

# Análise de dados para verificar a relação entre as plataformas mais utilizadas e as empresas mais lucrativas

Filipe de Araujo Mendes, Iasmin Correa Araújo, Thaís Ferreira da Silva,

Departamento de Ciência da Computação - Universidade Federal de Minas Gerais (UFMG)

flipeara@ufmg.br, iasminaraujo@ufmg.br, thaisfds@ufmg.br

**Resumo.** *Este artigo trata sobre o mercado de jogos que vem crescendo mais a cada ano em busca de analisar se as plataformas onde os jogos são disponibilizados impactam nos lucros obtidos por uma empresa. Dessa forma, foi utilizado a técnica de bootstrap para calcular intervalos de confiança com 95% de confiabilidade, analisando o passado e presente do mercado de jogos, para assim afirmar que a plataforma escolhida afeta significativamente no lucro de uma empresa e que a recomendação para os próximos anos é o PS4.*

**Abstract.** *This article deals with the gaming market that has been growing more each year in an attempt to analyze whether the platforms where the games are made available impact the profits obtained by a company. In this way, the bootstrap technique was used to calculate confidence intervals with 95% reliability, analyzing the past and present of the gaming market, in order to affirm that the chosen platform significantly affects the profit of a company and that the recommendation for the years to come is the PS4.*

## 1. Introdução

Em 1972, surgiu o primeiro videogame da história, o Magnavox Odyssey. Com o passar do tempo, os videogames evoluíram cada vez mais, e se antes apenas uma pequena parcela da população era adepta aos games, na atualidade temos cada vez mais usuários dos mais diversos tipos de jogos.

Nesse contexto, investimentos em videogames vêm sendo uma grande oportunidade para obter retornos financeiros. Surgem questionamentos sobre quais são as empresas mais lucrativas e de onde vem esse lucro: plataforma, gênero do jogo, região de lançamento, dentre outros aspectos.

No presente artigo, analisamos a base de dados Video Game Sales, disponibilizada no Kaggle, para chegar a conclusões sobre quais são as empresas ideais para investir, com foco na melhor plataforma para se investir, olhando para o passado e para datas mais atuais

O restante do trabalho está organizado da seguinte forma: exploramos a base de dados, realizamos testes com relação às distribuições de plataformas, utilizando Bootstrap, e obtemos alguns resultados, discutindo-os posteriormente.

## **2. Descrição do problema e relevância**

É evidente que, dadas as diversas plataformas existentes, deve existir uma preferência por alguma em relação às outras. Essa preferência pode se mostrar tanto em dados globais quanto em dados regionais. Isso porque a cultura de uma sociedade pode impactar nos jogos que ela utiliza.

Dessa forma, esses se tornam aspectos essenciais para quem quer realizar investimentos em jogos. Além disso, também podemos pensar na diversificação de investimentos: investir em mais de uma empresa pode ser mais lucrativo também.

## **3. Base de dados**

A base de dados utilizada foi obtida no site Kaggle que contém dados de videogames com vendas superiores a 100.000 cópias entre 1980 e 2016. O banco de dados é composto por onze tabelas, dentre as quais foram utilizadas as seguintes:

- Rank - Classificação das vendas gerais
- Name - O nome do jogo
- Platform - Plataforma de lançamento dos jogos (ou seja, PC, PS4, etc.)
- Year - Ano de lançamento do jogo
- Genre - Gênero do jogo
- Publisher - Empresa do jogo

## **4. Descrição dos dados**

Para facilitar a análise ao longo da pesquisa, optamos por realizar um filtro nos dados levando em consideração as empresas mais relevantes, e os dados entre os anos 2000 e 2015.

A redução do intervalo dos anos analisados é devido ao histograma gerado de vendas por anos, onde percebemos uma baixa nos dados do banco de dados nos extremos. Os dados coletados antes de 2000 foram descartados devido a baixa competitividade no mercado que era dominado pela Nintendo, já os dados após 2015 eram poucos e acabaram por se tornar irrelevantes. Isso pode ser observado na Imagem 1 onde temos o histograma de Vendas por Ano antes da modificação e Imagem 2 onde temos o histograma de Vendas por Ano no intervalo de 2000 a 2015.

