

Estudo Estatístico

July 5, 2024

1 Hábitos de Jogo da Turma TPSIP 11.23

O presente estudo visa entender os hábitos de jogo dos alunos e professora da turma TPSIP 11.23. A análise é baseada em três variáveis, horas de jogo diárias, quantidade de jogos testados ao longo deste ano e os gêneros preferidos. Esta análise serve para entender melhor o comportamento da turma em relação ao tema.

```
[ ]: import pandas as pd
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
```

```
[ ]: # Funções Auxiliares

# Calcular Frequência
def calculate_frequencies(column):
    if column.dtype == 'object':
        absolute_freq = column.value_counts().sort_index()
    else:
        absolute_freq = column.value_counts().sort_index()

    relative_freq = absolute_freq / len(column)
    cumulative_abs_freq = absolute_freq.cumsum()
    cumulative_rel_freq = relative_freq.cumsum()

    return pd.DataFrame({
        'Frequência Absoluta ': absolute_freq,
        'Frequência Absoluta Acumulada': cumulative_abs_freq,

        'Frequência Relativa': relative_freq,
        'Frequência Relativa Acumulada': cumulative_rel_freq
    })

# Calcular média, moda, mediana, quartis, extremos e amplitude
def calculate_statistics(column):
    return pd.Series({
        'Média': column.mean(),
        'Moda': stats.mode(column, keepdims=False)[0],
```

```

        'Mediana': column.median(),
        'Q1 (25%)': column.quantile(0.25),
        'Q2 (50%)': column.quantile(0.50),
        'Q3 (75%)': column.quantile(0.75),
        'Mínimo': column.min(),
        'Máximo': column.max(),
        'Amplitude': column.max() - column.min()
    })

# Criar gráfico de barras e circular
def create_charts(data, column_name, ylabel= "Frequência"):
    value_counts = data[column_name].value_counts().sort_index()

    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(15, 6))

    # Gráfico de barras
    value_counts.plot(kind='bar', ax=ax1)
    ax1.set_title(f'Gráfico de Barras - {column_name}')
    ax1.set_xlabel(column_name)
    ax1.set_ylabel(ylabel)
    ax1.tick_params(axis='x', rotation=45)

    # Gráfico circular
    value_counts.plot(kind='pie', ax=ax2, autopct='%1.1f%%', startangle=90)
    ax2.set_title(f'Gráfico Circular - {column_name}')
    ax2.set_ylabel('')

    plt.tight_layout()
    plt.show()

# Criar diagrama de extremos e quartis
def create_horizontal_boxplot(data, column_name):
    plt.figure(figsize=(10, 6))
    plt.boxplot(data[column_name].dropna(), vert=False, whis=float('inf'))
    plt.title(f'Diagrama de Extremos e Quartis - {column_name}')

    plt.plot(data[column_name].min(), 1, 'bo', markersize=10) # Blue dot for
    ↪min
    plt.plot(data[column_name].max(), 1, 'bo', markersize=10) # Blue dot for
    ↪max

    plt.xlabel(column_name)
    plt.yticks([])
    plt.tight_layout()
    plt.show()

```

```
[ ]: # Ler dados
infoStudents = pd.read_csv('./data/csvFile.csv', header=0)
infoStudents['Média de Horas de Jogo Por Dia'] = pd.
    ↳to_numeric(infoStudents['Média de Horas de Jogo Por Dia'], errors='coerce')
infoStudents['Total de Jogos Testados Este Ano'] = pd.
    ↳to_numeric(infoStudents['Total de Jogos Testados Este Ano'], errors='coerce')
```

2 Informação recolhida

Conseguimos recolher as seguintes informações referentes à turma TPSIP 11.23:

```
[ ]: pd.DataFrame(infoStudents)
```

```
[ ]:
```

	Nome	Média de Horas de Jogo Por Dia	\
0	Afonso	12	
1	Artur	0	
2	Catarina	0	
3	Daniel	4	
4	Eliezer	0	
5	Gilberto	6	
6	Gonçalo	2	
7	Gustavo	1	
8	Hélder	0	
9	Henrique	0	
10	João	3	
11	Pedro F.	1	
12	Pedro G.	3	
13	Pedro D.	0	
14	Rafael R.	10	
15	Rafael C.	1	
16	Simão	2	
17	Vanessa	3	
18	Vasco	9	
19	Daniela	2	

	Total de Jogos Testados Este Ano	Género Favorito
0	52	FPS
1	2	FPS
2	1	LÓGICA
3	5	FPS
4	0	LÓGICA
5	23	FPS
6	12	FPS
7	3	TERROR
8	1	RPG
9	0	RPG

