

**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES**

**ADA MARIA CAYRES FERNANDEZ
BEATRIZ THOMPSON SATHLER FREITAS
LEONARDO COPPOLA BIAZUCCI
MICHELLY RODRIGUES DA SILVA
THIAGO DE OLIVEIRA DEODATO
RAFAEL MARTINS CAMPOS**

T3 - TÉCNICA DE DEPENDÊNCIA: ANÁLISE DE VARIÂNCIA

**SÃO PAULO - SP
2021**

SUMÁRIO

1. INTRODUÇÃO	3
2. OBJETIVOS	3
3. PRÉ-PROCESSAMENTO	4
4. APLICAÇÃO DA ANÁLISE DE VARIÂNCIA	5
5. TESTE DE BARTLETT	10
6. APLICANDO A ANOVA	11
7. INTERPRETAÇÃO DOS RESULTADOS	12
8. CONCLUSÃO	13
9. REFERÊNCIAS	13

1. INTRODUÇÃO

Será feito neste trabalho o uso da técnica de dependência conhecida como análise de variância sobre um recorte de banco de dados que diz respeito a partidas jogadas em *League of Legends*, jogo online famoso mundialmente. O uso da análise de variância nos permitirá descobrir ao final da análise se há alguma diferença no desempenho entre os dois possíveis lados (azul ou vermelho) de uma partida do jogo. Esta informação poderá nos dar uma vantagem significativa no ambiente competitivo.

A análise será feita utilizando códigos na linguagem *Python* com auxílio das bibliotecas *pandas*, *numpy*, *matplotlib*, *pylab*, *scipy*, *statsmodels*, *bioinfokit*. A execução dos códigos será realizada no *Google Colaboratory* (ou simplesmente *Colab*).

2. OBJETIVOS

Este documento tem como objetivo principal o estudo da relação entre as variáveis do banco de dados dos primeiros 10 minutos das partidas do jogo online conhecido mundialmente, *League of Legends*. Com o uso da técnica da análise de variância, espera-se descobrir ao final da análise se há alguma diferença no desempenho entre os dois possíveis lados (azul ou vermelho) de uma partida do jogo. Esta análise é de profunda importância quando se está em um ambiente competitivo, pois ao saber se há alguma diferença entre os times, este conhecimento poderá servir como vantagem ao jogador.

Esse tipo de informação é algo que remete as chances e probabilidades de cada time vencer e isso foge um pouco do "meta" do jogo. O "meta" é o padrão de ações mais eficiente e que deverá ser seguido durante a partida. Sendo assim, não estão em questão ações no jogo e sim o time o qual o jogador estará. Desta forma, podemos dizer que a obtenção desse conhecimento é ainda mais eficaz, pois não é uma informação tão simples de ser obtida pelo jogador comum.

Neste documento, temos como foco descobrir se há alguma diferença no desempenho entre os dois possíveis lados (azul ou vermelho) de uma partida do jogo.

Assim, saberemos se há algum desbalanceamento entre as equipes e, em caso positivo, observar qual equipe tem vantagem.

3. PRÉ-PROCESSAMENTO

Antes de iniciar a análise, criamos o *data frame* (recorte do banco de dados) apenas com as onze colunas de interesse. Em seguida, traduzimos do inglês para o português do Brasil o nome das dez variáveis envolvidas na análise.

Note que pegamos 5 variáveis escolhidas do time azul no início do projeto, também para o time vermelho.

```
url_lol = 'https://raw.githubusercontent.com/michellyrds/mqaa/master/datasets/high_diamond_ranked_10min.csv'

url = url_lol
raw_data = pd.read_csv(url)

df = raw_data.filter(['blueWardsPlaced', 'blueKills', 'blueDeaths', 'blueEliteMonsters', 'blueGoldDiff',
                     'redWardsPlaced', 'redKills', 'redDeaths', 'redEliteMonsters', 'redGoldDiff', 'gameId'], axis=1)

df = df.rename(columns={'blueWardsPlaced' : 'azulSentinelas',
                        'blueKills' : 'azulAbates',
                        'blueDeaths' : 'azulMortes',
                        'blueEliteMonsters' : 'azulMonstrosEpicos',
                        'blueGoldDiff' : 'azulDiferencaOuros',
                        'redWardsPlaced' : 'vermelhoSentinelas',
                        'redKills' : 'vermelhoAbates',
                        'redDeaths' : 'vermelhoMortes',
                        'redEliteMonsters' : 'vermelhoMonstrosEpicos',
                        'redGoldDiff' : 'vermelhoDiferencaOuros'})

raw_data.var()
```

