

## 2ª Etapa

Eniedson Fabiano Pereira da Silva Junior - 119110512

João Paulo Alves Dantas - 119210258

1)

A indústria do entretenimento vem ganhando cada vez mais espaço com o passar dos anos, dentro dela, o mercado de jogos digitais se destaca com receita de quase 200 bilhões apenas esse ano e com uma estimativa de aumento de 2,1% ao ano, segundo o Global Games Market Reports de 2022. Um dos fatores determinantes para o sucesso ou não de novos jogos é o gênero, a partir disso, estudos e análises quanto às vendas a partir do mesmo é de extrema relevância, já que pode garantir investimentos mais promissores e também melhorar a qualidade de jogos de gêneros mais requisitados pelos clientes no mundo todo.

2)

- a) Jogos existentes no mercado relacionados a quantidade de cópias vendidas dos mesmos, em milhões. A variável analisada é contínua.
- b)
  - i) Objetivo Geral - Analisar as vendas dos jogos a partir dos seus respectivos gêneros.
  - ii) Objetivo Específico - Verificar quais gêneros são mais promissores a partir da quantidade de cópias vendidas.
- c) O gênero do jogo, que é uma variável do tipo categórica, e a quantidade de vendas do jogo, que é uma variável quantitativa.
- d) Existem muitos jogos espalhados pelo mundo e novos títulos são criados e lançados diariamente, dentre esses existem muitos jogos famosos, mas também existem jogos muito pouco vendidos e que não teriam impactos positivos nas análises realizadas, tendo em vista isso, a base de dados conta com uma amostragem que se baseia em selecionar apenas jogos que possuem mais que 100 mil cópias vendidas.
- e)
  - i) Para uma boa análise dos dados vários parâmetros podem ser utilizados, mas dentre eles, as médias e medianas podem ser de grande utilidade para identificar as medidas de centralidade dos dados em questão para cada grupo e analisar quais grupos possuem, no geral, são mais promissores, além disso, outro parâmetro interessante poderia ser o somatório de todas as cópias vendidas dentro de cada grupo, podendo verificar as maiores concentrações de vendas para os dados analisados;
  - ii) Nesse caso, para identificar os melhores gêneros, o desvio padrão é uma ótima estatística, já que pode diferenciar gêneros promissores na totalidade, que possuem um desvio padrão mais baixo, de grupos não tão promissores, mas que possuem parâmetros altos devido a outliers presentes nos mesmos, o que pode ser feito a partir de uma análise, por exemplo, usando o coeficiente de variação;
  - iii) A partir de todas as medidas e parâmetros, algumas das estimativas e inferências possíveis poderia se direcionar a avaliação das taxas de sucesso de um determinado jogo a ser desenvolvido ou então para a escolha de um novo gênero para direcionar investimentos na criação de um novo título.