Além disso, as melhores 25 empresas equivalem a aproximadamente 91% das vendas ao longo dos anos. Dessa forma reduzimos os dados de 435 empresas no nosso banco de dados para 25 empresas com o objetivo de acelerar as análises, visto que a remoção das outras empresas não ocasionou em nenhuma alteração significativamente grande nas vendas a serem analisadas.

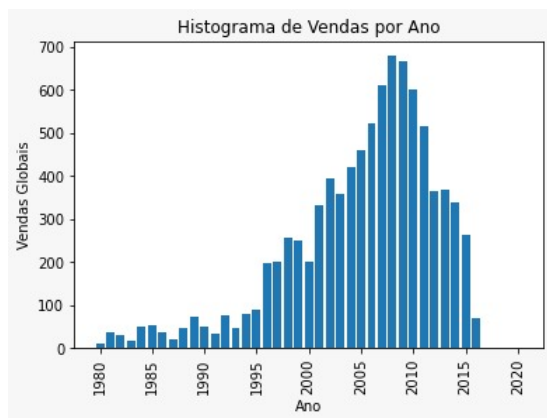


Imagem 1

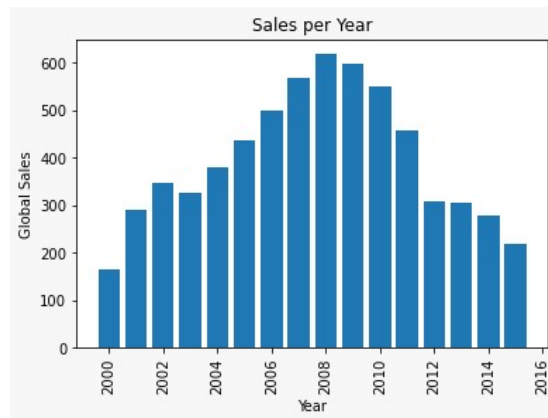


Imagem 2

Para as análises deste artigo, foram feitas 4 novas colunas com base nos dados presentes no site Kaggle, sendo eles: Main\_Plataform, Games\_Total, Games\_Main, Percentage\_Main

- Main\_Plataform - A principal plataforma utilizada pela empresa
- Games\_Total - Quantidade de jogos lançados no total
- Games\_Main - Quantidade de jogos lançados para a principal plataforma
- Percentage\_Main - Porcentagem dos jogos lançados na principal plataforma

	Publisher	Global_Sales	Main_Platform	Games_Total	Games_Main	Percentage_Main
15	Nintendo	1125.13	DS	478.0	149.0	31.171548
9	Electronic Arts	990.97	PS2	1201.0	196.0	16.319734
2	Activision	667.42	X360	875.0	131.0	14.971429
22	Ubisoft	445.91	DS	858.0	180.0	20.979021
17	Sony Computer Entertainment	418.80	PS2	494.0	204.0	41.295547

Tabela 1

## 5. Análise dos dados

### 5.1) Empresas mais lucrativas

Inicialmente buscamos compreender melhor com quais dados estávamos lidando, e o que poderia de certa forma ser relevante para o problema proposto. Ao analisarmos qual empresa estava no topo do ranking em cada ano percebemos que a Nintendo dominou o mercado entre 1983 e 2015, mas que com o passar dos anos algumas outras empresas apareceram no primeiro lugar como a Sony Computer Entertainment, que ficou no pódio duas vezes, e a Electronic Arts, que ficou no pódio 7 vezes, e atualmente ambas são grandes empresas no ramo dos jogos. Nessa primeira parte levamos em consideração a porcentagem das vendas anuais obtidas por cada empresa.

