ele carrega as variáveis que mais influenciaram na **perda de pontos**. Nota-se que despontam apenas as cargas de **Goals\_Conceded e Conversion\_Conceded**. Como poderia ser nomeado este fator? Talvez seria o caso de chamar o **ML3** de **Eficiência Ofensiva** e o **ML1** de **Eficiência Defensiva**.

Ao contrário da Análise Fatorial Confirmatória, onde o modelo privilegia aspectos estilísticos, para identificar padrões de comportamento relacionados com o modelo de jogo ds equipes, a Análise Fatorial Exploratória encontrou correlações mais fortes no que diz respeito às conversões (seja de gols feitos ou sofridos, seja de pontos), sendo que apenas o fator **ML2** assemelha-se ao modelo testado na AFC, ao correlacionar as variáveis de posse.

Abaixo apresento o diagrama do modelo resultante:

**Factor Diagram**



#### ##Ranking

Assim como na análise confirmatória, também é possível rankear as observações (neste caso, as equipes) na análise exploratória. Para isso, mantive a simplicidade matemática, sem criar pesos com multiplicações ou exponenciações. O ranking levará em conta a soma dos 3 fatores positivos (ML4, ML2 e ML3) subtraídos pelo ML1, que tem carga inversa (quanto menor a pontuação obtida, melhor, afinal ele encontra as correlações entre pontos e gols/conversão concedidos).

Comparar o resultado da análise confirmatória (cujo link está no 1º parágrafo) e desta análise exploratória é muito interessante e divertido. Novamente o Flamengo liderou em todos os fatores à exceção da posse, mas desta vez a pontuação nos demais fatores foi suficiente para manter a equipe à frente do campeão Atlético-MG.

Chama a atenção o rebaixado Grêmio, que terminou em 7º no ranking geral, e foi o 3º com a maior Expectativa - mais uma vez a ineficiência defensiva da equipe demonstra que este foi o fator determinante para a queda.

Outro rebaixado, o Bahia, também criou expectativa suficiente para escapar (13º), mas foi o 3º pior em todos os demais fatores - ou seja, ineficiente em todos os processos. O Sport Recife teve situação inversa: destacou-se pela eficiência defensiva (5º) e ficou em posições intermediárias em ataque e posse, porém terminou em 19º no ranking devido à baixíssima expectativa criada - mesmo obtendo performance acima do esperado, não foi o suficiente para evitar o rebaixamento.

Na outra ponta da tabela, o São Paulo foi a equipe com a maior diferença entre a classificação no Brasileiro (idêntica à análise confirmatória) e o ranking na análise exploratória (5º melhor, 4º em posse e em ataque, 6º em expectativa e 12º em defesa), indicando que poderia ter obtido melhor colocação no campeonato. Por outro lado, o Fortaleza foi uma equipe que criou alta expectativa (4º maior) e conseguiu confirmar em pontos, que o levaram para a Libertadores mesmo terminando em 11º no ranking da AFE, o que se deve principalmente à baixa eficiência defensiva (17º).

▲	Teams	Ranking_Total	Expectancy	Possession	Efficiency_Off	Efficiency_Def
1	Flamengo	75.42	25.13	30.89	30.89	-3.70
2	Atlético-MG	69.34	18.25	24.63	24.63	-7.21
3	Corinthians	34.71	1.38	28.19	28.19	-5.91
4	Palmeiras	17.86	4.65	-0.62	-0.62	0.15
5	São Paulo	10.52	5.84	16.02	16.02	-0.87
6	RB Bragantino	5.51	-5.37	-0.88	-0.88	1.37
7	Grêmio	5.20	12.18	9.50	9.50	11.05
8	Santos	4.51	-0.92	9.85	9.85	-1.69
9	Fluminense	2.47	2.76	-0.10	-0.10	-3.13
10	Internacional	-0.61	-7.27	3.31	3.31	-1.50
11	Fortaleza	-4.20	7.16	-7.05	-7.05	4.02
12	Athletico	-4.88	-0.31	-0.13	-0.13	3.00
13	América-MG	-5.66	6.99	-13.09	-13.09	-3.20
14	Atlético-GO	-12.25	-15.17	-5.05	-5.05	-8.42
15	Cuiabá	-19.00	-11.49	-11.98	-11.98	-6.42
16	Ceará	-19.17	-1.65	-20.74	-20.74	-3.86
17	Juventude	-21.13	-2.62	-12.94	-12.94	1.97
18	Bahia	-25.21	-1.66	-13.63	-13.63	8.50
19	Sport	-28.07	-12.13	-9.04	-9.04	-5.32
20	Chapecoense	-85.37	-25.77	-27.11	-27.11	21.19

São muitas as considerações possíveis no cruzamento dos resultados das duas análises apresentadas, assim como há outros tantos aspectos inexplorados para não tornar este

artigo ainda mais extenso e cansativo. Em breve publicarei os resultados de uma regressão múltipla baseada no mesmo banco de dados.