Figura 1: linhas de código do pré-processamento do banco de dados.

	azulSentinelas	azulAbates	azulMortes	azulMonstrosEpicos	azulDiferencaOuros	vermelhoSentinelas	vermelhoAbates	vermelhoMortes	vermelhoMonstrosEpicos	vermelhoDiferencaOuros
0	28	9	6	0	643	15	6	9	0	-643
1	12	5	5	0	-2908	12	5	5	2	2908
2	15	7	11	1	-1172	15	11	7	0	1172
3	43	4	5	1	-1321	15	5	4	0	1321
4	75	6	6	0	-1004	17	6	6	1	1004
5	18	5	3	1	698	36	3	5	0	-698
6	18	7	6	1	2411	57	6	7	0	-2411
7	16	5	13	0	-2615	15	13	5	1	2615
8	16	7	7	0	-1979	15	7	7	2	1979
9	13	4	5	1	-1548	16	5	4	0	1548

Figura 2: mostrando as 10 primeiras linhas do *data frame* traduzido.

4. APLICAÇÃO DA ANÁLISE DE VARIÂNCIA

Começamos avaliando os requisitos e pressupostos:

- As variáveis dependentes são quantitativas?
- A distribuição das variáveis dependentes segue uma distribuição normal dentro de cada grupo?
- A variância dos grupos são semelhantes?

As variáveis de interesse dessa análise são:

- *'azulAbates'* (quantidade de abates realizados pelo time azul);
- *'azulMortes'* (quantidade de mortes dos integrantes do time azul);
- *'vermelhoAbates'* (quantidade de abates realizados pelo time vermelho);
- *'vermelhoMortes'* (quantidade de mortes dos integrantes do time vermelho);

As variáveis explicativas, por sua vez, são as seguintes:

- *'azulSentinelas'* (sentinelas posicionadas pelo time Azul);
- *'azulMonstrosEpicos'* (monstros épicos derrotados pelo time Azul);
- *'azulDiferencaOuros'* (diferença de ouros do time Azul)
- *'vermelhoSentinelas'* (sentinelas posicionadas pelo time Vermelho);
- *'vermelhoMonstrosEpicos'* (monstros épicos derrotados pelo time Vermelho);
- *'vermelhoDiferencaOuros'* (diferença de ouros do time Vermelho).

Dados que as variáveis são quantitativas discretas, partimos para avaliação de sua normalidade, plotando o histograma dos grupos e seus gráficos Q-Q:

- **Histogramas das variáveis de interesse**

Primeiramente, plotamos histogramas das variáveis de interesse:

- Histograma do 'azulAbates'

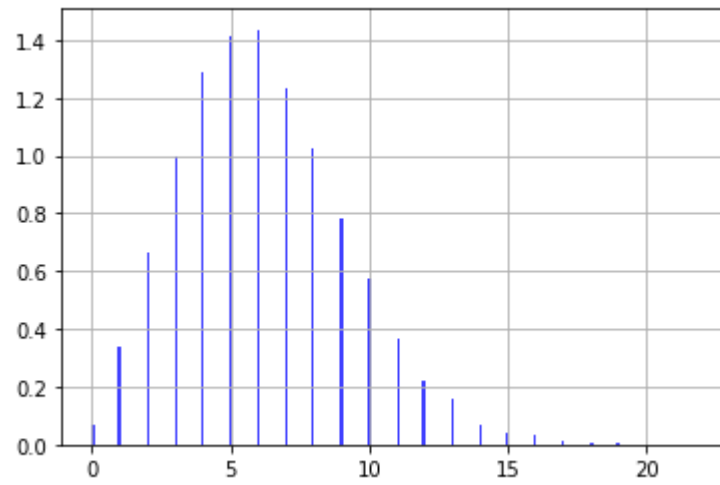


Figura 3: histograma da variável 'azulAbates'

