

Análise Estatística do Tamanho Médio de Trabalhos Processados em um Servidor de Jogos Online

Coleta e interpretação de dados

Statistical Analysis of the Average Size of Jobs Processed on an Online Game Server

Data collection and interpretation

I.A.C. Regô , J. P. S. Oliveira, R. F. Leão, T. O. D. S. C. Chacon

¹ Universidade Federal de Pernambuco - Centro de Informática

² Recife, Brasil

Iacr; jpso2; rfl5; tossc / @cin.ufpe.br

Resumo. Em Digitaland, um servidor de jogos online, o mundo virtual ganha vida. Nesse reino digital, jogadores de todo o planeta se reúnem para participar de aventuras épicas, enfrentar desafios emocionantes e buscar tesouros inimagináveis. No centro desse universo, está o servidor de jogos Reino dos Desafios, que atende aos jogadores com uma infraestrutura sólida e confiável. Porém, em meio a toda a diversão e emoção, algo curioso começou a chamar a atenção dos administradores do servidor. Eles notaram que em determinados momentos, o servidor tinha comportamentos diferentes. Os jogadores começaram a relatar atrasos nas ações do jogo, e a administração decidiu investigar a fundo esse mistério. Para desvendar os segredos dessa dinâmica intrigante, os administradores começaram a coletar dados sobre o comportamento dos jogadores, incluindo a média de jogadores online por dia. Eles também dividiram os dias da semana, em dias úteis e finais de semana. O objetivo principal era determinar se havia uma diferença significativa na carga do servidor entre finais de semana e dias úteis, conduzindo um teste de hipótese. O resultado indicou que não havia evidências significativas para rejeitar a hipótese nula. Isso significa que a média do tamanho médio de trabalhos nos finais de semana não é significativamente maior do que nos dias úteis. Com esse enigma decifrado, os administradores estão mais determinados do que nunca a manter o servidor no auge de seu desempenho. Os aventureiros virtuais podem continuar suas jornadas épicas em "Digitaland" com a certeza de que o "Reino dos Desafios" permanecerá um lugar emocionante e confiável onde histórias extraordinárias aguardam serem escritas e conquistas inesquecíveis aguardam serem alcançadas.

Palavras-chave: servidor de jogos online, administradores, comportamento, dias da semana, teste de hipótese, desempenho.

Abstract. On Digitaland, an online game server, the virtual world comes to life. In this digital realm, players from all over the planet come together to participate in epic adventures, face exciting challenges and seek unimaginable treasures. At the center of this universe is the Reino dos Desafios game server, which serves players with a solid and reliable infrastructure. However, amidst all the fun and excitement, something curious began to catch the attention of the server administrators. They noticed that at certain times, the server behaved differently. Players began to report delays in game actions, and the administration decided to investigate this mystery further. To unlock the secrets of this intriguing dynamic, administrators began collecting data on player behavior, including the average number of players online per day. They also divided the days of the week into weekdays and weekends. The main objective was to determine if there was a significant difference in server load between weekends and weekdays by conducting a hypothesis test. The result indicated that there was no significant evidence to reject the null hypothesis. This means that the average job size on weekends is not significantly higher than on weekdays. With this enigma solved, administrators are more determined than ever to keep the server at peak performance. Virtual adventurers can continue their epic journeys in "Digitaland" safe in the knowledge that the "Kingdom of Challenges" will remain an exciting and trustworthy place where extraordinary stories wait to be written and unforgettable achievements wait to be achieved.

Keywords: online game server, administrators, behavior, days of the week, hypothesis testing, performance.

1 Introdução

Nas vastas terras virtuais de Digitaland, o servidor de jogos online, "Reino dos Desafios", é o palco de aventuras emocionantes para jogadores de todo o mundo. No entanto, recentemente, a administração notou flutuações no desempenho do servidor, levando a atrasos nas ações dos jogadores. Determinados a resolver esse mistério, eles coletaram dados sobre o comportamento dos jogadores, focando na carga do servidor durante os dias da semana e os fins de semana. Através de um teste de hipótese e análise estatística, eles buscaram determinar se havia diferenças significativas. Os resultados forneceram insights valiosos, enquanto os aventureiros continuam suas jornadas em Digitaland, confiantes de que o "Reino dos Desafios" permanece emocionante e confiável.

Na Seção II deste artigo, a metodologia para tal investigação é apresentada, incluindo a realização da coleta de dados, as bibliotecas usadas para análise e as condições dos testes de hipóteses. A análise descritiva dos dados e a visualização por meio de gráficos como boxplot, histograma e gráfico de linhas estão na Seção III. Ainda nesta seção da discussão dos resultados, foram conduzidos dois testes de hipóteses: o primeiro para verificar a normalidade da distribuição dos dados e o segundo para avaliar se a média de acessos é maior durante o fim de semana do que durante a semana. A Seção IV traz as conclusões do trabalho. Por fim, a Seção V apresenta o dataset e o repositório criado para o desenvolvimento do projeto.