	Publisher	Year	Global_Sales	Percentage
	Nintendo	1983.0	10.96	100.000000
	Nintendo	1984.0	45.56	90.956279
	Nintendo	1985.0	49.95	93.381941
	Nintendo	1986.0	16.18	44.438341
	Nintendo	1987.0	11.95	60.475709
	Nintendo	1988.0	36.44	78.416182
	Nintendo	1989.0	63.88	87.711108
	Nintendo	1990.0	35.49	75.849541
	Nintendo	1991.0	15.97	57.260667
	Nintendo	1992.0	38.11	60.053577
	Nintendo	1993.0	20.04	45.942228
	Nintendo	1994.0	24.99	37.732146
	Sony Computer Entertainment	1995.0	18.45	24.289100
	Nintendo	1996.0	73.70	38.519835
	Sony Computer Entertainment	1997.0	43.90	22.604397
	Nintendo	1998.0	48.41	19.421488

	Nintendo	1999.0	65.33	26.605579
	Nintendo	2000.0	34.05	17.471394
	Nintendo	2001.0	45.37	13.743487
	Electronic Arts	2002.0	73.01	18.472788
	Electronic Arts	2003.0	69.83	19.513763
	Electronic Arts	2004.0	67.33	16.057332
	Nintendo	2005.0	127.47	27.714484
	Nintendo	2006.0	205.61	39.461462
	Nintendo	2007.0	104.18	17.047667
	Nintendo	2008.0	91.22	13.437233
	Nintendo	2009.0	128.89	19.315151
	Electronic Arts	2010.0	81.38	13.553168
	Electronic Arts	2011.0	72.25	14.143372
	Activision	2012.0	50.34	15.386496
	Take-Two Interactive	2013.0	52.41	15.711374
	Electronic Arts	2014.0	45.96	15.330220
	Electronic Arts	2015.0	45.42	18.849602

Tabela 2

## 5.2) Principal plataforma de cada empresa

Em seguida tentamos encontrar uma relação entre as empresas e as principais plataformas onde os seus jogos foram disponibilizados para o público. Dessa forma, criamos 4 novas colunas, semelhantes às da tabela 1, e geramos uma matriz de correlação presente na Imagem 3 para verificar como as vendas e plataformas se relacionam. Podemos observar que Games\_Main e Games\_Total tem uma correlação igual a 0.87, o que nos dá um indício de que os aspectos se relacionam.

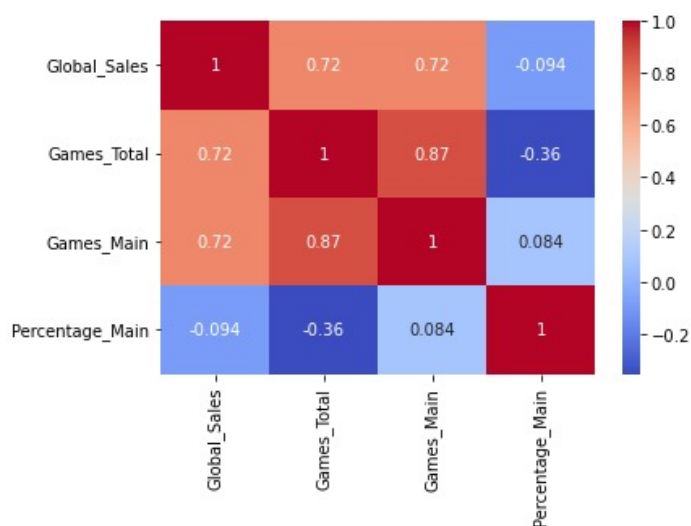
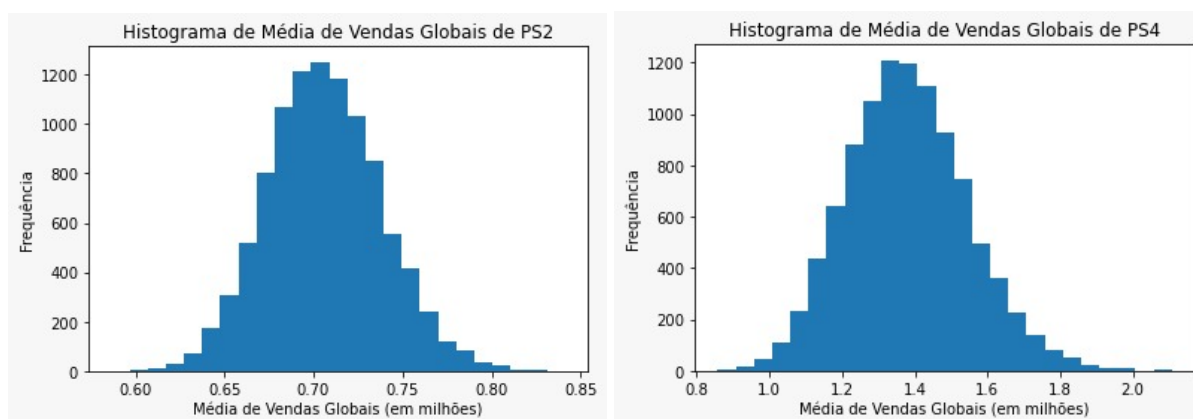