10	3	FPS
11	5	ARCADE
12	1	FPS
13	1	RPG
14	120	RPG
15	10	RPG
16	4	FPS
17	12	LÓGICA
18	3	TERROR
19	30	LÓGICA

```
[ ]: # Calcular Frequências
resultsFrequency = {}
for column in infoStudents.columns[1:]:
    resultsFrequency[column] = calculate_frequencies(infoStudents[column])

# Calcular Estatísticas
statistic_results = {}
for column in ["Média de Horas de Jogo Por Dia", "Total de Jogos Testados Este_
↪Ano"]:
    statistic_results[column] = calculate_statistics(infoStudents[column])
```

2.1 Média de Horas de Jogo Por Dia

```
[ ]: pd.DataFrame(resultsFrequency["Média de Horas de Jogo Por Dia"])
```

```
[ ]:
                                     Frequência Absoluta  \
Média de Horas de Jogo Por Dia
0                                     6
1                                     3
2                                     3
3                                     3
4                                     1
6                                     1
9                                     1
10                                    1
12                                    1

                                     Frequência Absoluta Acumulada  \
Média de Horas de Jogo Por Dia
0                                     6
1                                     9
2                                    12
3                                    15
4                                    16
6                                    17
```

9	18
10	19
12	20

	Frequência Relativa \
Média de Horas de Jogo Por Dia	
0	0.30
1	0.15
2	0.15
3	0.15
4	0.05
6	0.05
9	0.05
10	0.05
12	0.05

	Frequência Relativa Acumulada
Média de Horas de Jogo Por Dia	
0	0.30
1	0.45
2	0.60
3	0.75
4	0.80
6	0.85
9	0.90
10	0.95
12	1.00

Total Frequência Absoluta: 20

Total Frequência Relativa: 1

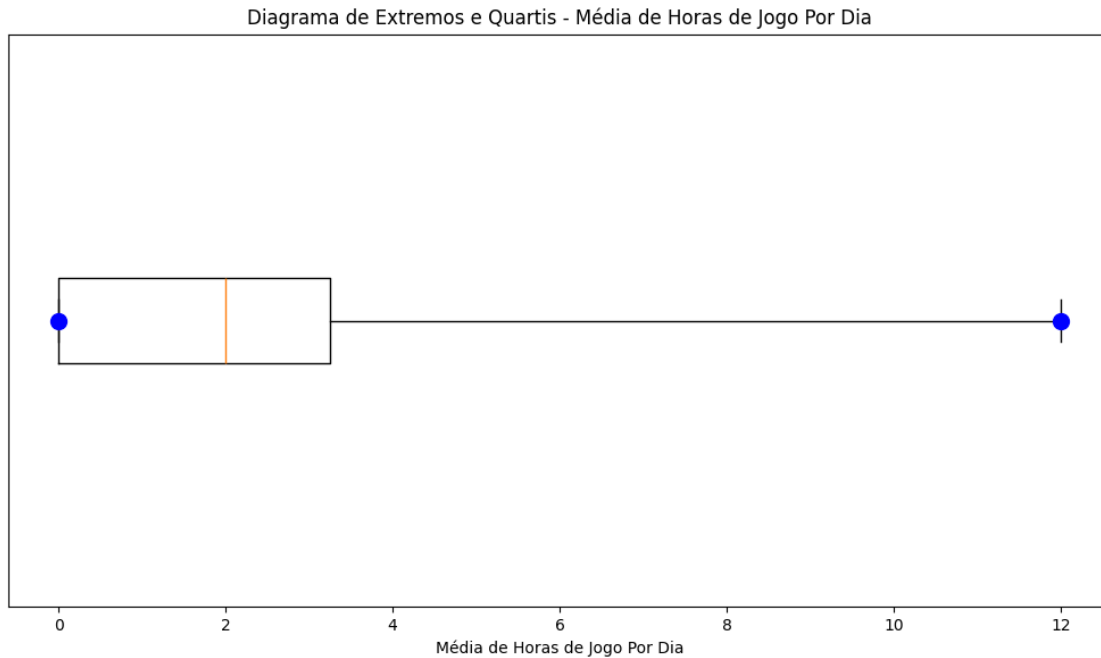
```
[ ]: pd.DataFrame(statistic_results["Média de Horas de Jogo Por Dia"])
```

```
[ ]:
Média      2.95
Moda       0.00
Mediana    2.00
Q1 (25%)   0.00
Q2 (50%)   2.00
Q3 (75%)   3.25
Mínimo     0.00
Máximo     12.00
Amplitude  12.00
```