2 Metodologia

2.1 Coleta de dados

A coleta de dados foi realizada de forma sistemática e abrangente para obter informações detalhadas sobre o comportamento dos jogadores e a carga do servidor. Foi feito um monitoramento contínuo do servidor, o que possibilitou o registro da média de acessos de jogadores por dia, durante um tempo significativo. Isso ajudou a capturar variações diárias e semanais no comportamento do servidor. Para avaliar se havia diferenças significativas, os dados foram categorizados em dois grupos: dias da semana (incluindo segunda-feira a sexta-feira) e fins de semana (sábado e domingo). Essa divisão permitiu uma comparação direta entre os períodos. Essa abordagem na coleta de dados permitiu aos administradores obter uma visão abrangente do comportamento do servidor, identificar padrões e tendências, e, finalmente, conduzir uma análise estatística confiável para responder à questão principal sobre a diferença na carga do servidor entre os dias da semana e os fins de semana.

2.2 Bibliotecas usadas na análise de dados

As análises do projeto foram estruturadas em um arquivo com a extensão .ipynb, usando a linguagem de programação Python. Para calcular as medidas de centralidade e dispersão foram usadas NumPy e SciPy (modulo kurtosis). Para criar visualização dos dados de forma gráfica, aproveitaram-se das bibliotecas Matplotlib e Seaborn. Além disso, para ajudar a trabalhar no teste de hipótese foram usados SciPy, modulo stats, e o modulo norm para trabalhar com distribuição normal.

2.3 Teste de hipótese I – Distribuição normal

A normalidade da distribuição dos dados coletados foi atestada utilizando o gráfico Q-Q (quantile-quantile plot) e a função normaltest da biblioteca Scipy para calcular o valor-p, em um intervalo de confiança de 95%, sendo Hipótese Nula (H_0): A distribuição é normal e a Hipótese Alternativa (H_a): A distribuição não é normal. O teste usado foi o "teste de normalidade de D'Agostino e Pearson". A estatística do teste é baseada nas medidas de assimetria e curtose dos dados, que indicam o grau de desvio da distribuição dos dados em relação a uma distribuição normal.

2.4 Teste de hipótese 2 – Grau de acesso nos dias de semana

Foi realizado um teste de hipótese independente para verificar se o número de acessos no fim de semana é maior do que nos dias úteis da semana, em um intervalo de confiança de 95%. Usamos a biblioteca NumPy para dividir os dados entre dias úteis e

final de semana. Depois, é comparado o valor p com o α , se for maior, a hipótese nula (H_0) é rejeitada.

3 Resultado e discussão

3.1 Análise descritiva dos dados

Os tipos de dados analisados são do tipo quantitativo e contínuo já que são sobre o tamanho médio de trabalho, o que representa a media diária de jogadores online simultaneamente. Foram calculadas as medidas de centralidade e de dispersão, a fim de estudar a tendência central da base de dados, e apresentadas na Tabela 1.

Table 1. Medidas centrais de dispersão

Medidas	Valores
Média	234.79
Mediana	234.139
Moda	35,845
Desvio padrão	63.01
Variância	3971,4141
Coeficiente de variação	26.84
Q1	191.55
Q2	234.13
Q3	277.72
Curtose	-0.06

Utilizando as medidas de centralidade e de dispersão, foram elaborados os gráficos de boxplot (Figura 1) e histograma (Figura 2).

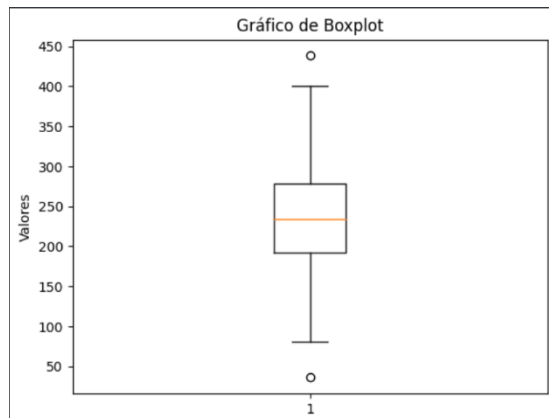


Fig. 1. Gráfico de Boxplot

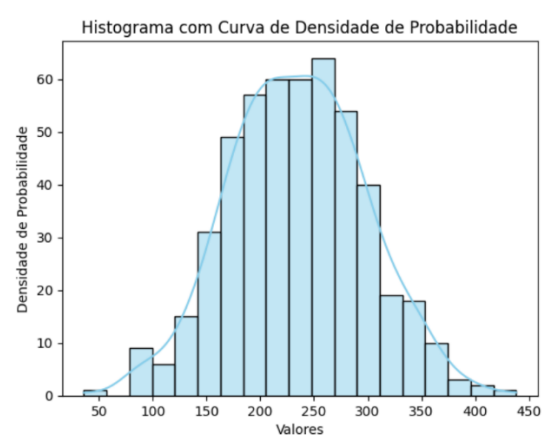


Fig. 2. Histograma



Fig. 3. Gráfico de linhas

A Figura 1 é um gráfico BoxPlot, ele começa no valor mínimo da base de dados e termina no valor máximo, o que mostra a pequena variação entre eles e baixo desvio padrão. Além de uma dispersão simétrica entre a mediana, sem outliers.