Imagem 3

Ainda sobre as vendas por plataforma, podemos fazer o bootstrap para cada amostra.



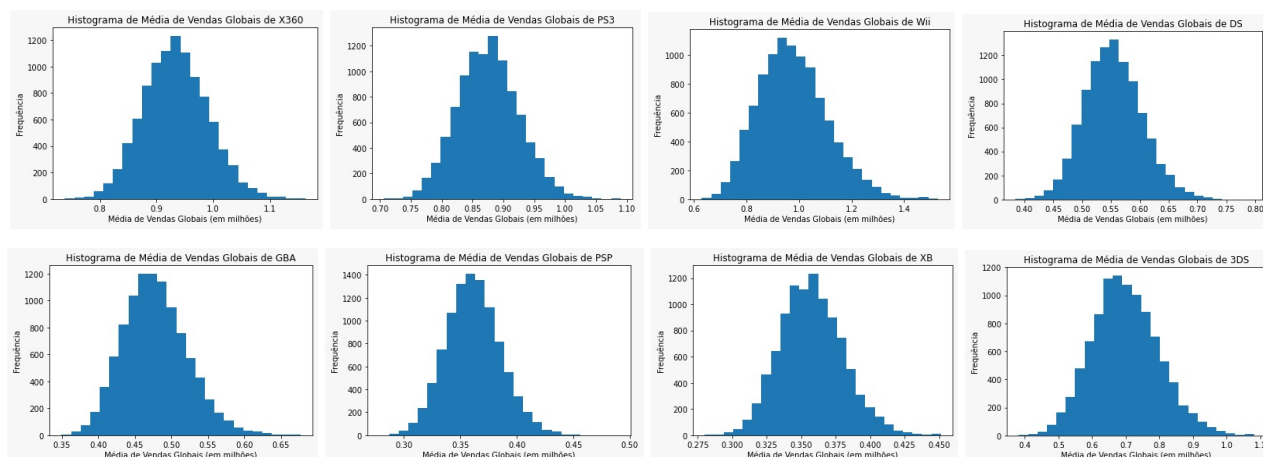


Imagem 4

Dado o bootstrap, conseguimos algumas estatísticas, apresentadas na tabela abaixo. Porém, apenas obter dados das plataformas ainda não auxilia tanto, pois podemos ter plataformas que venderam mais em anos específicos e não vendem mais atualmente. Além disso, o jogo que lançou em 2008, por exemplo, continua vendendo nos anos seguintes, então comparar um jogo lançado em 2008, que até 2016 teve 8 anos para vender, com um jogo lançado em 2015, o qual teve somente 1 ano para vender, não nos dá uma ideia muito boa sobre as vendas. Por isso, juntamente às estatísticas, geramos a média e o desvio padrão dos anos para comparar plataformas que venderam em períodos similares, para saber, naquele período, qual seria a melhor plataforma para se lançar um jogo.