- Histograma do 'azulMortes'

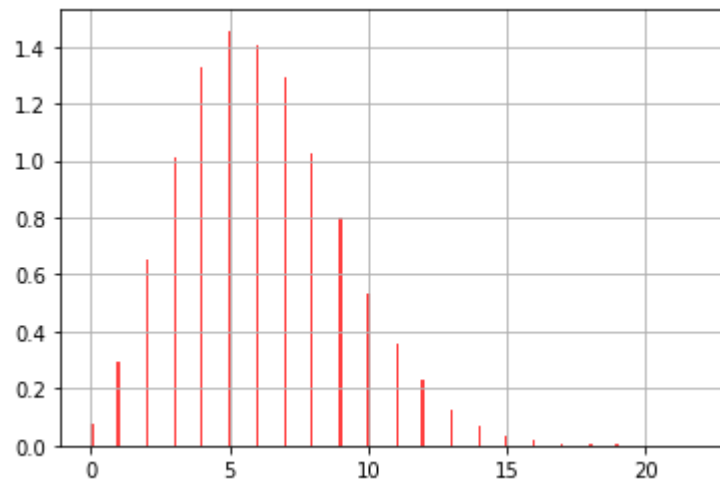


Figura 4: histograma da variável 'azulMortes'

- Histograma do *'vermelhoAbates'*

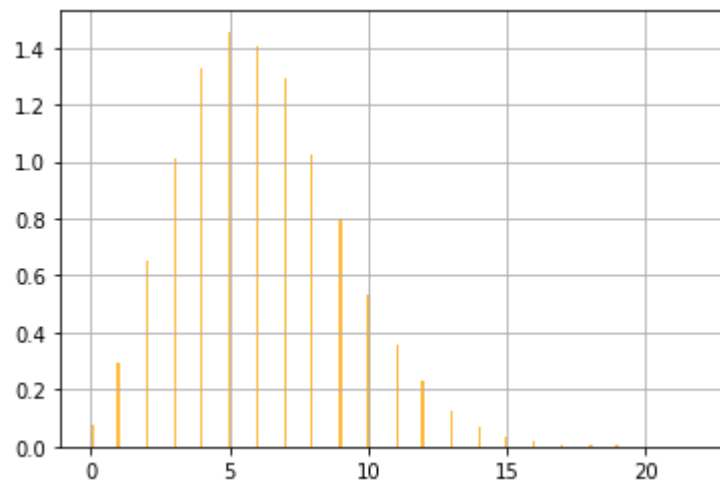


Figura 5: histograma da variável *'vermelhoAbates'*

- Histograma do *'vermelhoMortes'*

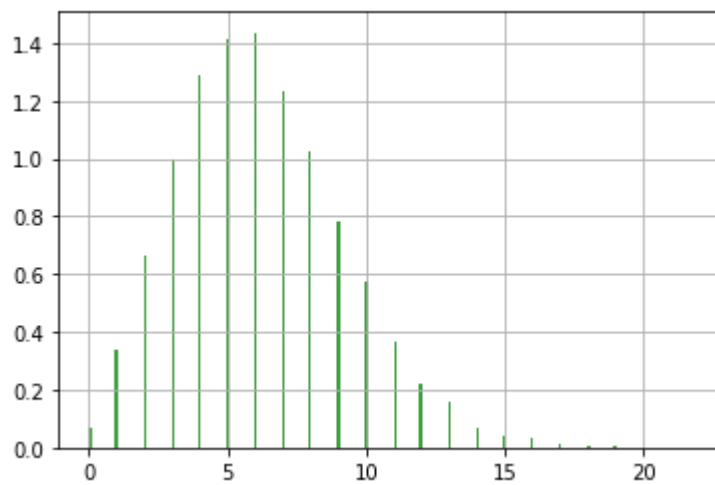


Figura 6: histograma da variável *'vermelhoMortes'*

- **Gráficos Q-Q das variáveis de interesse**

Na sequência, plotamos os gráficos Q-Q das variáveis de interesse:

