

TESTES DE HIPÓTESE APLICADOS A RESULTADOS DE EXAMES ESTUDANTIS NOS ESTADOS UNIDOS DA AMÉRICA: UMA BREVE ANÁLISE DA SEGREGAÇÃO RACIAL E SOCIOECONÔMICA NO PAÍS

Vitor Eduardo Centurione MAGALHÃES

- **RESUMO:** O sistema estudantil americano é ponto de debate no atual cenário social e étnico do país. Segundo análise do Hamilton Project do ano de 2014, 14% da população dos Estados Unidos era composta por imigrantes. Este número, que simboliza um total de 43,5 milhões de indivíduos, impacta de forma direta o sistema educacional básico americano que, muitas vezes, acaba por segregar àqueles que não são nativos. Diante do cenário descrito e, conforme o banco de dados disponibilizado no KAGGLE, foram observadas informações de desempenho de 1000 alunos, abrangendo os períodos de dezembro de 2017 à dezembro de 2018. Para realização do presente estudo foram utilizadas técnicas de testes de hipótese do tipo T-Student, além do cálculo da média amostral para modelagem e análise dos dados.
- **PALAVRAS-CHAVE:** Teste de hipótese; distribuição normal; segregação racial; Estados Unidos; educação; pobreza; imigração.

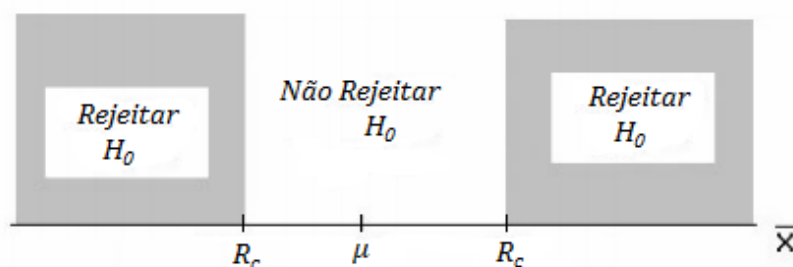
1. Introdução

Uma hipótese estatística é uma afirmação ou conjectura sobre o parâmetro, ou parâmetros, da distribuição de probabilidades de uma característica “X” da população ou de uma variável aleatória.

Os testes de hipótese têm como objetivo auxiliar na tomada de decisão relacionada a alguma característica de uma população, baseando-se em dados obtidos através apenas de uma parcela de indivíduos dessa mesma população com um grau de incerteza tolerável.

Ou seja, um teste de uma hipótese estatística consiste no procedimento utilizado para decisão por “ H_0 ”, chamado de hipótese nula, ou por “ H_a ”, chamado de hipótese alternativa.

É chamada de região crítica o conjunto de valores assumidos pela variável aleatória ou estatística de teste para os quais a hipótese nula é rejeitada.



$$\begin{aligned} \text{Se } \bar{X} \notin R_c &\rightarrow \text{Rejeita} - \text{se } H_0 \\ \text{Se } \bar{X} \in R_c &\rightarrow \text{Não rejeita} - \text{se } H_0 \end{aligned}$$

Para que haja realização correta de um teste de hipóteses, a hipótese nula “H₀” deve ser, primeiramente, interpretada como verdadeira e, através do teste realizado na hipótese alternativa “H_a” pode-se chegar a conclusão de rejeitar, ou não a hipótese nula “H₀”.

Tais testes podem ser realizados para os seguintes comportamentos: Teste bilateral; Teste unilateral esquerdo; Teste unilateral direito. Descrevendo matematicamente cada um deles temos que

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_a: \mu &\neq \mu_0 \end{aligned} \quad (1)$$

para Teste Bilateral;

$$\begin{aligned} H_0: \mu &\geq \mu_0 \\ H_a: \mu &< \mu_0 \end{aligned} \quad (2)$$

para Teste Unilateral Esquerdo e

$$\begin{aligned} H_0: \mu &\leq \mu_0 \\ H_a: \mu &> \mu_0 \end{aligned} \quad (3)$$

para Teste Unilateral Direito.