Não foram utilizados intervalos visto que era um pequeno grupo de valores únicos. A média de 2,95 horas indica que os alunos passam cerca de 3 horas por dia a jogar. A mediana de 2 horas mostra que metade dos alunos joga até 2 horas por dia, o que sugere que a maioria dos alunos tem um

comportamento moderado em termos de tempo de jogo. A moda de 0 horas revela que uma parte significativa da turma não joga regularmente. A distribuição dos quartis demonstra que 75% dos alunos jogam até 3,5 horas por dia, e apenas uma minoria joga mais de 3,5 horas. A amplitude de 0 a 12 horas destaca uma variação significativa nos hábitos de jogo, com alguns alunos a dedicar uma quantidade substancial de tempo aos jogos, enquanto que outros não jogam.

```
[ ]: create_horizontal_boxplot(infoStudents, "Média de Horas de Jogo Por Dia")
```



```
[ ]: create_charts(infoStudents, "Média de Horas de Jogo Por Dia", ylabel="Nº de Alunos")
```

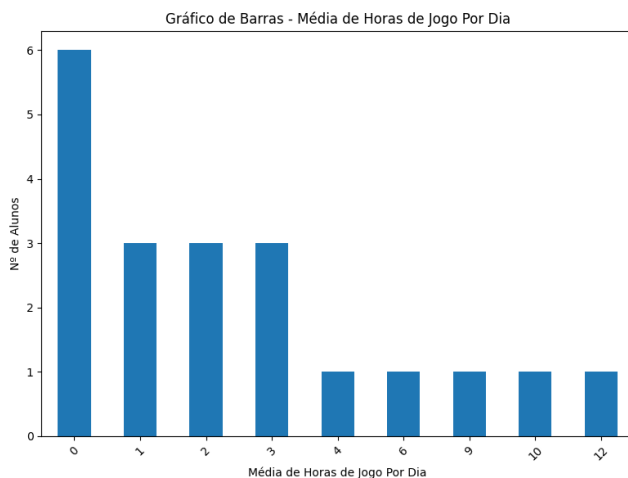
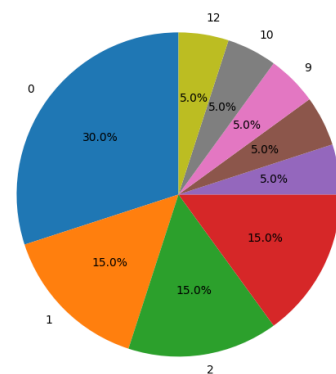


Gráfico Circular - Média de Horas de Jogo Por Dia



2.2 Quantidade de Jogos Testados

```
[ ]: pd.DataFrame(resultsFrequency["Total de Jogos Testados Este Ano"])
```

```
[ ]:
```

Total de Jogos Testados Este Ano	Frequência Absoluta \
0	2
1	4
2	1
3	3
4	1
5	2
10	1
12	2
23	1
30	1
52	1
120	1

Total de Jogos Testados Este Ano	Frequência Absoluta Acumulada \
0	2
1	6
2	7
3	10
4	11
5	13
10	14
12	16
23	17
30	18
52	19
120	20

Total de Jogos Testados Este Ano	Frequência Relativa \
0	0.10
1	0.20
2	0.05
3	0.15
4	0.05
5	0.10
10	0.05
12	0.10
23	0.05

30	0.05
52	0.05
120	0.05

	Frequência Relativa Acumulada
Total de Jogos Testados Este Ano	
0	0.10
1	0.30
2	0.35
3	0.50
4	0.55
5	0.65
10	0.70
12	0.80
23	0.85
30	0.90
52	0.95
120	1.00