- Gráfico Q-Q do 'azulAbates'

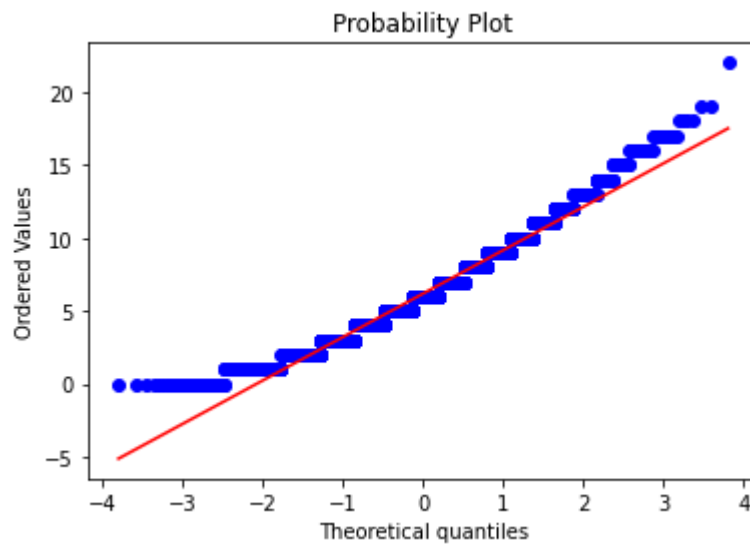


Figura 7: Gráfico Q-Q da variável 'azulAbates'

- Gráfico Q-Q do 'azulMortes'

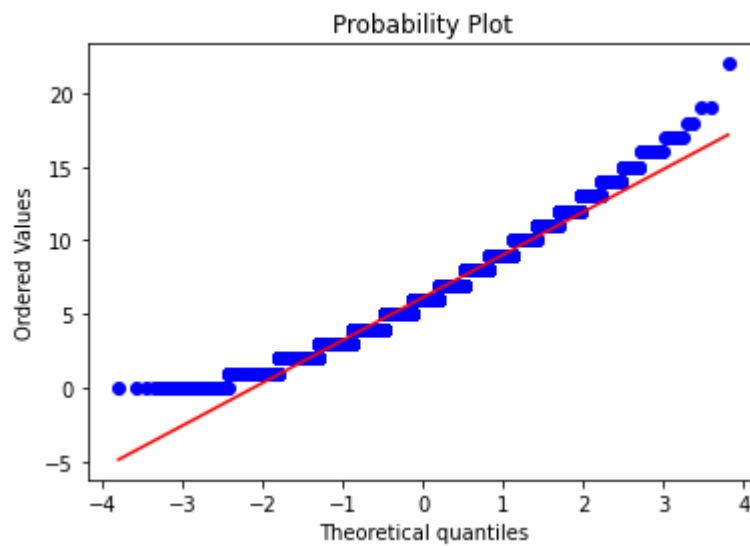


Figura 8: Gráfico Q-Q da variável 'azulMortes'

- Gráfico Q-Q do 'vermelhoAbates'

