

DENILSON JUNIO MARQUES SOARES

**TEORIA CLÁSSICA DOS TESTES E TEORIA DE RESPOSTA AO ITEM
APLICADAS EM UMA AVALIAÇÃO DE MATEMÁTICA BÁSICA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2018

DENILSON JUNIO MARQUES SOARES

**TEORIA CLÁSSICA DOS TESTES E TEORIA DE RESPOSTA AO ITEM
APLICADAS EM UMA AVALIAÇÃO DE MATEMÁTICA BÁSICA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 20 de fevereiro de 2018.

Ana Carolina Campana Nascimento

Silvana Cláudia dos Santos

Paulo César Emiliano
(Orientador)

*A meus pais, Denilson e Eliane,
pelo amor incondicional*

AGRADECIMENTOS

Agradeço primeiramente a Deus, que de forma singular me conduziu até aqui. Sei que ainda há um vasto caminho a percorrer, mas estou certo que Sua companhia me fortalece, me dando paz, saúde e sabedoria.

Também agradeço a meus pais, Denilson e Eliane, a minha irmã, Raquel, e a minha noiva, Talita, pelo importante e constante apoio que foram fundamentais em certos momentos.

Ao Professor Paulo César Emiliano, pelo apoio e incentivo, pela paciência, solicitude e principalmente pela confiança depositada em mim.

Aos professores e servidores do departamento de estatística da UFV por sempre estarem à disposição para esclarecimentos e por exercerem brilhantemente suas funções. Vocês são exemplos a serem seguidos por mim.

Ao departamento de matemática da UFV, na pessoa da professora doutora Lana Mara Rodrigues dos Santos, pela disponibilidade dos dados que foram utilizados neste trabalho.

Aos membros da banca que com observações relevantes puderam acrescentar e melhorar este trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por conceder a bolsa e financiar a pesquisa.

Aos amigos de Viçosa, Ubá, Piumhi e a todos que direta ou indiretamente, contribuíram para a realização deste trabalho.

Muito obrigado!

“A persistência é o menor caminho do êxito.”

Charles Chaplin

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	x
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	5
2.1 Teoria clássica dos testes (TCT)	5
2.1.1 Postulados da teoria clássica dos testes	6
2.1.2 Precisão de um teste	7
2.1.2.1 Coeficiente alfa de Cronbach	9
2.1.2.2 Fórmula vinte de Kuder-Richardson	10
2.1.2.3 Fórmula de predição de Spearman-Brown	10
2.1.3 Análise de itens pela teoria clássica dos testes	11
2.1.3.1 Estatísticas descritivas	11
2.1.3.2 Índice de dificuldade	12
2.1.3.3 Discriminação do item	13
2.1.4 Alguns problemas enfrentados pela TCT	19
2.2 Teoria de resposta ao item (TRI)	21
2.2.1 Pressupostos da TRI	21
2.2.2 Modelos unidimensionais	25
2.2.2.1 Modelo logístico de um parâmetro	26
2.2.2.2 Modelo logístico de dois parâmetros	28
2.2.2.3 Modelo logístico de três parâmetros	31
2.2.3 Estimação	32
2.2.4 Critérios para seleção de modelos	35
2.2.4.1 Critérios de informação de Akaike (AIC) e bayesiano (BIC)	35
2.2.4.2 Teste da razão de verossimilhanças	35
2.2.5 Qualidade do ajuste	36
2.2.6 Função de informação	36
3 MATERIAL E MÉTODOS	39
3.1 Material	39
3.2 Métodos	40

3.2.1	Análise dos itens pela TCT	41
3.2.2	Análise dos itens pela TRI	42
4	RESULTADOS E DISCUSSÃO	44
5	CONCLUSÕES	57
	REFERÊNCIAS	59
	APÊNDICE	65