Total Frequência Absoluta: 20

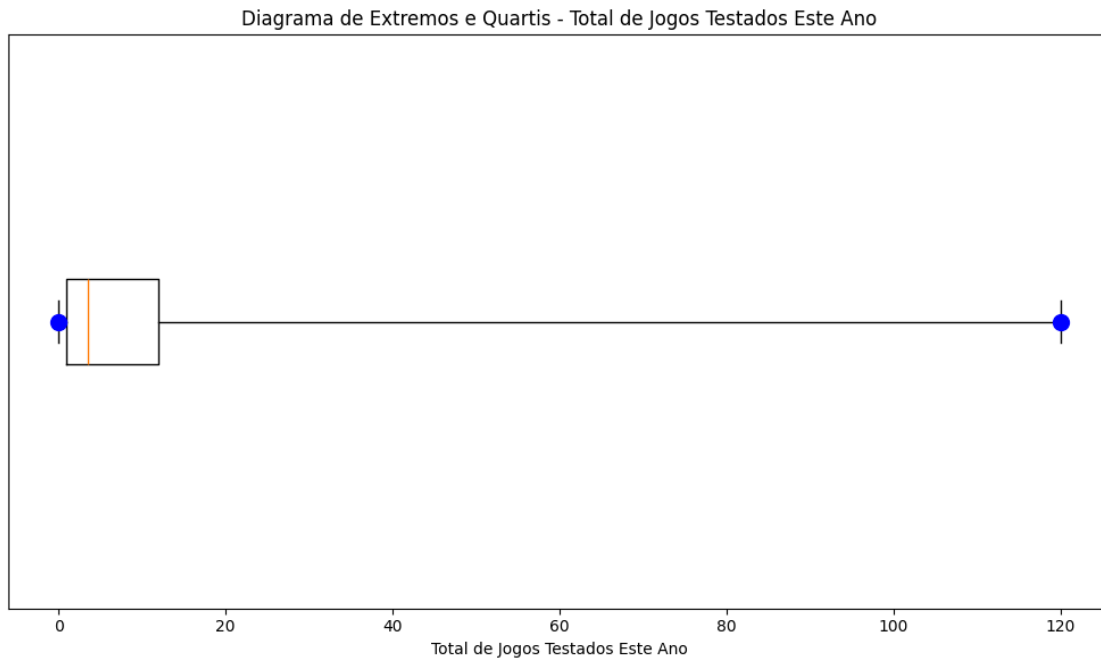
Total Frequência Relativa: 1

```
[ ]: pd.DataFrame(statistic_results["Total de Jogos Testados Este Ano"])
```

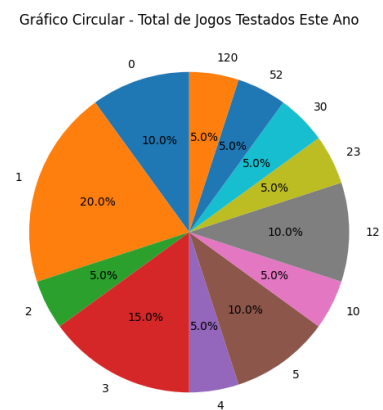
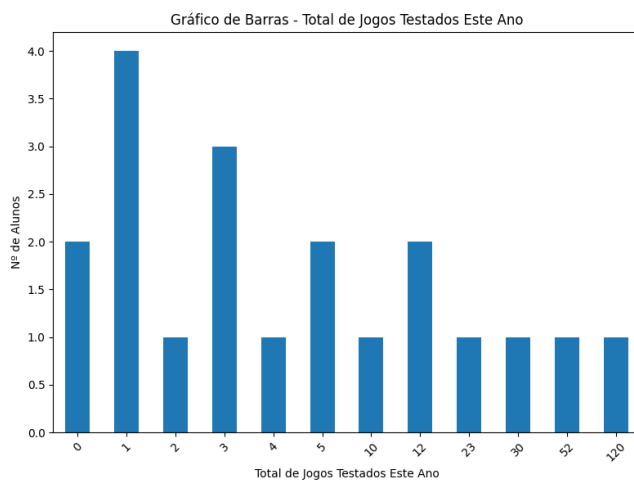
```
[ ]:
Média      14.4
Moda       1.0
Mediana    3.5
Q1 (25%)   1.0
Q2 (50%)   3.5
Q3 (75%)  12.0
Mínimo     0.0
Máximo    120.0
Amplitude  120.0
```

Visto que os valores são bastante concentrados, existia uma distância muito acentuada entre os valores e o conjunto de valores únicos é reduzido, decidimos utilizar números inteiros em vez de intervalos, caso contrário existiriam muitos intervalos vazios. A média de 14,4 jogos testados durante este ano indica que os alunos testaram mais de 2 jogos por mês. No entanto, a mediana de 3,5 jogos mostra que metade da turma testou menos de 4 jogos, o que sugere que um pequeno grupo de alunos jogou um grande número de jogos. A moda de 1 jogo revela que é comum que os alunos testem pelo menos um jogo por ano. A grande amplitude, de 0 a 120 jogos, indica que alguns alunos exploram muitos jogos, mas outros não testaram nenhum. Esta variação sugere diferentes níveis de interesse e envolvimento em jogos na turma TPSIP 11.23.

```
[ ]: create_horizontal_boxplot(infoStudents, "Total de Jogos Testados Este Ano")
```

```
[ ]: create_charts(infoStudents, "Total de Jogos Testados Este Ano", yLabel="Nº de Alunos")
```