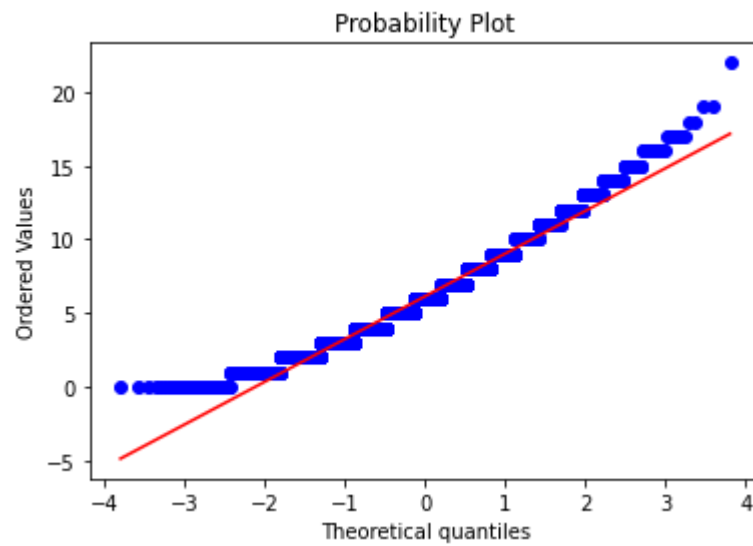


Figura 9: Gráfico Q-Q da variável 'vermelhoAbates'

- Gráfico Q-Q do 'vermelhoMortes'

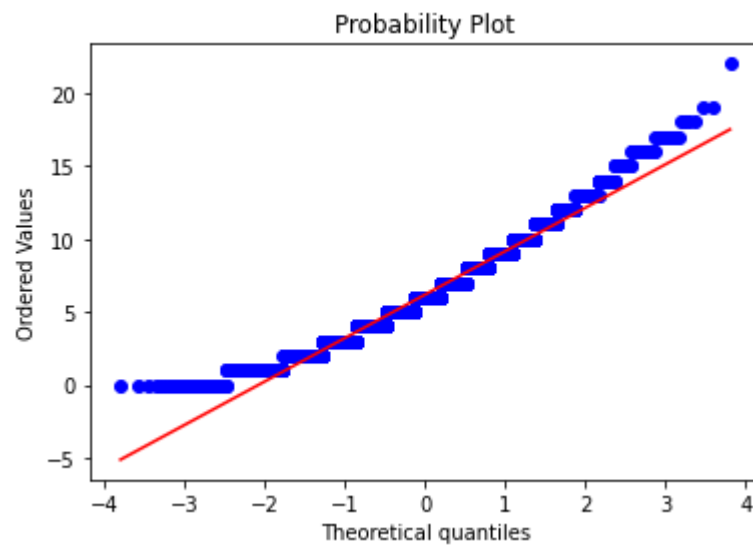


Figura 10: Gráfico Q-Q da variável 'vermelhoMortes'

Assim, vemos que ambas variáveis possuem distribuição normal.

Por fim, avaliando sua variância:

```
pd.set_option('display.float_format', '{:.5f}'.format)
df.var()

azulSentinelas      324.69072
azulAbates           9.06629
azulMortes           8.60729
azulMonstrosEpicos   0.39128
azulDiferencaOuros   6018922.19608
vermelhoSentinelas   340.67660
vermelhoAbates        8.60729
vermelhoMortes        9.06629
vermelhoMonstrosEpicos 0.39248
vermelhoDiferencaOuros 6018922.19608
dtype: float64
```

Figura 11: variância das variáveis envolvidas no análise

5. TESTE DE BARTLETT

Para testar a homocedasticidade da variância dos grupos, usaremos o teste de Bartlett (supondo a distribuição normal das variáveis).

Teste de hipótese (se p-value for maior que 0.05, assumimos que os dados são heterocedástico):

- Hipótese nula: as variâncias são iguais entre os grupos.
- Hipótese alternativa: as variâncias não são iguais entre os grupos.
- Teste entre *'azulAbates'* e *'vermelhoAbates'*

```
trt1 = df['azulAbates']
trt2 = df['vermelhoAbates']

stat, p = bartlett(trt1, trt2)

print('Teste de Bartlett\n T: {}, p-value:{}'.format(stat, p))

Teste de Bartlett
T: 6.664628758367618, p-value:0.009834513995856923
```