	Platform	Global_Sales	Mean	Std	CI_Lower	CI_Upper	Year_Mean	Year_Std
8	PS2	1105.90	0.706197	0.031999	0.646470	0.770776	2004.432674	2.477310
13	X360	926.22	0.933635	0.056426	0.827820	1.049803	2009.703629	2.437693
9	PS3	894.11	0.876469	0.048970	0.785535	0.977383	2010.490196	2.342780
12	Wii	834.41	0.976208	0.133563	0.746817	1.264069	2008.909730	1.631573
1	DS	727.12	0.554877	0.050173	0.462873	0.659553	2008.019069	1.755368
3	GBA	268.33	0.477383	0.044512	0.398232	0.573049	2003.203209	1.488548
11	PSP	259.88	0.362323	0.024544	0.317200	0.413050	2007.986072	2.344720
14	XB	233.36	0.357636	0.022947	0.315660	0.405675	2003.648773	1.381560
10	PS4	220.74	1.381130	0.164455	1.078619	1.719103	2014.462500	0.652884
0	3DS	214.85	0.696830	0.099684	0.515356	0.904289	2012.759740	1.439576

Tabela 3

Dessa forma, podemos agrupar plataformas, levando em consideração a média e desvio padrão do ano de lançamento dos jogos, presentes na Tabela 3, para poder compará-las entre si. Pegando plataformas com períodos de tempos similares resulta em jogos que tiveram, aproximadamente, o mesmo tempo no mercado para vender, fazendo sentido, assim, comparar as vendas entre eles.

GRUPO 1	GRUPO 2	GRUPO 3
PS2 (2004,4 ; 2,5)	DS (2008,02; 1,75)	X360 (2009,7; 2,43)
GBA (2003,2 ; 1,5)	Wii (2008,9; 1,63)	PS3 (2010, 5; 2,34)
XB (2003,6 ; 1,4)	-	-

Tabela 4

Assim, podemos obter a melhor plataforma de cada grupo, demonstrados na Tabela 4. Para o grupo 1, de 2003, a plataforma a qual o intervalo de confiança ficou mais acima foi o PS2, mas ainda assim, talvez não seja tão justo colocá-la com o GBA e o XB, já que a média do ano é 1 a mais e o desvio padrão do PS2 é quase o dobro das outras duas plataformas, então poderíamos dividir o grupo 1 em grupo 1a (PS2) e grupo 1b (GBA e XB), para então compará-las entre si.

### 5.3) Plataformas recentes mais lucrativas

Considerando que temos as plataformas mais populares de cada período, para realizar inferências precisamos de observar dados que estejam atualizados. Por exemplo: PS2 foi a plataforma que mais vendeu de forma geral, mas ela foi lançada em 2000, não sendo mais tão utilizada hoje em dia. Assim, optou-se por analisar um recorte dos dados mais atuais, a partir de 2013.

Observamos que as 15 empresas que mais venderam representam aproximadamente 91% das vendas, e as 5 plataformas com mais vendas representaram 94% delas. Rodamos novamente o bootstrap para considerar as estatísticas a partir de 2013.