2.3 Gênero Favorito

```
[ ]: pd.DataFrame(resultsFrequency["Gênero Favorito"])
```

```
[ ]:          Frequência Absoluta   Frequência Absoluta Acumulada \
Gênero Favorito
```

ARCADE	1	1
FPS	8	9
LÓGICA	4	13
RPG	5	18
TERROR	2	20

	Frequência Relativa	Frequência Relativa Acumulada
Gênero Favorito		
ARCADE	0.05	0.05
FPS	0.40	0.45
LÓGICA	0.20	0.65
RPG	0.25	0.90
TERROR	0.10	1.00

Total Frequência Absoluta: 20

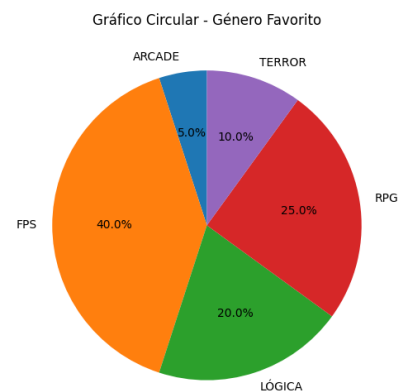
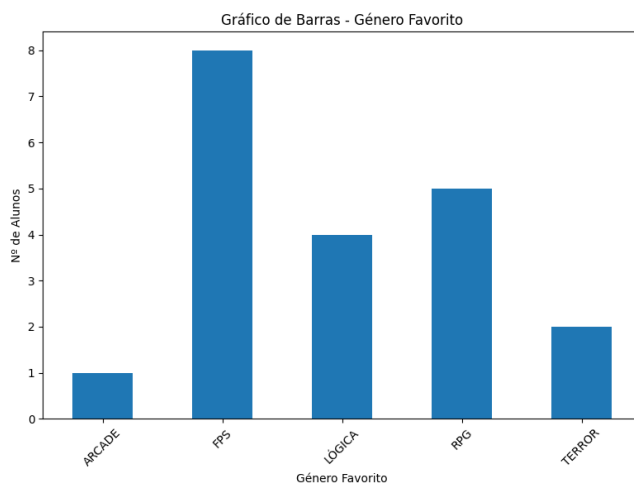
Total Frequência Relativa: 1

Moda: FPS

Os dados mostram que os jogos de FPS são os mais populares, preferidos por 40% dos alunos. Os jogos de RPG também têm uma presença significativa, com 25%. Jogos de lógica são escolhidos por 20% dos alunos, seguidos por jogos de terror (10%) e arcade (5%).

Esta distribuição indica uma preferência clara por jogos que oferecem ação e estratégia (FPS e RPG), jogos de lógica e terror também têm uma base considerável. Uma menor preferência por jogos de arcade pode sugerir que jogos mais modernos e complexos têm uma maior aderência.

```
[ ]: create_charts(infoStudents, "Gênero Favorito", ylabel="Nº de Alunos")
```



2.4 Análise

As frequências relativas das horas de jogo diárias mostram que 30% dos alunos não jogam diariamente, enquanto que 45% jogam entre 1 e 3 horas por dia. Apenas uma pequena fração (15%) joga mais de 9 horas por dia.

Quanto à quantidade de jogos testados, a maioria dos alunos testou entre 0 e 5 jogos ao longo do ano. Apenas uma pequena fração testou uma quantidade significativamente maior de jogos (até 120), o que indica uma diversidade de comportamento em relação a experimentar jogos.

O gênero de jogo preferido pelos alunos é o FPS (40%), seguido de RPG (25%) e lógica (20%).

2.5 Conclusão

A análise dos dados revela que a turma possui uma variada gama de comportamentos e preferências em relação a jogos. A maioria dos alunos joga de forma moderada, dedicando até 3,5 horas por dia aos jogos. O experimentar de jogos é comum, mas varia amplamente entre os alunos. Os jogos de FPS são os mais populares, seguidos por RPG e Lógica.

Estas informações podem ser úteis para entender os hábitos de jogo dos alunos, e permitem adaptar atividades e discussões sobre jogos de forma mais direcionada e eficaz.

2.6 Trabalho Elaborado Por:

- Artur Santos
- Pedro Dias
- Vanessa Ferreira