Além disso, para que haja execução correta, um teste de hipótese deve seguir as seguintes etapas:

- I – Definição das hipóteses nula e alternativa (“H₀” e “H_a” respectivamente).
- II – Identificação da estatística do teste e caracterização da sua distribuição.
- III – Definição da regra de decisão, com a especificação do nível de significância do teste.
- IV – Cálculo da estatística de teste e tomada de decisão.

No presente estudo, afim de observar os diferentes rendimentos escolares dos alunos, baseando-se em gênero e etnia, será realizado um teste do tipo T-Student para uma amostra de 30 indivíduos de cada categoria. Cada situação será comparada com os resultados obtidos tanto para todos os da mesma população (mesmo gênero e grupo de etnia), quanto para toda a população geral.

Matematicamente, o teste T-Student é aplicável para variância desconhecida e amostras pequenas. Considerando o fato do conjunto de estudo possuir apenas 1000 indivíduos, a amostra de 30 indivíduos por categoria, com nível de significância de 95%, deve apresentar um resultado satisfatório para observações.

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \quad (4)$$

Conforme mostrado em (4), a formulação matemática do teste é bem simples, onde:

\bar{X} : Média Amostral.

μ_0 : Valor fixo usado para comparação com a média da amostra.

s : Desvio padrão amostral.

n : Tamanho da amostra.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (5)$$

Quanto maior o valor de “T” encontrado, mais confiança temos ao rejeitar a hipótese nula “H₀”, ou seja, mais certeza temos ao afirmar que $\bar{X} \leq \mu_0$ não é verdadeiro.

2. Desenvolvimento e implementação do algoritmo de precisão para cálculo dos testes de hipótese para cada uma das variáveis aleatórias

2.1 Preparação e análise geral dos dados

Para demonstrar a metodologia aplicada e verificar o desempenho do teste para cada uma das categorias, é considerado um conjunto de dados composto por todos os indivíduos da amostra segmentados pelo gênero (482 indivíduos do sexo masculino e 518 indivíduos do sexo feminino). Os dados de estudo são: gênero, grupo étnico, pontuação na prova de matemática, pontuação na prova de escrita e pontuação na prova de leitura.

O objetivo de análise dessas informações é estudar o rendimento dos diferentes grupos no exame, podendo assim sugerir uma possível segregação de algum grupo étnico, mostrando uma possível marginalização por parte da sociedade (os grupos étnicos foram segmentados por letras, para que assim haja interpretação puramente técnica sobre os fatos coletados e não ocorra nenhum tipo de violação ou discriminação). A segmentação em gênero também é útil para demonstrar melhor os resultados obtidos.

Após os primeiros procedimentos de segmentação dos dados, foram obtidas as seguintes informações à respeito da população geral:

Tabela 1 – Medidas coletadas da amostra geral

População	Disciplina	Média	Mediana	Desvio Padrão	Variância
Geral	Matemática	66,089	66	15,163	229,919
	Leitura	69,169	70	14,600	213,165
	Escrita	68,054	69	15,195	230,908
Masculino	Matemática	68,728	69	14,356	206,102
	Leitura	65,473	66	13,931	194,095
	Escrita	63,311	64	14,113	199,200
Feminino	Matemática	63,632	65	15,491	239,985
	Leitura	72,608	73	14,378	206,733
	Escrita	72,467	74	14,844	220,369

A coleta dessas informações é útil para, ao fim do teste de hipótese, observarmos qual das medidas mostra-se mais próxima e, portanto, eficiente em relação ao resultado obtido para cada grupo étnico.

Para cada grupo étnico, segmentado anteriormente por gênero, foi separada uma amostra de 10 indivíduos afim de aferir se, com nível de significância de 95%, pode-se rejeitar a hipótese nula que, afirma que a média em cada disciplina para todos os grupos é de 60 pontos.

Para que o teste seja realizado da forma correta, antes verificou-se a normalidade das distribuições para cada grupo e disciplina. Além disso, verificou-se que, para nenhum dos segmentos existiam outliers, portanto sua observação gráfica através de boxplot só ficará disponível através do código que seguirá após o artigo.

Os gráficos de verificação de normalidade obtidos foram os que seguem com as respectivas legendas de cada grupo e gênero.