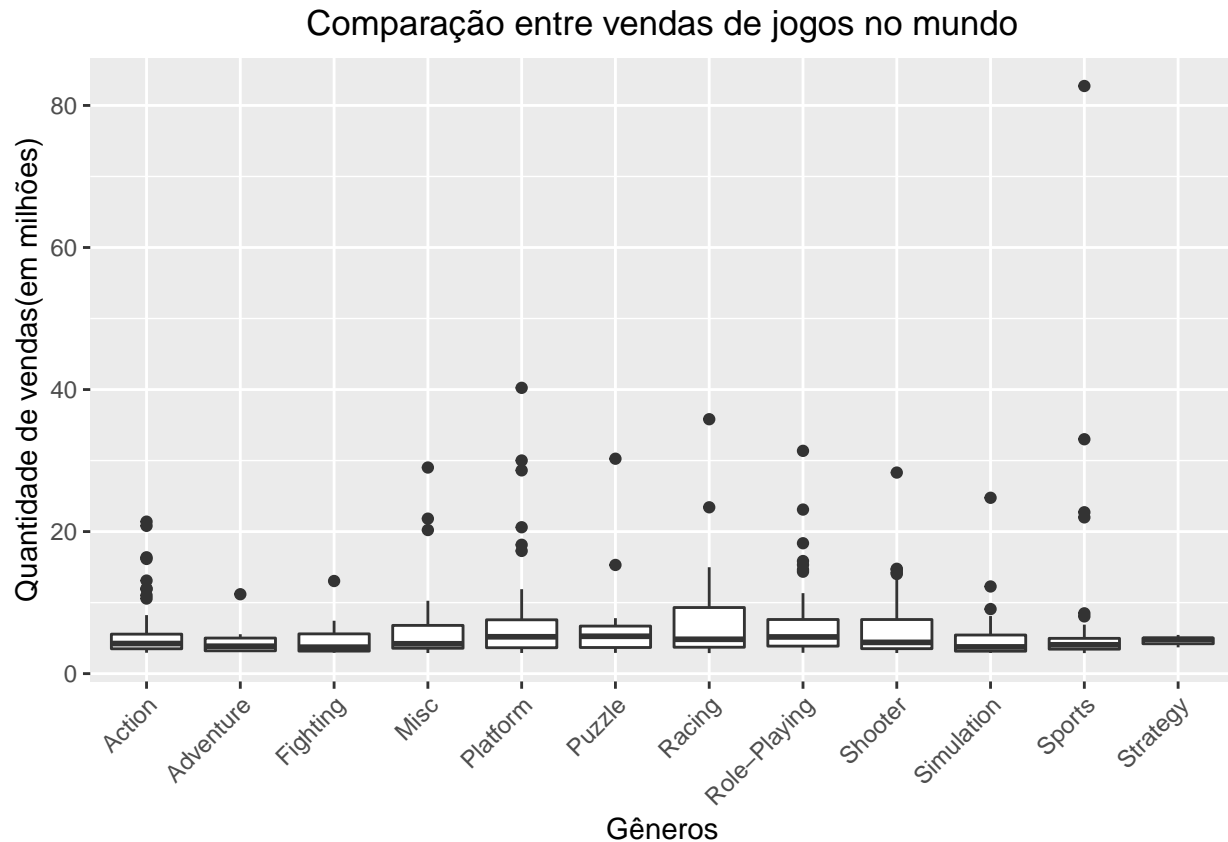
3)

- a) Resposta para a questão foi discutida nos itens 1 e 2 desta mesma atividade.
- b) Video Game Sales

4)

## Global

Na primeira parte iremos analisar os números no cenário mundial, o primeiro passo é exibir esses dados de maneira a analisar as médias, medianas e o IQR. Para isso escolhemos uma visualização baseada em um boxplot, que podemos ver a seguir:



Analisando o resultado exibido no gráfico podemos chegar a suposição que de fato o gênero não influencia nas vendas totais de um jogo, já que todas as médias apresentam resultados bem-parecidos, e somente alguns casos específicos fogem dessa curva, levando a acreditar que outros fatores influenciem mais que o gênero.

Para dar continuidade a nossos testes é importante testar a normalidade dos dados, e para isso utilizaremos o teste de Shapiro-Wilk nos grupos.

Genre	W	P.value
Sports	0.3127973	0.000000000000004011810
Platform	0.6247107	0.000000000081621580876
Racing	0.6565865	0.000000113140017376444
Role-Playing	0.6849663	0.000000000714109395788
Puzzle	0.6397710	0.000080025473704374074
Misc	0.5892975	0.000000000933466579956
Shooter	0.7477099	0.000000000416368016223
Simulation	0.5926912	0.000003749323319654595
Action	0.6253278	0.000000000000004746311
Fighting	0.7249293	0.000021905807885605515
Adventure	0.6919917	0.001827851138423381826

Strategy	0.9692752	0.870581673605300365715
----------	-----------	-------------------------

Como o valor de  $p$  foi menor que 0,05 em pelo menos um dos grupos, e seguindo a orientação do Canal Pesquisa, escolhemos o teste de Kruskal-Wallis para determinar de fato se essas variáveis são independentes.

Ao aplicar esse teste obtemos um valor de  $p$  igual a 0.0545696, ou seja, maior que o valor definido de 0,05, e por esse motivo podemos bater o martelo e definir que no cenário global gênero não é um fator determinante para as vendas de um jogo.