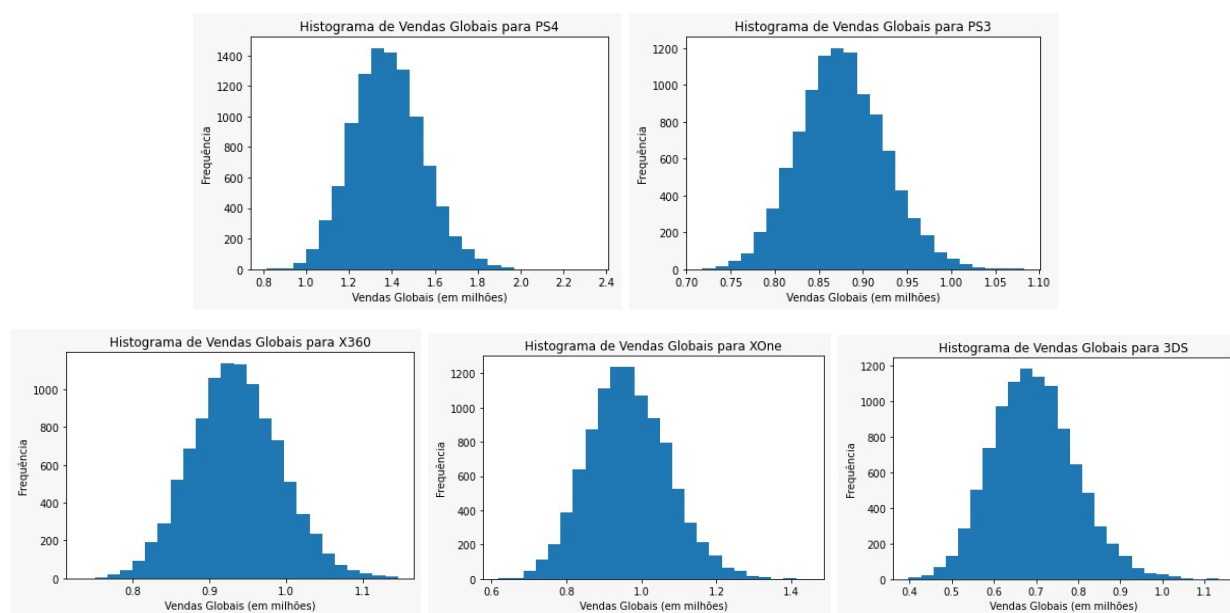


Imagem 5

Com isso, também temos novas estatísticas a observar:

	Platform	Global_Sales	Mean	Std	CI_Lower	CI_Upper	Year_Mean	Year_Std
4	PS4	222.78	1.379934	0.163421	1.073438	1.723128	2014.462500	0.652884
3	PS3	167.15	0.876264	0.048889	0.784831	0.976248	2010.490196	2.342780
7	X360	127.60	0.934625	0.057667	0.825069	1.050716	2009.703629	2.437693
8	XOne	123.76	0.964869	0.106258	0.766454	1.186313	2014.307692	0.702666
0	3DS	108.09	0.695576	0.098347	0.517140	0.901240	2012.759740	1.439576

Tabela 5

Dessa forma, podemos novamente agrupar os dados nos anos similares e atuais. Assim, deixamos PS4(2014,5; 0,65) e XOne(2014,3; 0,7).

## 6. Resultados

### 6.1) Melhores plataformas historicamente

Observando as Tabelas 3 e 4, conseguimos separar as plataformas em “Gerações”, para então compará-las. Como explicado anteriormente, iremos separar o grupo 1 em grupo 1a e grupo 1b, classificando o grupo 1b como 1ª geração, o grupo 2 como 2ª e o grupo 3 como 3ª. Não faz muito sentido comparar as plataformas que não se encaixam dentro dessas 3 gerações, pelo tempo que o jogo ficou no mercado, como explicado mais cedo, então ficaremos restritos a esses grupos.

Na primeira geração, composta pelo GBA e pelo XB, é possível perceber que o intervalo de confiança do GBA é mais acima. Simplificadamente, isso nos diz que, caso dois jogos, um de GBA e outro de XB, fossem comparados, muito provavelmente o jogo que mais vendeu seria o do GBA. Podemos também fazer a inferência de que, caso pudéssemos escolher lançar um jogo para o GBA ou o XB, se escolhêssemos o GBA, o jogo venderia mais. Portanto, o GBA foi a plataforma mais bem sucedida entre as duas.

A mesma análise pode ser feita para as gerações 2 e 3. O Wii teve um intervalo de confiança muito acima do DS, portanto têm maiores médias de vendas e foi mais bem sucedido. Na terceira geração, os intervalos possuem uma interseção muito maior, mas os números do X360 ainda são levemente maiores, indicando que, apesar de não ser tão significativa a diferença, ele ainda foi mais bem sucedido que o PS3 em média de vendas.