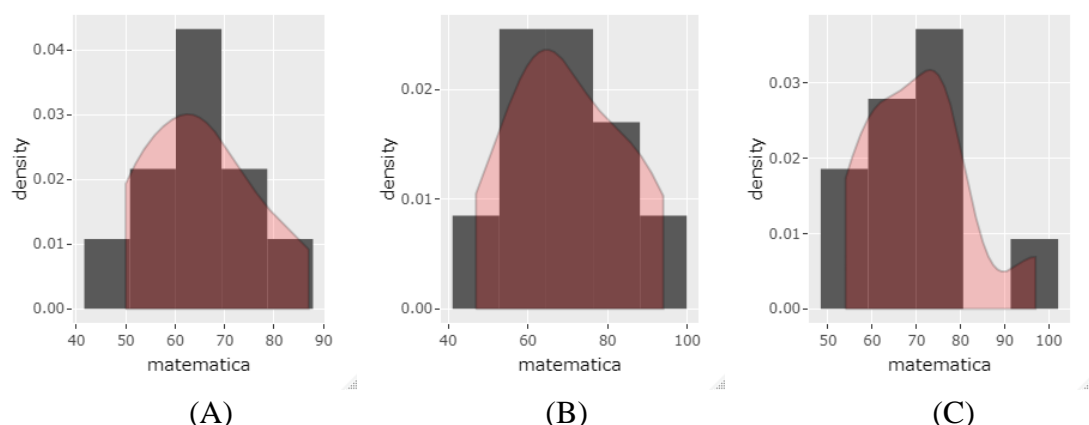


Figura 1 - Normalidade dos dados de matemática dos Grupos A, B e C masculinos

Conforme pode-se observar, todos os gráficos referentes aos desempenhos de matemática dos grupos A, B e C, das amostras masculinas, apresentam proximidade à normalidade e, portanto, não serão problema no momento de realização do teste T-Student.

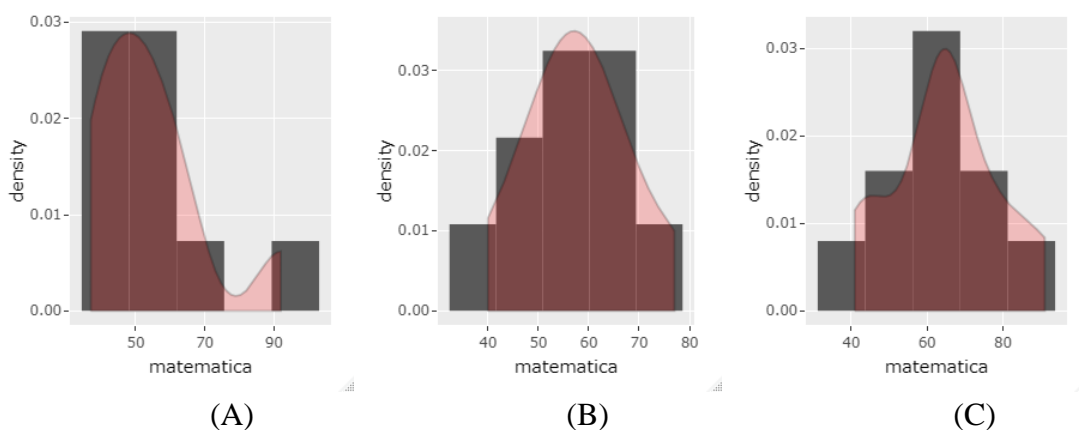


Figura 2 - Normalidade dos dados de matemática dos Grupos A, B e C femininos

O mesmo pode ser observado nos gráficos obtidos através da amostra feminina para o desempenho na disciplina de matemática. Em ambos os casos existe um comportamento similar um pouco mais distante da normalidade (grupo C – masculino, grupo A - feminino), porém esse comportamento não afetará nosso elemento de estudo, podendo assim ser desprezado.