A Figura 2 é um histograma que avalia em função de distribuição de probabilidade, ele também é outra forma de identificar se é uma distribuição normal ou não. Nesse caso, fica evidenciado que é normal, já que os valores estão distribuídos simetricamente em torno do valor médio (isso será melhor trabalhado no teste de hipótese I).

3.2 Teste de Hipótese I – Distribuição normal

Inicialmente, a normalidade dos dados foi investigada utilizando o módulo `normal_test` e a função `qq_plot` da biblioteca Scipy, sendo plotado o gráfico Q-Q (Figura 4)

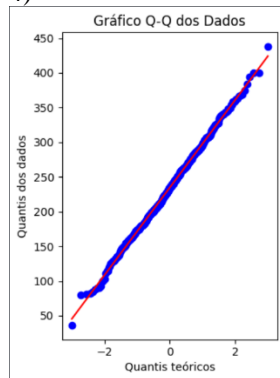


Fig. 4. Gráfico Q-Q dos dados

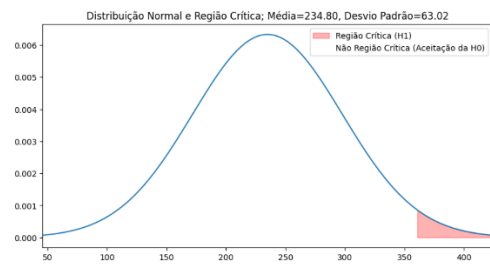
A Figura 4 mostra a normalidade da distribuição por meio de um gráfico Q-Q, assim, quanto mais perto os pontos estão da linha (que representa os quantis teóricos), mais próxima é a distribuição dos dados na normalidade. Ou seja, Podemos observar que os pontos estão perto da linha de referencia, evidenciando uma distribuição normal.

Além disso, também foi calculado o valor- $p = 0.9313$, considerando a Hipótese Nula (H_0): A distribuição é normal e a Hipótese Alternativa (H_a): A distribuição não é normal, para o intervalo de 95% de confiança. O valor- p é maior que o valor de alfa (0.05). Portanto, podemos inferir que há normalidade na distribuição da base de dados, aceitando a Hipótese Nula (H_0).

3.3 Teste de Hipótese II – Acesso nos dias da semana

Foi realizado um teste para verificar se o número de acessos era maior no fim de semana do que durante a semana. Assim, foi considerado Hipótese nula (H_0): A média do tamanho dos trabalhos durante a semana é maior do que nos fins de semana e Hipótese Alternativa (H_a): A média do tamanho médio de trabalhos nos finais de semana é significativamente maior do que nos dias úteis. A Figura 5 mostra a

distribuição. Como foi feito um teste independente, calculado o p valor = 0.9154 em um intervalo de confiança de 95%, ficou evidenciado que não temos evidências para rejeitar a hipótese nula, já que a média do tamanho dos trabalhos no fim de semana não é significativamente maior do que nos dias úteis.



4 Conclusões

Em "Digitaland", onde o universo virtual ganha vida, o servidor "Reino dos Desafios" permaneceu como uma âncora de diversão e aventura para jogadores de todo o mundo. O que antes parecia um mistério intrigante envolvendo o desempenho do servidor nos fins de semana e dias úteis agora está mais claro.

Após uma investigação metódica e um teste de hipótese, os administradores concluíram que não havia evidências significativas para rejeitar a hipótese nula. Isso significa que a média do tamanho médio de trabalhos nos fins de semana não é significativamente maior do que nos dias úteis.

Com esse enigma decifrado, os administradores estão mais determinados do que nunca a manter o servidor no auge de seu desempenho. Os aventureiros virtuais podem continuar suas jornadas épicas em "Digitaland" com a certeza de que o "Reino dos Desafios" permanecerá um lugar emocionante e confiável onde histórias extraordinárias aguardam serem escritas e conquistas inesquecíveis aguardam serem alcançadas.

5 Observações

Toda estrutura do Código e o material do projeto estará no repositório do GitHub: <https://github.com/rafaelaleao5/Projeto-estatistica>

References

1. Conheça o pacote Numpy - Python: aprenda a usar, <https://didatica.tech/o-pacote-numpy-python-para-machine-learning/>

2. Support | Minitab, <https://support.minitab.com/pt-br/minitab/21/help-and-how-to/statistics/basic-statistics/how-to/normality-test/interpret-the-results/key-results/#:~:text=Um%20n%C3%ADvel%20de%20signific%C3%A2ncia%20de,quando%20eles%20realmente%20a%20seguem.&text=Se%20o%20valor%20de%20p,n%C3%A3o%20seguem%20a%20distribui%C3%A7%C3%A3o%20normal.>
3. O que é boxplot? Aprenda a analisar esse tipo de gráfico!, <https://www.voitto.com.br/blog/artigo/boxplot>
4. O que é histograma, <https://www.tibco.com/pt-br/reference-center/what-is-a-histogram-chart#:~:text=Um%20histograma%20%C3%A9%20um%20gr%C3%A1fico,a%20cada%20um%20desses%20compartimentos.>