

Brasileiro_2021_CFA

Eduardo Cecconi

1/4/2022

Análise Fatorial Confirmatória

Dados do Campeonato Brasileiro 2021 extraídos do InStat

Como parte do treinamento em programação e em estatística, submeti dados de scout de 379 dos 380 jogos do Campeonato Brasileiro 2021 (o InStat não disponibilizou os dados do jogo Chapecoense 1x1 Athletico, pela 26ª rodada) a diversas técnicas, a começar pela Análise Fatorial Confirmatória.

O primeiro passo foi realizar o data wrangling para estruturar o banco de dados e proporcionar a elaboração das métricas e dos indicadores que posteriormente passarão pelo processo de validação.

Originalmente com 115 variáveis, constatei a ausência de pelo menos 4 informações relevantes e fáceis de serem adicionadas ao banco, com o uso da função **case_when()**, a partir da criação de uma variável chave para facilitar o código - identificar em cada jogo o mandante (H, de Home), e o visitante (A, de Away):

- Pontos (Points)
- Gols Concedidos (Goals_Conceded)
- Chutes Concedidos (Shots_Conceded)
- Finalizações Concedidas em Bolas Paradas (Set_Pieces_Conceded)

Das 120 variáveis resultantes, filtrei apenas 39 para construir as métricas, em um novo banco de dados, onde foram transformadas em 12 indicadores-chave em um arquivo com 24 variáveis no total. Os termos utilizados no código estão em inglês (como **CFA**, de **Confirmatory Factor Analysis**), para facilitar a futura tradução do artigo.

Em resumo, estes são os conceitos utilizados na elaboração das métricas:

- **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.

Como as correlações, variâncias e covariâncias são a base deste método, podemos explorar as respectivas tabelas, aplicadas ao novo banco de dados:

Além disso, também é importante ajustar as relações de medidas entre métricas que resultaram em escalas distintas, o que poderia causar problemas para interpretar os resultados, fechando com o arredondamento para duas casas decimais.

Serão testadas **quatro variáveis latentes** (os indicadores-chave, ou fatores) em separado, e depois todas agrupadas em um único modelo: **Ataque, Defesa, Posse e Bolas Paradas**. A aritmética é simples, apenas somas, porque carecem de maior estudo estatístico para envolverem pesos atribuídos com multiplicações, etc.

ATAQUE

O modelo de ataque relaciona quatro variáveis: **gols marcados, conversão, chutes e assistências para finalização**.

Os índices de ajuste foram baixos: CFI = 0.696 (ideal acima de 0.9); TLI = 0.088 (ideal acima de 0.9); RMSEA = 0.435 (ideal entre 0.05 e 0.08) e SRMR = 0.183 (ideal inferior a 0.05).

Não é motivo para desânimo, entretanto. Futebol, sabe-se, é um esporte onde a imprevisibilidade interfere mais do que em outros esportes, e muito mais se comparado a dados oriundos do mundo empresarial. Muitas vezes a elaboração de indicadores no futebol não é relevante para prever resultados, e sim para **identificar padrões de comportamento**, para **identificar o estilo** de cada equipe, e aí sim agregar informação relevante para a comissão técnica basear decisões.

```
## lavaan 0.6-9 ended normally after 100 iterations
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters      8
##
##      Number of observations          758
##
## Model Test User Model:
##
##      Test statistic                288.719
##      Degrees of freedom              2
##      P-value (Chi-square)           0.000
##
## Model Test Baseline Model:
##
##      Test statistic                949.481
##      Degrees of freedom              6
##      P-value                        0.000
##
## User Model versus Baseline Model:
##
##      Comparative Fit Index (CFI)    0.696
##      Tucker-Lewis Index (TLI)      0.088
##
## Loglikelihood and Information Criteria:
##
##      Loglikelihood user model (H0)   -5029.148
##      Loglikelihood unrestricted model (H1) -4884.789
##
##      Akaike (AIC)                   10074.297
##      Bayesian (BIC)                  10111.342
##      Sample-size adjusted Bayesian (BIC) 10085.939
##
## Root Mean Square Error of Approximation:
##
##      RMSEA                          0.435
##      90 Percent confidence interval - lower 0.393
##      90 Percent confidence interval - upper 0.478
##      P-value RMSEA <= 0.05           0.000
##
## Standardized Root Mean Square Residual:
##
```

```

##      SRMR                                0.183
##
## Parameter Estimates:
##
##      Standard errors                                Standard
##      Information                                Expected
##      Information saturated (h1) model            Structured
##
## Latent Variables:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      Attack =~
##      Goals      1.000
##      Conversion -0.311  0.205  -1.515  0.130    NA    NA
##      Attacks_Shot -0.086  0.047  -1.842  0.065    NA    NA
##      Shot_Assist -0.113  0.066  -1.699  0.089    NA    NA
##
## Variances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      .Goals      3.577  1.682   2.127  0.033   3.577   3.255
##      .Conversion  1.302  0.175   7.434  0.000   1.302   1.226
##      .Attacks_Shot 1.017  0.054  18.904  0.000   1.017   1.018
##      .Shot_Assist 13.991  0.721  19.417  0.000  13.991   1.002
##      Attack     -2.478  1.662  -1.491  0.136    NA    NA

