

# Análise dos Resultados do TreasureHunt de 2023

Ricardo de la Rocha Ladeira

28 de novembro, 2023

## Introdução

Este documento registra os resultados e a análise estatística sobre os dados de desempenho nas competições, a confiabilidade dos questionários aplicados e os resultados dos questionários relativos às competições geradas pelo TreasureHunt em 2023. O objetivo é manter a análise dos dados e sua interpretação no mesmo documento, para minimizar a desconexão entre esses dois aspectos. A interpretação dos dados se deu com base no que consta na referência abaixo:

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *biometrics*, JSTOR, p. 159–174, 1977.

Neste ano o jogo foi aplicado em uma turma no IFC campus Blumenau:

- turma de Segurança Computacional (Optativa), ministrada em 2023/2 para alunos da 8ª fase do curso de Bacharelado em Ciência da Computação, chamada aqui de **SUP**.

*Nota:* este relatório foi baseado no documento *Análise da Confiabilidade*, de Rafael R. Obelheiro, data: 11 de dezembro de 2017.

## Resultados de Desempenho

Foram aplicados oito exercícios distintos, acrescidos de um nono desafio escondido (*easteregg*). Na turma **SUP**, o tempo de aula destinado à atividade foi de 3 horas e 16 minutos.

A tabela a seguir resume os resultados de desempenho da turma. Cabe ressaltar que a taxa de submissões corretas foi próxima de 100%, o que indica compreensão adequada sobre o encontro de *flags*. Uma das submissões incorretas veio da digitação incorreta do ID do problema. Nesse sentido, deve-se manter os esforços acerca das explicações sobre as regras do jogo e o formato da flag em competições futuras com o objetivo de obter taxas superiores.

Indicador	SUP
Média de acertos da turma	5,67
Taxa de desempenho da turma	70,88%
Total de submissões da turma	53
Total de submissões corretas da turma	51
Taxa de submissões corretas da turma	96,23%

Obs.: não foi considerado o nono desafio na taxa de desempenho da turma. Caso ele seja considerado, a taxa de desempenho é de 63,00%.

Pelo que se pode observar nos registros de submissões, não houve tentativa de compartilhamento de *flags*, de contas, de comprometimento da plataforma ou de ataques de força bruta na tentativa de obter alguma resposta.

## Análise dos Questionários

Nesta seção analisamos os questionários aplicados, iniciando pela subseção de dados de entrada, onde listamos os dados brutos obtidos através das respostas dos jogadores nos questionários pré- e pós-competição. Os dados foram mapeados diretamente em números de 1 a 5, conforme Escala de Likert, onde 1 corresponde à menor resposta possível de ser sinalizada (muito baixa, muito desmotivador, discordo totalmente) e 5 corresponde à maior (muito alta, muito motivador, concordo totalmente). A segunda subseção traz a análise da confiabilidade do questionário. A terceira e última subseção descreve os resultados obtidos.

### Dados de entrada

Todos os dados de entrada referentes aos questionários (pré- e pós-competição) respondidos por estudantes que participaram do TreasureHunt em 2023 estão no *data frame* `quest`:

##	ID	Q11pre	Q12pre	Q13pre	Q14pre	Q15pre	Q16pre	Q17pre	Q21pre	Q11	Q12	Q13	Q14
## 1	A	3	4	5	3	5	3	4	2	5	5	5	4
## 2	B	4	4	5	3	5	4	5	3	4	4	4	2
## 3	C	4	5	5	4	4	4	5	4	4	5	5	3
## 4	D	4	4	3	4	4	3	5	2	4	4	4	4
## 5	E	4	4	4	2	5	1	3	3	4	5	4	1
## 6	F	4	5	5	3	5	2	4	4	5	5	5	4
## 7	G	5	4	3	2	5	3	5	1	2	4	5	3
## 8	H	4	5	4	1	4	1	4	3	3	5	5	4
## 9	I	4	4	5	1	5	1	4	3	5	5	3	3
##		Q15	Q16	Q17	Q21	Q31	Q32	Q41	Q42				
## 1		5	4	5	4	3	3	4	4				

```
## 2 4 1 4 3 4 3 4 4
## 3 5 2 5 4 4 4 4 4
## 4 5 4 5 2 4 3 4 4
## 5 4 1 4 3 4 5 2 4
## 6 4 5 4 4 3 3 4 4
## 7 4 4 4 4 2 3 2 4
## 8 5 3 5 2 2 3 4 5
## 9 5 4 5 2 4 4 4 4
```

Todos os registros são referentes à turma SUP.