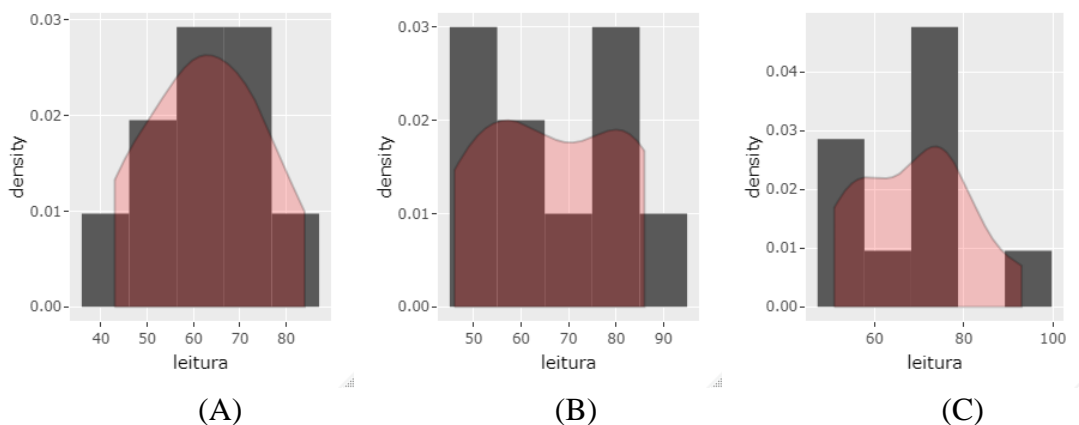


Figura 3 - Normalidade dos dados de leitura dos Grupos A, B e C masculinos

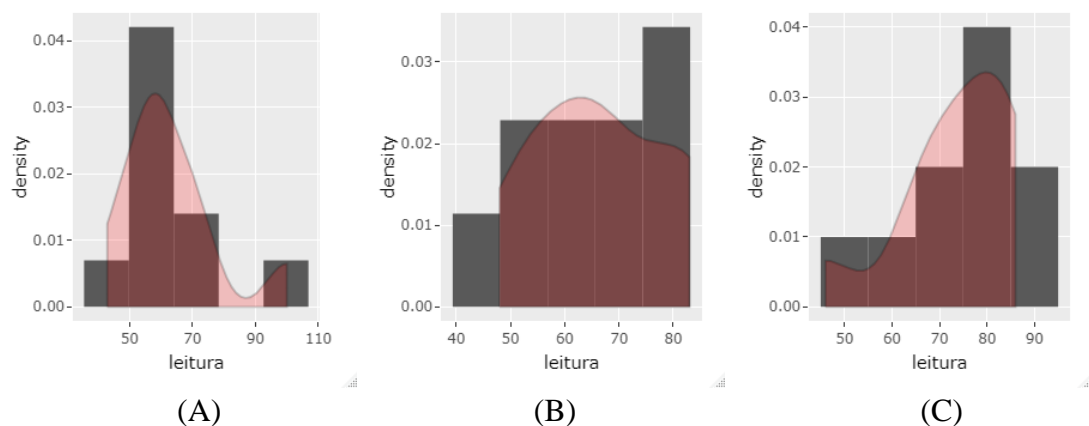


Figura 4 - Normalidade dos dados de leitura dos Grupos A, B e C femininos

Assim como nos dados de matemática, apesar dos grupos B e C neste caso se distanciarem um pouco mais do modelo em ambos os gêneros, é possível observar que apresentam certa normalidade e, assim será considerado para realização do teste em cada um dos segmentos.