```

DEFESA

O modelo de defesa relaciona seis variáveis: **gols sofridos, conversão concedida, pressão, contenção, chutes permitidos e PPDA**. Comparado ao ataque, a variável latente defensiva alcançou índices de ajuste melhores: CFI = 0.787 (ideal acima de 0.9); TLI = 0.645 (ideal acima de 0.9); RMSEA = 0.179 (ideal entre 0.05 e 0.08) e SRMR = 0.108 (ideal inferior a 0.05). Não houve registro de variância negativa.

```

## lavaan 0.6-9 ended normally after 26 iterations
##
##      Estimator                                ML
##      Optimization method                    NLMINB
##      Number of model parameters              12
##
##      Number of observations                  758
##
## Model Test User Model:
##
##      Test statistic                226.962
##      Degrees of freedom              9
##      P-value (Chi-square)           0.000
##
## Model Test Baseline Model:
##
##      Test statistic                1038.594
##      Degrees of freedom             15
##      P-value                        0.000
##
## User Model versus Baseline Model:

```

```

##
## Comparative Fit Index (CFI) 0.787
## Tucker-Lewis Index (TLI) 0.645
##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0) -6104.384
## Loglikelihood unrestricted model (H1) -5990.903
##
## Akaike (AIC) 12232.768
## Bayesian (BIC) 12288.336
## Sample-size adjusted Bayesian (BIC) 12250.231
##
## Root Mean Square Error of Approximation:
##
## RMSEA 0.179
## 90 Percent confidence interval - lower 0.159
## 90 Percent confidence interval - upper 0.199
## P-value RMSEA <= 0.05 0.000
##
## Standardized Root Mean Square Residual:
##
## SRMR 0.108
##
## Parameter Estimates:
##
## Standard errors Standard
## Information Expected
## Information saturated (h1) model Structured
##
## Latent Variables:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## Defense =~
## Goals_Conceded 1.000 1.401 1.336
## Pressing -0.045 0.018 -2.501 0.012 -0.064 -0.064
## Conversn_Cncdd 0.384 0.051 7.585 0.000 0.538 0.522
## Contention -0.082 0.021 -3.910 0.000 -0.114 -0.115
## PPDA 0.047 0.018 2.590 0.010 0.066 0.066
## Shots_Conceded 0.268 0.041 6.575 0.000 0.375 0.375
##
## Variances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Goals_Conceded -0.863 0.229 -3.772 0.000 -0.863 -0.785
## .Pressing 0.995 0.051 19.512 0.000 0.995 0.996
## .Conversn_Cncdd 0.773 0.051 15.048 0.000 0.773 0.728
## .Contention 0.985 0.050 19.594 0.000 0.985 0.987
## .PPDA 0.994 0.051 19.515 0.000 0.994 0.996
## .Shots_Conceded 0.858 0.047 18.458 0.000 0.858 0.859
## Defense 1.962 0.227 8.640 0.000 1.000 1.000

```

POSSE

O modelo de posse relaciona seis variáveis: **tempo de posse, velocidade de circulação, proteção, construção, progressão e imposição**. Temos uma tendência natural a associar o conceito de imposição com a defesa, mas como os dados envolvem todos os tipos de ações que o InStat considera “duelos” (muitos dos duelos ofensivos e aéreos não têm relação com ações defensivas, portanto), esta métrica serve como referência de obtenção de posse.