Porem o valor ainda é bem próximo ao nosso limiar, então para fazer uma análise mais precisa podemos ainda utilizar o teste de Pairwise Wilcox, que vai comparar as vendas entre cada um dos gêneros. Aplicando esse teste nos nossos dados obtemos o seguinte resultado:

Genre1	Genre2	value	Relevancia
Adventure	Action	0.7280435	Não
Fighting	Action	0.5509424	Não
Fighting	Adventure	0.9468510	Não
Misc	Action	0.8391801	Não
Misc	Adventure	0.7280435	Não
Misc	Fighting	0.5026927	Não
Platform	Action	0.2688766	Não
Platform	Adventure	0.6417719	Não
Platform	Fighting	0.2327402	Não
Platform	Misc	0.6553777	Não
Puzzle	Action	0.7280435	Não
Puzzle	Adventure	0.7021248	Não
Puzzle	Fighting	0.6700006	Não
Puzzle	Misc	0.8303474	Não
Puzzle	Platform	0.9468510	Não
Racing	Action	0.3685761	Não
Racing	Adventure	0.6419665	Não
Racing	Fighting	0.2327402	Não
Racing	Misc	0.6419665	Não
Racing	Platform	0.9468510	Não
Racing	Puzzle	0.9468510	Não
Role-Playing	Action	0.2688766	Não
Role-Playing	Adventure	0.6417719	Não
Role-Playing	Fighting	0.2327402	Não
Role-Playing	Misc	0.6691571	Não
Role-Playing	Platform	0.9933051	Não
Role-Playing	Puzzle	0.9468510	Não
Role-Playing	Racing	0.9468510	Não
Shooter	Action	0.6417719	Não
Shooter	Adventure	0.6691571	Não
Shooter	Fighting	0.2652894	Não
Shooter	Misc	0.8391801	Não
Shooter	Platform	0.7771064	Não
Shooter	Puzzle	0.9729863	Não
Shooter	Racing	0.7280435	Não
Shooter	Role-Playing	0.7874555	Não
Simulation	Action	0.7771064	Não
Simulation	Adventure	0.9933051	Não
Simulation	Fighting	0.9059909	Não

Simulation	Misc	0.6979054	Não
Simulation	Platform	0.5026927	Não
Simulation	Puzzle	0.7280435	Não
Simulation	Racing	0.5442272	Não
Simulation	Role-Playing	0.4463233	Não
Simulation	Shooter	0.6691571	Não
Sports	Action	0.7021248	Não
Sports	Adventure	0.9661898	Não
Sports	Fighting	0.7771064	Não
Sports	Misc	0.6691571	Não
Sports	Platform	0.2327402	Não
Sports	Puzzle	0.6419665	Não
Sports	Racing	0.2327402	Não
Sports	Role-Playing	0.2327402	Não
Sports	Shooter	0.4753215	Não
Sports	Simulation	0.9468510	Não
Strategy	Action	0.8370172	Não
Strategy	Adventure	0.7771064	Não
Strategy	Fighting	0.7021248	Não
Strategy	Misc	0.8391801	Não
Strategy	Platform	0.7771064	Não
Strategy	Puzzle	0.8391801	Não
Strategy	Racing	0.8590205	Não
Strategy	Role-Playing	0.7821499	Não
Strategy	Shooter	0.9933051	Não
Strategy	Simulation	0.7771064	Não
Strategy	Sports	0.6700006	Não