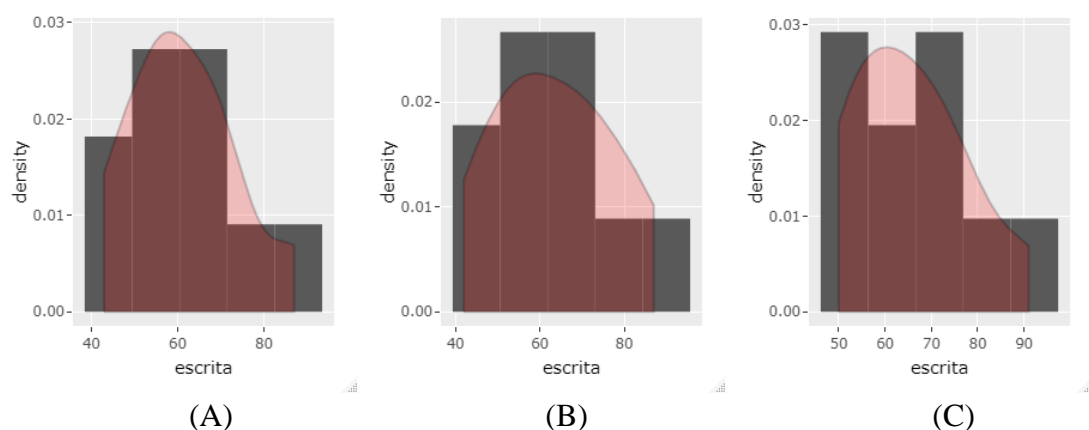


Figura 5 - Normalidade dos dados de escrita dos Grupos A, B e C masculinos

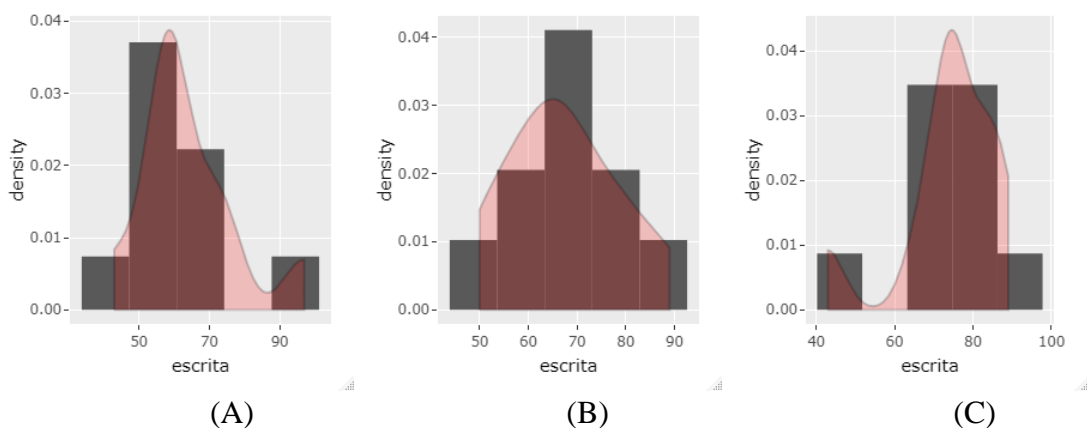


Figura 6 - Normalidade dos dados de escrita dos Grupos A, B e C femininos

Para os dados de escrita das amostras masculina e feminina, podemos observar nitidamente uma normalidade nos dados mais amigável do que nas duas ocasiões anteriores no caso dos três grupos étnicos.

2.2 Realização dos testes de hipótese com nível de significância de 95%

Através da biblioteca Tidyverse da linguagem de programação R, os testes de hipótese do tipo T-Student podem ser facilmente executados desde que os parâmetros de entrada (argumentos da função) satisfaçam as condições matemáticas do teste.

A função pede como argumentos um valor recebido como “x”, correspondente ao vetor numérico com os valores da métrica que será testada; um parâmetro chamado “alternative” que é responsável por determinar se o teste realizado será bicaudal, unicaudal direito ou unicaudal esquerdo; um termo chamado de “mu”, correspondente à média amostral a ser contestada para o valor “ H_0 ”.

A saída de informações do teste nos fornece os valores de “T”, conforme (4), o p-valor (valor que representa o menor nível de significância com que se rejeita a hipótese nula “ H_0 ”), o intervalo de confiança e a média amostral do conjunto. Desses valores, o fundamental para nossa interpretação é o p-valor, pois baseado nele vamos rejeitar, ou não, a hipótese nula de que a pontuação média é de 60 pontos.