Embora novamente sem atingir os índices ideais de ajuste, o desempenho no teste também melhorou (comparado à defesa, à exceção do RMSEA):
CFI = 0.833 (ideal acima de 0.9); TLI = 0.721 (ideal acima de 0.9); RMSEA = 0.242 (ideal entre 0.05 e 0.08) e SRMR = 0.081 (ideal inferior a 0.05).

```
## lavaan 0.6-9 ended normally after 41 iterations
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters      12
##
##      Number of observations          758
##
## Model Test User Model:
##
##      Test statistic                  410.171
##      Degrees of freedom                9
##      P-value (Chi-square)             0.000
##
## Model Test Baseline Model:
##
##      Test statistic                  2414.314
##      Degrees of freedom               15
##      P-value                          0.000
##
## User Model versus Baseline Model:
##
##      Comparative Fit Index (CFI)      0.833
##      Tucker-Lewis Index (TLI)         0.721
##
## Loglikelihood and Information Criteria:
##
##      Loglikelihood user model (H0)     -2417.196
##      Loglikelihood unrestricted model (H1) -2212.110
##
##      Akaike (AIC)                     4858.391
##      Bayesian (BIC)                    4913.960
##      Sample-size adjusted Bayesian (BIC) 4875.855
##
## Root Mean Square Error of Approximation:
##
##      RMSEA                            0.242
##      90 Percent confidence interval - lower 0.223
##      90 Percent confidence interval - upper 0.263
##      P-value RMSEA <= 0.05             0.000
```

```

##
## Standardized Root Mean Square Residual:
##
##   SRMR                                0.081
##
## Parameter Estimates:
##
##   Standard errors                Standard
##   Information                    Expected
##   Information saturated (h1) model Structured
##
## Latent Variables:
##
##           Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
## Possession =~
##   Possession_Tim    1.000
##   Passing_Speed     1.258    0.095   13.202   0.000    0.095   0.453
##   Ball_Care         12.861    0.339   37.924   0.000    0.976   0.977
##   Build_Up          10.055    0.379   26.526   0.000    0.763   0.764
##   Progression        6.905    0.440   15.700   0.000    0.524   0.524
##   Imposition         0.756    0.487    1.552   0.121    0.057   0.057
##
## Variances:
##
##           Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
##   .Possession_Tim    0.002    0.000   14.051   0.000    0.002   0.239
##   .Passing_Speed     0.035    0.002   19.250   0.000    0.035   0.794
##   .Ball_Care         0.046    0.015    3.133   0.002    0.046   0.046
##   .Build_Up          0.416    0.023   17.745   0.000    0.416   0.417
##   .Progression       0.724    0.038   19.134   0.000    0.724   0.725
##   .Imposition        0.995    0.051   19.465   0.000    0.995   0.997
##   Possession         0.006    0.000   14.841   0.000    1.000   1.000

```

BOLAS PARADAS

Este modelo tem a relação entre apenas duas métricas: *finalizações realizadas e concedidas em bolas paradas*. Embora seja um item muito relevante sob a ótica dos treinadores, sempre atentos a padrões de bolas paradas que possam ser prevenidos ou aproveitados nas partidas, estatisticamente o modelo não conseguiu ser identificado.

Para o fator ser testado, foi preciso restringir as cargas com o uso do **a**, um conceito com o qual ainda não estou familiarizado, e que retira potencial informativo deste indicador. Por enquanto, com os dados disponíveis no banco extraído para realizar este teste, das bolas paradas não se consegue obter informação relevante.

```

## lavaan 0.6-9 ended normally after 13 iterations
##
##   Estimator                ML
##   Optimization method      NLMINB
##   Number of model parameters      3
##
##   Number of observations          758
##
## Model Test User Model:
##
##   Test statistic              0.000