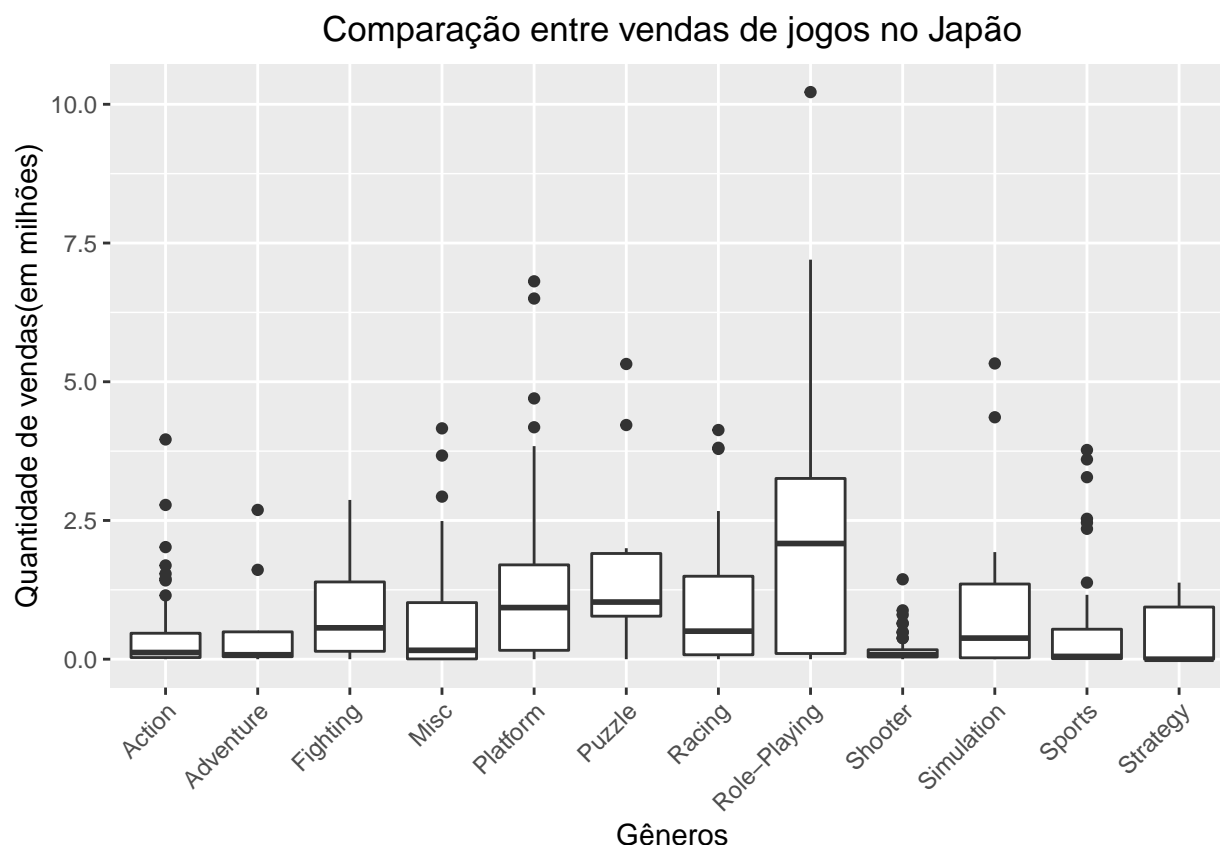
---

E com isso de fato podemos notar que nenhuma das comparações possui um resultado relevante, nesse caso menor que 0.05, para fazer-nos acreditar que essas variáveis são dependentes.

## Japão

Saindo um pouco do âmbito mundial e entrando em um mercado mais nichado e gostos mais bem definidos, o Japão, será que os nossos testes terão os mesmos resultados, ou será que encontraremos uma análise diferente?

Utilizando a mesma base de dados, porem agora olhando somente para as vendas em territorio japones encontramos o seguinte **boxplot**:



De cara já podemos notar que o resultado foi bem diferente que no exemplo anterior, aqui temos gêneros que possuem medianas visivelmente maiores que outros, mas será que isso já é o suficiente para tirar alguma conclusão? Da mesma forma que no anterior vamos prosseguir com os testes, começando novamente pelo de normalidade, para sabermos se devemos continuar utilizando os mesmos testes.

Genre	W	P.value
Sports	0.5858790	0.000000000012228525265
Platform	0.7690365	0.000000044166912781942
Racing	0.7934709	0.000019210835312851075
Role-Playing	0.8691679	0.000015784858675447007
Puzzle	0.8349910	0.027148302386793658758
Misc	0.7069434	0.000000059302487042848
Shooter	0.6400286	0.000000000002337199671
Simulation	0.7091305	0.000071560018458789482
Action	0.6299792	0.000000000000005905058
Fighting	0.8589037	0.003212977306670562749
Adventure	0.6439790	0.000515757515160936955
Strategy	0.7647707	0.040548191067493431772

O resultado do teste de Shapiro foi semelhante ao encontrado no primeiro testes, por isso vamos seguir usando o teste de Kruskal-Wallis para testar a dependência entre essas duas variáveis. Aplicando o teste nos nossos dados obtemos o valor de p igual a  $9.5001426 \times 10^{-13}$ , valor infinitamente menor que o limiar de 0,05, evidenciando que de fato o número total de vendas de um jogo é influenciado pelo seu gênero, quando o alvo de estudo é o território japonês.