## Análise da Confiabilidade dos Questionários

### Satisfação

As questões relativas à satisfação são 1.1, 1.2, 1.3, 1.5 e 1.7. Serão consideradas apenas as respostas do questionário pós-competição.

```
library(psych)
satisf <- quest[, c("Q11", "Q12", "Q13", "Q15", "Q17")]
alfa.satisf <- psych::alpha(satisf, na.rm = T, check.keys = T)

## In factor.stats, I could not find the RMSEA upper bound . Sorry about that

## In smc, smcs < 0 were set to .0
## In smc, smcs < 0 were set to .0
## In smc, smcs < 0 were set to .0
## In smc, smcs < 0 were set to .0
## In smc, smcs < 0 were set to .0

alfa.satisf.coef <- alfa.satisf$total$raw_alpha
```

O coeficiente  $\alpha$  de Cronbach é  $\alpha = 0.64$ , o que indica confiabilidade substancial, conforme LANDIS & KOCH (1977).

### Conscientização

As questões relativas à conscientização são 1.6 e 2.1. Serão consideradas apenas as respostas do questionário pós-competição.

```
conscient <- quest[, c("Q16", "Q21")]
alfa.consc <- psych::alpha(conscient, na.rm = T)
alfa.consc.coef <- alfa.consc$total$raw_alpha
```

O coeficiente  $\alpha$  de Cronbach é  $\alpha = 0.14$ , o que indica confiabilidade pequena.

### Nível de dificuldade do desafio

As questões relativas ao nível de dificuldade do desafio são 3.1 e 3.2, presentes apenas no questionário pós-competição.

```
dific <- quest[, c("Q31", "Q32")]
alfa.dific <- psych::alpha(dific, na.rm = T)
alfa.dific.coef <- alfa.dific$total$raw_alpha
```

O coeficiente  $\alpha$  de Cronbach é  $\alpha = 0.69$ , o que indica consistência substancial.

### Fatores de motivação

As questões relativas à competitividade e composição de problemas como fatores de motivação são 4.1 e 4.2, presentes apenas no questionário pós-competição.

```
motiv <- quest[, c("Q41", "Q42")]
alfa.motiv <- psych::alpha(motiv, na.rm = T)
alfa.motiv.coef <- alfa.motiv$total$raw_alpha
```

O coeficiente  $\alpha$  de Cronbach é  $\alpha = 0.22$ , o que indica consistência razoável.

### Comparação com os valores de anos anteriores

A tabela abaixo apresenta os valores obtidos calculando os índices gerais de  $\alpha$  de Cronbach nas competições de cada ano.

Fator	2017	2018	2019	2020	2021	2022	2023
Satisfação	0.78	0.92	0.76	0.65	0.77	0.00	0.64
Conscientização	0.60	0.44	0.68	0.39	0.67	0.00	0.14
Dificuldade	0.72	0.60	0.75	0.93	0.73	0.00	0.69
Motivação	0.55	0.86	0.61	0.00	0.86	0.00	0.22

Obs. 1: os valores de 2017 dizem respeito a três turmas (BCC, TADS e FIC).

Obs. 2: os valores de 2018 dizem respeito à turma de BCC da UDESC.

Obs. 3: os valores de 2019 dizem respeito à turma TADS, do IFC, e à turma de BCC, da UDESC.

Obs. 4: os valores de 2020 dizem respeito a somente uma turma do IFC, em uma aplicação remota do jogo.

Obs. 5: os valores de 2021 dizem respeito a duas turmas, uma de ensino superior e uma de ensino médio.

Obs. 6: os valores de 2022 ainda não foram calculados e por esse motivo estão

zerados. Em 2022 o jogo foi aplicado em duas turmas, uma de ensino superior e uma de ensino médio.

O conjunto de exercícios utilizados e os contextos de aplicação são diferentes, portanto, a tabela serve apenas para registrar os valores.

## Resumo e Discussão

A tabela abaixo traz o resumo dos valores de consistência interna por fator nos questionários em 2023. A coluna *SUP* traz os resultados obtidos através das respostas da turma do curso superior.

Fator	SUP
Satisfação	0.64
Conscientização	0.14
Dificuldade	0.69
Motivação	0.22

Os dados obtidos apresentaram consistências distintas, por isso, não permite sugerir consistência interna aos questionários dentro de cada fator. O fator para o qual mais houve confiabilidade nos questionários foi *dificuldade* no geral. O fator *conscientização* foi o que obteve o menor valor  $\alpha$  de Cronbach no geral, 0,14.

Turmas de Computação não costumam ser muito numerosas, o que representa um desafio para o cálculo do  $\alpha$  de Cronbach. Quanto maior o tamanho da amostra, mais confiável costuma ser o resultado. Isso pode ajudar a explicar a maior disparidade entre os dados obtidos com a turma *SUP*. Conforme dito na literatura, amostras maiores (e questionários também) tendem a tornar os resultados mais confiáveis. Há diversos estudos sobre qual deve ser o menor tamanho da amostra para aplicação deste coeficiente, embora não haja um consenso.