```

```

## Degrees of freedom                                0
##
## Model Test Baseline Model:
##
## Test statistic                                    18.626
## Degrees of freedom                                1
## P-value                                            0.000
##
## User Model versus Baseline Model:
##
## Comparative Fit Index (CFI)                      1.000
## Tucker-Lewis Index (TLI)                        1.000
##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0)                    -3032.736
## Loglikelihood unrestricted model (H1)            -3032.736
##
## Akaike (AIC)                                     6071.472
## Bayesian (BIC)                                   6085.364
## Sample-size adjusted Bayesian (BIC)             6075.838
##
## Root Mean Square Error of Approximation:
##
## RMSEA                                             0.000
## 90 Percent confidence interval - lower           0.000
## 90 Percent confidence interval - upper           0.000
## P-value RMSEA <= 0.05                            NA
##
## Standardized Root Mean Square Residual:
##
## SRMR                                             0.000
##
## Parameter Estimates:
##
## Standard errors                                Standard
## Information                                    Expected
## Information saturated (h1) model                Structured
##
## Latent Variables:
##
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## Set_Pieces =~
## Set_Pc_sht (a) 1.000 NA NA
## St_Pcs_Cnc (a) 1.000 NA NA
##
## Variances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Set_Piece_shot 3.744 0.225 16.662 0.000 3.744 1.156
## .Set_Pics_Cncdd 3.744 0.225 16.662 0.000 3.744 1.156
## Set_Pieces -0.505 0.119 -4.238 0.000 NA NA

```


MODELO FINAL

Com a irrelevância estatística do modelo de bolas paradas, devido à superficialidade dos dados disponíveis, este fator foi excluído do modelo final, evitando problemas na execução do código.