Os valores encontrados para os P-valores das amostras segmentadas foram inseridos na seguinte tabela:

Tabela 2 – P-Valor encontrado para cada amostra segmentada

Gênero	Disciplina	Grupo	P-Valor	Média Amostral
Masculino	Matemática	A	0,1931	65,2
		B	0,0647	69,8
		C	0,0252	70,5
	Leitura	A	0,5505	62,5
		B	0,1959	66,7
		C	0,0571	68,9
	Escrita	A	0,7587	61,3
		B	0,4873	63,3
		C	0,1990	65,6
Feminino	Matemática	A	0,2712	54,0
		B	0,5305	57,8
		C	0,3396	64,8
	Leitura	A	0,5761	62,9
		B	0,1269	66,7
		C	0,0065	73,8
	Escrita	A	0,4080	64,0
		B	0,0814	67,3
		C	0,0072	74,1

Conforme a teoria, os P-Valores encontrados servem como parâmetro para rejeitar ou não a hipótese nula “ H_0 ”. Dito isso, podemos fazer algumas análises com relação tanto ao P-Valor, quanto à média amostral encontrada.

Para que seja mais fácil explicitar as evidências, o seguinte quadro traz de forma objetiva qual o significado aplicável do p-valor para o caso de estudo em questão.

Tabela 3 – Relação entre o P-Valor e H_0

Comportamento do P-Valor	Evidenciação contra H_0
$p - \text{valor} \geq 0,10$	Não existe evidência contra H_0
$p - \text{valor} < 0,10$	Fraca evidência contra H_0
$p - \text{valor} < 0,05$	Evidência significativa contra H_0
$p - \text{valor} < 0,01$	Evidência altamente significativa contra H_0
$p - \text{valor} < 0,001$	Evidência extremamente significativa contra H_0

Observando então essas informações e fazendo a análise detalhada dos dados contidos na Tabela 2, temos que:

Para o grupo A – masculino, tanto na disciplina de matemática, quanto nas disciplinas de leitura e escrita, não rejeita-se H_0 , ou seja, não é possível afirmar, com nível de significância de 95%, que a média de pontos seja diferente de 60 pontos. No caso da disciplina de matemática, a probabilidade de uma amostra tirada dessa população apresentar a média observada é de 19,31%, o que não nos permite tirar muitas conclusões a respeito do grupo A - masculino de modo geral. O mesmo vale para as duas outras disciplinas desse grupo e gênero, uma vez que os p-valores apresentados não evidenciam nada contra a hipótese nula “ H_0 ”.

Para o grupo B – masculino, as disciplinas de leitura e escrita não nos permitem concluir que existem evidências contra a hipótese nula “ H_0 ”, o que significa que, mesmo que a média amostral encontrada seja diferente da hipótese nula, não se pode afirmar que, com nível de significância de 95%, “ H_0 ” pode ser rejeitada. Contudo para a disciplina de matemática temos o apontamento de que exista uma evidência mesmo que fraca, contra “ H_0 ”, uma vez que a probabilidade de uma amostra tirada dessa população apresentar a média observada é de apenas 6,47%.

Para o grupo C – masculino, pode-se observar que, para disciplina de matemática, temos uma evidência significativa para rejeição da hipótese nula “ H_0 ”, já que a probabilidade de uma amostra tirada dessa população apresentar a média observada é de apenas 2,52%, o que nitidamente mostra um comportamento muito atípico na população. Para a disciplina de leitura, existe uma evidência fraca de rejeição da hipótese nula, porém é válido salientar que o p-valor encontrado aproxima-se muito da conclusão de evidência significativa contra “ H_0 ”. Ainda assim, não satisfaz a condição de atender um nível de significância de 95%. Para a disciplina de escrita nada se pode evidenciar contra “ H_0 ”.

Para o grupo A – feminino, nenhum dos p-valores encontrados para as disciplinas mostram evidência suficientemente concreta para rejeição da hipótese nula “ H_0 ”, ou seja, não se pode rejeitar a hipótese de que a média de pontos em todas as disciplinas seja de 60 pontos.

Para o grupo B – feminino, pode se observar que, nem na disciplina de matemática, nem na de leitura, foi possível evidenciar algo contra a hipótese nula “ H_0 ”. O mesmo não se repete para as médias da disciplina de escrita, que demonstra uma evidência fraca contra a hipótese nula, já que possui probabilidade de ocorrência de apenas 8,14%.