### 6.2) Melhor plataforma futura

Analogamente, podemos fazer a comparação entre o que podemos chamar de 4ª geração, o PS4 e o XOne. Como pode ser observado na Tabela 5, o intervalo de confiança das duas plataformas é maior do que os intervalos anteriores, o que indica uma maior variação nas vendas para essas plataformas, e também existe uma interseção, entretanto, tal interseção é pequena, ficando bem claro que o PS4 possui uma média de vendas bem maior que o XOne, sendo então a melhor plataforma.

Assim, queremos, sabendo que o PS4 é a plataforma mais bem sucedida atualmente, olhar para as empresas que mais publicam jogos para ela. Para isso, geramos outras estatísticas, considerando tanto números relativos quanto absolutos de vendas de jogos para PS4 para cada uma das Top 15 empresas atuais.

Publisher	Global_Sales	Total_Games	PS4_Games	Percentage_PS4
Electronic Arts	143.52	107	20	18.691589
Activision	120.50	104	17	16.346154
Take-Two Interactive	96.85	46	8	17.391304
Ubisoft	78.42	92	19	20.652174
Warner Bros. Interactive Entertainment	53.34	62	12	19.354839
Sony Computer Entertainment	40.08	23	13	56.521739
Square Enix	31.33	52	12	23.076923
Namco Bandai Games	29.50	156	16	10.256410
Bethesda Softworks	25.12	22	6	27.272727
Konami Digital Entertainment	16.78	40	4	10.000000
Sony Computer Entertainment Europe	14.48	8	4	50.000000
Capcom	13.20	38	6	15.789474
Disney Interactive Studios	11.65	21	2	9.523810

Tabela 6

Com isso, investir em empresas que lançam mais jogos para o PS4 seria uma estratégia melhor do que investir no seu concorrente direto, o XOne, já que seriam lançados jogos para o PS4 que com alta probabilidade, traria mais lucro que os jogos feitos para o XOne.

### 6.3) Crescimento do mercado

Com exceção do PS2, podemos observar uma movimentação de um deslocamento nos intervalos de confiança de uma geração para outra, tendendo a mais vendas para gerações mais novas. Isso pode ser explicado pelo fato de que, com o passar do tempo, o mercado dos jogos se expandiu, e houve um aumento nas vendas em geral. A única exceção a esse ponto é o PS2, que é antigo e teve um número de vendas significativo em comparação com sua geração. Isso pode ser explicado porque existe uma infinidade de jogos para PS2, que foi um console acessível, marcando toda uma geração com seus icônicos jogos para cada filme e desenho animado lançado na época.

## 7. Conclusão e trabalho futuros

Neste trabalho foram analisados aspectos da base de dados Video Game Sales, considerando principalmente qual o impacto de determinadas plataformas para as vendas de videogames. Fazendo tal análise, é possível observar qual empresa mais publica jogos para determinada plataforma e encontrar as empresas mais lucrativas com base em tal aspecto. Isso auxiliaria no objetivo inicial deste trabalho: recomendar empresas a se investir com base na análise dos dados



das vendas de videogames.

Porém, existem outros aspectos ainda não investigados que podem ser observados para melhores recomendações, como os impactos das regiões e dos gêneros de jogos produzidos. Esses podem ser posteriormente melhor explorados ao longo de uma nova investigação com novos métodos da ciência de dados. Assim, teremos recomendações de investimentos cada vez mais sofisticadas.

## **8. Referências**

- 1) Base de dados: <https://www.kaggle.com/datasets/gregorut/videogamesales>
- 2) Bootstrap: <https://www.educative.io/answers/what-is-the-bootstrap-method-in-data-science>
- 3) História do Videogame:  
<https://www.microlins.com.br/blog/tecnologia/a-evolucao-dos-videogames/>