A ordem das métricas dentro dos indicadores também foi alterada, porque na definição dos modelos a primeira variável recebe carga padrão = 1, e se tornou importante que os gols feitos e sofridos não passassem por este filtro.

```
## lavaan 0.6-9 ended normally after 204 iterations
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters    35
##
##      Number of observations        758
##
## Model Test User Model:
##
##      Test statistic                2468.783
##      Degrees of freedom            101
##      P-value (Chi-square)          0.000
##
## Model Test Baseline Model:
##
##      Test statistic                5952.620
##      Degrees of freedom            120
##      P-value                       0.000
##
## User Model versus Baseline Model:
##
##      Comparative Fit Index (CFI)    0.594
##      Tucker-Lewis Index (TLI)       0.518
##
## Loglikelihood and Information Criteria:
##
##      Loglikelihood user model (H0)   -13547.078
##      Loglikelihood unrestricted model (H1) -12312.687
##
##      Akaike (AIC)                  27164.157
##      Bayesian (BIC)                 27326.231
##      Sample-size adjusted Bayesian (BIC) 27215.091
##
## Root Mean Square Error of Approximation:
##
##      RMSEA                          0.176
##      90 Percent confidence interval - lower 0.170
##      90 Percent confidence interval - upper 0.182
##      P-value RMSEA <= 0.05           0.000
##
## Standardized Root Mean Square Residual:
##
##      SRMR                           0.164
```

```

##
## Parameter Estimates:
##
## Standard errors          Standard
## Information              Expected
## Information saturated (h1) model Structured
##
## Latent Variables:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## Attack =~
##   Conversion      1.000          NA      NA
##   Goals           -3.404      2.082    -1.635    0.102      NA      NA
##   Attacks_Shot     0.307      0.051     6.037    0.000      NA      NA
##   Shot_Assist      0.411      0.113     3.631    0.000      NA      NA
## Defense =~
##   Conversn_Cncdd   1.000          0.538    0.522
##   Goals_Conceded   2.598      0.339     7.666    0.000     1.397    1.333
##   Pressing         -0.127      0.044    -2.866    0.004    -0.068   -0.068
##   Contention       -0.205      0.045    -4.535    0.000    -0.110   -0.110
##   PPDA              0.128      0.044     2.898    0.004     0.069    0.069
##   Shots_Conceded   0.704      0.056    12.570    0.000     0.379    0.379
## Possession =~
##   Possession_Tim   1.000          0.076    0.874
##   Passing_Speed     1.249      0.095    13.116    0.000     0.095    0.451
##   Ball_Care        12.813      0.337    38.017    0.000     0.974    0.975
##   Build_Up         10.047      0.378    26.591    0.000     0.764    0.764
##   Progression       6.933      0.439    15.808    0.000     0.527    0.527
##   Imposition        0.771      0.487     1.585    0.113     0.059    0.059
## kpi =~
##   Possession        1.000          0.408    0.408
##   Attack            1.265      2.196     0.576    0.565      NA      NA
##   Defense           1.107      1.688     0.655    0.512     0.064    0.064
##
## Variances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Conversion        1.286      0.155     8.284    0.000     1.286    1.210
## .Goals              3.689      1.637     2.254    0.024     3.689    3.356
## .Attacks_Shot       1.020      0.054    18.815    0.000     1.020    1.021
## .Shot_Assist       13.997      0.721    19.406    0.000    13.997    1.003
## .Conversn_Cncdd     0.773      0.051    15.137    0.000     0.773    0.728
## .Goals_Conceded    -0.854      0.225    -3.801    0.000    -0.854   -0.777
## .Pressing           0.995      0.051    19.518    0.000     0.995    0.995
## .Contention         0.986      0.050    19.586    0.000     0.986    0.988
## .PPDA               0.994      0.051    19.519    0.000     0.994    0.995
## .Shots_Conceded     0.856      0.046    18.448    0.000     0.856    0.857
## .Possession_Tim     0.002      0.000    13.955    0.000     0.002    0.237
## .Passing_Speed      0.035      0.002    19.244    0.000     0.035    0.796
## .Ball_Care          0.049      0.015     3.381    0.001     0.049    0.050
## .Build_Up           0.415      0.023    17.719    0.000     0.415    0.416
## .Progression        0.721      0.038    19.114    0.000     0.721    0.722
## .Imposition         0.995      0.051    19.465    0.000     0.995    0.997
## .Attack             -0.225      0.137    -1.644    0.100      NA      NA
## .Defense             0.288      0.049     5.885    0.000     0.996    0.996
## .Possession          0.005      0.002     2.899    0.004     0.833    0.833

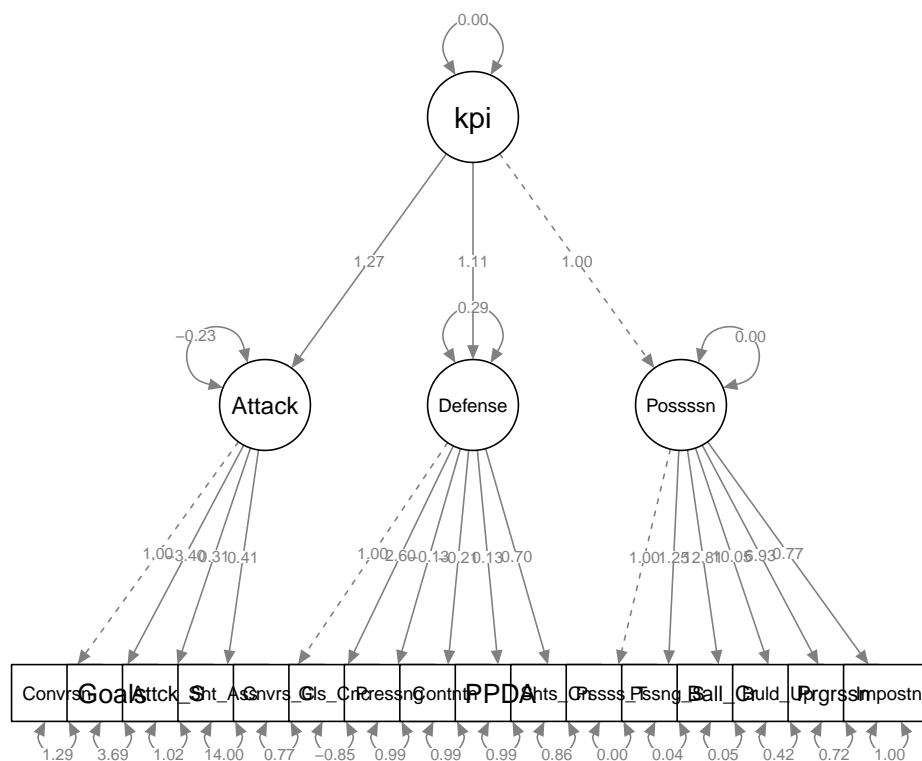
```

kpi 0.001 0.002 0.588 0.557 1.000 1.000

Como esperado, os índices de ajuste ficaram baixos, o que não invalida essa metodologia de análise de indicadores aplicada ao futebol. É preciso ampliar o estudo sobre as possibilidades matemáticas estatisticamente amparadas na formulação das relações entre as métricas que compõem os indicadores, atribuindo diferentes pesos; e também é importante reiterar a análise como instrumento de identificação de padrões e de estilos de jogo, sem deixar de olhar para volumes e eficiências.