Figura 12: teste de hipótese entre *'azulAbates'* e *'vermelhoAbates'*

- Teste entre 'azulMortes' e 'vermelhoMortes'

```
from scipy.stats import bartlett

trt1 = df['azulMortes']
trt2 = df['vermelhoMortes']

stat, p = bartlett(trt1, trt2)

print('Teste de Bartlett\n T: {}, p-value:{}'.format(stat, p))

Teste de Bartlett
T: 6.664628758367618, p-value:0.009834513995856923
```

Figura 13: teste de hipótese entre 'azulMortes' e 'vermelhoMortes'

6. APLICANDO A ANOVA

Avaliamos 2 (dois) grupos: time azul e time vermelho, a partir das variáveis de interesse: 'azulAbates', 'azulMortes', 'vermelhoAbates' e 'vermelhoMortes'.

O interesse estava em verificar se há significância estatística no número de abates e mortes dos times azul e vermelho. Importante verificar também diferenças das variáveis métricas abates e mortes.

Dado que as amostras são relacionadas, o teste utilizado será ANOVA para medidas repetidas.

```
fvalue, pvalue = stats.f_oneway(df['azulMortes'], df['vermelhoMortes'])
print(fvalue, pvalue)

1.1961757324375164 0.2741007030743987
```

Figura 14: F-Value e P-Value

```
df_mv = df.filter(['azulMortes', 'vermelhoMortes', 'azulAbates', 'vermelhoAbates'])
df_mv = pd.melt(df_mv.reset_index(), id_vars=['index'], value_vars=['azulMortes',
                                                                    'vermelhoMortes',
                                                                    'azulAbates', 'vermelhoAbates'])
df_mv.columns = ['index', 'treatments', 'value']
model = ols('value ~ (treatments)', data=df_mv).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table

# https://www.reneshbedre.com/blog/anova.html
```

	sum_sq	df	F	PR(>F)
treatments	21.140703	3.0	0.79745	0.495068
Residual	349159.163883	39512.0	NaN	NaN

Figura 15: aplicação da ANOVA

```
from bioinfokit.analys import stat
res = stat()
res.anova_stat(df=df_mv, res_var='value', anova_model='value ~ (treatments)')
res.anova_summary

# https://www.reneshbedre.com/blog/anova.html
```

	df	sum_sq	mean_sq	F	PR(>F)
treatments	3.0	21.140703	7.046901	0.79745	0.495068
Residual	39512.0	349159.163883	8.836788	NaN	NaN

Figura 16: aplicação da ANOVA

7. INTERPRETAÇÃO DOS RESULTADOS

Nessa análise, o nosso objetivo principal é o de julgar se há alguma diferença no desempenho entre os dois possíveis lados (time azul ou time vermelho) de uma partida de League of Legends, informação que dependendo do resultado obtido, pode nos dar uma vantagem significativa no ambiente competitivo.

O resultado obtido após a aplicação da ANOVA foi de que nós temos um F de 0,79745 e um F crítico de 0,495068. Como nós temos um F maior do que o F crítico, podemos assumir que existe uma diferença significativa nas estatísticas entre os 2 lados analisados.

Considerando os resultados, podemos assumir que os lados azul e vermelho não são balanceados entre si, e que um time poder escolher em qual deseja jogar é uma vantagem competitiva relevante. As diferenças na ordem de escolha dos personagens e na distância dos pontos de interesse no mapa, de acordo com os cálculos, foram diferenças que afetaram significativamente o desempenho de ambos os lados, um positivamente e o outro negativamente.

8. CONCLUSÃO

Desse modo é possível confirmar com os resultados obtidos nessa análise que de fato existe um desbalanceamento, levando em conta as variáveis analisadas, a pessoa que jogar no equipe azul não teria o mesmo resultado que uma pessoa que joga na equipe vermelha. Isso impacta no resultado das estatísticas dos abates e mortes das duas equipes, consequentemente, pode alterar o resultado de uma partida do League of Legends dependendo de qual equipe o jogador é colocado.

9. REFERÊNCIAS

1. GOOGLE. **Google Colaboratory** - Disponível em: <https://colab.research.google.com>. Acesso em: 30 out. 2021.