## Resultados dos Questionários

A tabela a seguir exhibe os resultados dos questionários aplicados com a turma *SUP*. Os valores presentes nas células *Pré* e *Pós* representam a média obtida para cada pergunta do questionário. A coluna *Evolução* contém a diferença entre as médias gerais do questionário pós-competição e do questionário pré-competição. A coluna *Significativa?* indica se essa diferença é estatisticamente significativa ( $p \leq 0,05$ ), calculada por meio do teste de Wilcoxon.

Embora tenha havido diferenças em todas as questões em termos absolutos, estatisticamente não é possível afirmar que há diferença de resultados.

No que diz respeito ao aprendizado (questão 1.4), a percepção de dificuldade das atividades de segurança diminuiu em termos absolutos, mas não em termos estatísticos. O ideal é que este indicador fique próximo ao neutro, pois atividades avaliativas não devem ser muito fáceis ou muito difíceis.

Com relação às questões sobre conscientização (1.6 e 2.1), as médias gerais ficaram entre *neutro* e *concordo parcialmente* após a aplicação da competição.

De forma geral, o tempo destinado à atividade e a dificuldade dos exercícios obtiveram média entre os valores *neutro* e *concordo parcialmente* (questões 3.1 e 3.2).

O resultado de motivação com a competitividade ficou entre *neutro* e *motivador*, e de motivação com a solução dos exercícios compostos ficou entre *motivador* e *muito motivador*, sugerindo que a atividade foi positiva neste aspecto.

As médias foram calculadas e arredondadas em duas casas decimais, da seguinte maneira: `mean ( na.omit ( quest$COLUNA ) )`. Exemplo:

```
media <- round(mean(na.omit(quest$Q11pre)), digits = 2)
```

Questão	Pré	Pós	Evolução	Significativa?
1.1	4.00	4.00	+0.00	Não (p = 0.7615)
1.2	4.33	4.67	+0.34	Não (p = 0.1857)
1.3	4.33	4.44	+0.11	Não (p = 0.8826)
1.4	2.56	3.11	+0.55	Não (p = 0.2896)
1.5	4.67	4.56	-0.11	Não (p = 0.6762)
1.6	2.44	3.11	+0.67	Não (p = 0.2723)
1.7	4.33	4.56	+0.23	Não (p = 0.5510)
2.1	2.78	3.11	+0.33	Não (p = 0.5163)
3.1	-	3.33	-	-
3.2	-	3.44	-	-
4.1	-	3.56	-	-
4.2	-	4.11	-	-

## Discussão

Uma das possíveis razões para que quase todos os resultados não apontem diferença estatisticamente significativa é o tamanho da amostra, limitando o poder do teste. Sugere-se sempre que a amostra seja grande o suficiente para estimar as variações, mas dado o contexto de ser uma disciplina ofertada para a oitava e última fase do curso, com evasão alta, não se conseguiu uma quantidade expressiva de jogadores. Estiveram presentes nove jogadores, exatamente os nove estudantes que frequentaram regularmente as aulas de Segurança Computacional. Ainda assim, considera-se que a aplicação do jogo foi positiva.

## Conclusão

A tabela abaixo mostra o resumo das conclusões que podem ser extraídas da análise dos dados:

Resultado estatístico	Conclusão
$\alpha$ de Cronbach igual a 0,64 em satisfação	As respostas obtidas mostraram que as questões relativas a este atributo foram consistentes substancialmente.
$\alpha$ de Cronbach = 0,14 em conscientização	As respostas obtidas mostraram que as questões relativas a este atributo obtiveram consistência pequena.
$\alpha$ de Cronbach igual a 0,69 em dificuldade	As respostas obtidas mostraram que as questões relativas a este atributo foram substancialmente consistentes.
$\alpha$ de Cronbach = 0,22 em motivação	As respostas obtidas mostraram que as questões relativas a este atributo obtiveram consistência razoável.
Diferenças não significativas nas questões ( $p > 0.05$ )	Em números absolutos, os estudantes melhoraram a percepção sobre a preparação para participação de competições de Segurança Computacional, com exceção da questão 1.5. No entanto, as diferenças não foram estatisticamente significativas e não é possível afirmar estatisticamente que há diferença entre as respostas.

Com relação aos resultados dos questionários, não houve diferença estatisticamente significativa entre as respostas dos questionários pré- e pós-competição. Embora a amostra seja pequena e possa limitar os resultados, o indicador com pequeno decréscimo entre as percepções pré- e pós-competição (1.5) será reavaliado para aplicação da competição em 2024.

Espera-se que em 2024 seja possível aplicar o jogo novamente em uma turma de ensino médio, bem como aplicar novamente em uma turma de curso superior.