Índice	Resultado	Parâmetro
CFI	594	> 0.9
TLI	0.518	> 0.9
RMSEA	0.176	< 0.05
SRMR	0.164	< 0.08

A visualização do modelo é simples, com os indicadores de primeira ordem ataque, posse e defesa relacionados com o KPI (Key Performance Indicator), que soma todos os demais:



Ranking

Apesar de a AFC não ter encontrado indicadores altos de relevância para o modelo, o objetivo da análise fatorial neste caso, desde o início, foi a criação de um ranking das equipes do Brasileiro 2021. Para isto, os fatores latentes precisam ser construídos não apenas apontando as relações (como ocorreu no modelo), mas também seguindo as operações matemáticas - na defesa, por exemplo, os gols sofridos, o perigo concedido e os chutes permitidos subtraem o valor total das ações defensivas como pressão, contenção e PPDA. E, devido à importância do tema para a análise de comportamentos, incluí o indicador de bolas paradas no cálculo, contrariando o modelo avaliado e ignorando a superficialidade dos dados.

E, para minha surpresa, **o ranking resultante** de todo este processo apresentado através da análise fatorial confirmatória é **exatamente igual à colocação final das 20 equipes do Brasileiro**:

▲	Teams	KPI_Total	Attack_Total	Defense_Total	Possession_Total	Set_Pieces
1	Atlético-MG	25.50	18.08	6.49	23.74	15.65
2	Flamengo	15.31	22.99	-4.44	25.62	23.52
3	Palmeiras	11.40	4.67	9.73	15.13	14.41
4	Fortaleza	5.13	7.31	-4.21	1.80	11.24
5	Corinthians	4.33	1.76	-2.43	-2.00	-8.05
6	RB Bragantino	3.54	-5.14	8.81	12.28	-1.65
7	Fluminense	1.99	4.57	-5.94	-3.94	1.34
8	América-MG	1.19	6.97	-5.22	-1.08	5.94
9	Atlético-GO	1.19	-13.06	17.62	-8.69	-14.90
10	Santos	-1.16	-2.22	-1.40	-6.81	-5.67
11	Ceará	-1.17	0.29	8.88	-2.96	4.29
12	Internacional	-2.72	-9.62	0.29	1.81	-1.69
13	São Paulo	-2.73	3.91	-7.63	-10.61	2.28
14	Athletico	-3.21	-0.91	-2.10	-0.98	-3.94
15	Cuiabá	-3.52	-9.04	14.01	-7.76	-15.22
16	Juventude	-4.29	-0.66	-2.36	-5.80	-6.50
17	Grêmio	-6.62	9.53	-13.73	1.78	5.73
18	Bahia	-6.63	-0.33	-8.44	-0.15	-7.08
19	Sport	-10.55	-12.06	-0.44	-17.30	-10.65
20	Chapecoense	-28.29	-26.46	-7.96	-14.33	-9.08

Mesmo sem levar em consideração itens como pontos e expectativa de pontos, que serão abordados em artigo exercitando a regressão múltipla, e apesar dos índices baixos nos testes de ajuste dos fatores, o ranking bate com a classificação do 1º ao 20º.

É curioso como o Flamengo lidera 3 dos 4 indicadores-chave, à exceção da defesa, o que foi determinante para o Atlético-MG terminar em 1º no ranking. Outro aspecto interessante do ranking é observado no Grêmio: 3º em ataque, 6º em bolas paradas e 7º em posse, mas o 20º lugar em defesa o colocou em 17º no ranking geral.

Gráficos e demais análises das variáveis originais e das métricas, proporcionando mais subsídios para a análise do desempenho das equipes do Brasileiro 2021, serão também disponibilizados em **artigos onde testarei não apenas a regressão múltipla, mas também a análise fatorial exploratória**.