Agora para sabermos quais gêneros levam vantagem ou desvantagem em relação a outros vamos aplicar novamente o teste de Pairwise Wilcox, que agora provavelmente vai nos demonstrar resultados relevantes, diferente do que aconteceu no estudo anterior.

Var1	Var2	value	Relevancia
Fighting	Action	0.01077359452329	Sim
Platform	Action	0.00000741537758	Sim
Platform	Misc	0.01366889699904	Sim
Puzzle	Action	0.00886443573062	Sim
Racing	Action	0.02690236343655	Sim
Role-Playing	Action	0.00000523328631	Sim
Role-Playing	Misc	0.00067048730074	Sim
Shooter	Fighting	0.00026415320411	Sim
Shooter	Platform	0.00000007761107	Sim
Shooter	Puzzle	0.00152245806517	Sim
Shooter	Racing	0.00112458687789	Sim
Shooter	Role-Playing	0.00000154768945	Sim
Sports	Fighting	0.01077359452329	Sim
Sports	Platform	0.00012794411386	Sim
Sports	Puzzle	0.01956250019497	Sim
Sports	Racing	0.01366889699904	Sim
Sports	Role-Playing	0.00001426116396	Sim
Adventure	Action	0.81311743359774	Não
Fighting	Adventure	0.28375341104008	Não
Misc	Action	0.91864399262561	Não
Misc	Adventure	0.91864399262561	Não
Misc	Fighting	0.14082062299164	Não
Platform	Adventure	0.16752123629894	Não
Platform	Fighting	0.53579586622058	Não
Puzzle	Adventure	0.25976950952714	Não
Puzzle	Fighting	0.29410553110611	Não
Puzzle	Misc	0.07129536791782	Não
Puzzle	Platform	0.57967447168975	Não
Racing	Adventure	0.36073788774976	Não
Racing	Fighting	0.87161787720090	Não
Racing	Misc	0.17673888849098	Não
Racing	Platform	0.36073788774976	Não
Racing	Puzzle	0.29410553110611	Não
Role-Playing	Adventure	0.07129536791782	Não
Role-Playing	Fighting	0.10774488001067	Não
Role-Playing	Platform	0.16752123629894	Não
Role-Playing	Puzzle	0.68899088931845	Não
Role-Playing	Racing	0.07129536791782	Não
Shooter	Action	0.20707904567062	Não
Shooter	Adventure	0.91864399262561	Não
Shooter	Misc	0.58222783145406	Não
Simulation	Action	0.46434513256315	Não
Simulation	Adventure	0.69792361104023	Não
Simulation	Fighting	0.53579586622058	Não
Simulation	Misc	0.65128193121220	Não
Simulation	Platform	0.29410553110611	Não
Simulation	Puzzle	0.28375341104008	Não
Simulation	Racing	0.68274857487949	Não

Simulation	Role-Playing	0.07129536791782	Não
Simulation	Shooter	0.20707904567062	Não
Sports	Action	0.28083165617477	Não
Sports	Adventure	0.69792361104023	Não
Sports	Misc	0.91864399262561	Não
Sports	Shooter	0.57967447168975	Não
Sports	Simulation	0.33704591648427	Não
Strategy	Action	0.57967447168975	Não
Strategy	Adventure	0.60868894158390	Não
Strategy	Fighting	0.29410553110611	Não
Strategy	Misc	0.53579586622058	Não
Strategy	Platform	0.20707904567062	Não
Strategy	Puzzle	0.20853070144825	Não
Strategy	Racing	0.30412751741868	Não
Strategy	Role-Playing	0.07129536791782	Não
Strategy	Shooter	0.68274857487949	Não
Strategy	Simulation	0.48144884934318	Não
Strategy	Sports	0.49652680684255	Não

---

Como era esperado, algumas combinações possuem resultado relevante, como no caso de ‘Shooter’ e ‘Role-Playing’, o teste de Pairwise apenas nos diz que estatisticamente falando jogos desses dois gêneros tendem a ter uma diferença de vendas, para saber qual deles é mais propício a atingir lucros maiores podemos olhar para o **boxplot** exibido no começo da análise. Analisando esses dois gêneros podemos ver que jogos de ‘Role-Playing’ são mais propícios a atingirem lucros maiores que jogos de ‘Shooter’.

Todas as análises e gráficos foram feitos utilizando o software R.