Para o grupo C – feminino, no que diz respeito a disciplina de matemática, nada se pode evidenciar contra a hipótese nula “ H_0 ”, ou seja, não se pode rejeitar a média de 60 pontos para essa disciplina, uma vez que a probabilidade de uma amostra tirada dessa população apresentar a média observada é de 33,96%. Contudo, é interessante observar que, tanto para leitura quanto para escrita, os p-valores encontrados mostram evidências extremamente fortes de rejeição de “ H_0 ”, o que nos diz que, para o caso de leitura podemos afirmar, com nível de significância de

99,35%, que a hipótese nula deve ser rejeitada e que, para o caso de escrita podemos afirmar, com nível de significância de 99,28%, que a hipótese nula deve ser rejeitada.

Com isso podemos estabelecer algumas relações entre os grupos étnicos e entre os gêneros. Uma das observações cabíveis é a de que o grupo A, tanto para o sexo feminino, quanto para o sexo masculino, apresentou uma média amostral muito próxima e não rejeitou " H_0 ", demonstrando assim que, para ambos os sexos a hipótese nula fornece resultado para média em literatura satisfatório. Os p-valores de literatura para o grupo A são muito próximos também, o que demonstra que, a probabilidade de uma amostra tirada dessa população apresentar a média observada, é razoavelmente alta.

Também é possível interpretar que, os menores p-valores encontrados são pertencentes ao grupo C, tanto para o sexo feminino, quanto para o masculino. Isso mostra que, naquele grupo, a média encontrada não condiz com aquela proposta anteriormente como 60 pontos, mesmo que apenas para matemática no caso masculino e, literatura e escrita para o caso feminino. Nos casos em questão é possível observar as maiores médias amostrais do estudo, o que possivelmente esboça que, naquele grupo a hipótese nula " H_0 " foi rejeitada pois pontuações mais altas ocorrem com maior frequência.

Outra observação interessante a se fazer é a de que no grupo B, para ambos os sexos, na disciplina de literatura, a mesma média amostral foi encontrada. Apesar dos p-valores apenas confirmarem que não se pode rejeitar a hipótese nula " H_0 ", podemos observar certa proximidade entre eles, com variação de apenas 6,90%, o que possivelmente indique que nesse grupo sejam mais frequentes notas intermediárias.

3. Conclusões

Conforme pôde-se observar no decorrer do estudo, se seguidos os procedimentos da forma correta, os testes de hipótese do tipo T-Student podem demonstrar informações bastante satisfatórias para uma análise inicial. Apenas respeitando as normalidades e filtrando os outliers é possível obter diversas informações de interesse à respeito do dataset. Conforme foi feita a segmentação dos grupos de interesse através do apoio do software R, as etapas de estudo pontual de cada caso e a aplicação do teste para cada uma das amostras de 10 indivíduos foi bastante conclusiva e mostrou-se muito útil, principalmente quando o conjunto de dados é grande demais para ser estudado por completo.

Através dos dados obtidos nas Tabelas 1, 2 e 3, é possível observar que a discrepância de pontuação entre os diferentes grupos étnicos estudados, dos diferentes gêneros, não foi exorbitante, apesar de pouco poder se concluir sobre estes.

Referências

- TRIOLA, Mario F. et al. **Introdução a Estatística**. 7ª Edição. Livros Técnicos e Científicos Editora S. A. Rio de Janeiro, 1998.
- Bussab, WO; Morettin, PA. **Estatística Básica**. São Paulo: Editora Saraiva, 2006 (5ª Edição).
- Montgomery, DC; Runger, GC. **Estatística aplicada e probabilidade para engenheiros**
- <http://leg.ufpr.br/~paulojus/CE003/ce003/node6.html>
- https://repositorioaberto.uab.pt/bitstream/10400.2/5952/10/R_text_v9_ReposAb.pdf
- https://www.ime.unicamp.br/~hlachos/Inferencia_Hipo1.pdf
- <https://www2.unifap.br/herondino/files/2014/04/8-DISTRIBUI%C3%87%C3%83O-T-STUDENT.pdf>

https://www.hamiltonproject.org/charts/immigrants_are_both_more_likely_to_lack_a_high_school_degree_and_more_likely_to_be_unemployed

https://www.hamiltonproject.org/charts/high_skilled_immigration_increases_innovation