LISTA DE FIGURAS

Figura 1	Representação gráfica do método de Kelley.	14
Figura 2	Curvas para itens baseadas na proporção de acertos.	15
Figura 3	Exemplo de uma curva característica do item.	26
Figura 4	Índice de dificuldade de dois itens de um teste.	27
Figura 5	Índice de discriminação negativo.	29
Figura 6	Curvas características dos itens para diferentes valores do coeficiente de discriminação a	30
Figura 7	Parâmetros da curva característica do item.	32
Figura 8	Função de informação para quatro itens.	38
Figura 9	Modelo do banco de dados.	40
Figura 10	Distribuição do número de acertos dos estudantes.	44
Figura 11	Item 3 e histograma das alternativas marcadas.	46
Figura 12	Item 16 e histograma das alternativas marcadas.	47
Figura 13	Item 17 e histograma das alternativas marcadas.	47
Figura 14	Curva característica do teste para o modelo logístico de dois parâmetros.	49
Figura 15	Item 15 e histograma das alternativas marcadas.	52
Figura 16	Curvas características dos itens.	53
Figura 17	Item 4 e histograma das alternativas marcadas.	53
Figura 18	Item 18 e histograma das alternativas marcadas.	54
Figura 19	Item 46 e histograma das alternativas marcadas.	54
Figura 20	Curvas de informação.	55

LISTA DE TABELAS

Tabela 1	Índice de reprovação em cálculo I na UFV - MAT 140	2
Tabela 2	Índice de reprovação em cálculo I na UFV - MAT 146	2
Tabela 3	Exemplos de índices de discriminação	14
Tabela 4	Classificação do item do teste, de acordo com o índice de discriminação	14
Tabela 5	Correlação ponto-bisserial para um teste de dez itens	18
Tabela 6	Estatística Kaiser-Meyer-Olkin (KMO)	23
Tabela 7	Carga fatorial: orientações para o tamanho da amostra	24
Tabela 8	Classificação para o índice de discriminação	30
Tabela 9	Tópicos que compõem o teste analisado	39
Tabela 10	Análise clássica do teste	45
Tabela 11	Comparação entre os modelos logísticos de 1 e 2 parâmetros	48
Tabela 12	Comparação entre os modelos logísticos de 2 e 3 parâmetros	48
Tabela 13	Teste χ^2 de Bock para a qualidade do ajuste	50
Tabela 14	Parâmetros de discriminação (a) e dificuldade (b) dos itens	51

RESUMO

SOARES, Denilson Junio Marques, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica.** Orientador: Paulo César Emiliano. Coorientadores: Sebastião Martins Filho e Lana Mara Rodrigues dos Santos.

O departamento de matemática da Universidade Federal de Viçosa tem realizado algumas medidas de prevenção à reprovação dos estudantes na disciplina de cálculo, cujos índices têm assustado os professores da disciplina. Uma dessas medidas está na elaboração de avaliações de matemática básica, compostas por 50 itens de múltipla escolha cada, que objetivam analisar o conhecimento prévio do estudante matriculado, bem como estimar sua habilidade em matemática e, com isto, identificar os estudantes propensos à reprovação a fim de proporcionar uma intervenção pedagógica capaz de reverter esta situação. Tendo em vista a importância desta avaliação, que acontece no início de cada período letivo, para o processo de ensino-aprendizagem da disciplina, o presente trabalho teve como objetivo oferecer uma análise estatística destas avaliações com base na teoria clássica dos testes (TCT) e na teoria de resposta ao item (TRI). A principal diferença entre essas teorias está no fato de que a primeira oferece uma abordagem em que se investiga o escore total do teste, ao passo que na segunda, se investigam as propriedades de cada item, individualmente. A metodologia se desenvolveu inicialmente como bibliográfica através de um referencial teórico que englobou ambas as teorias utilizadas e, na sequência, com um estudo de caso, em que foi utilizado uma dessas avaliações na coleta e análise dos dados. Os resultados dos desempenhos dos estudantes na TRI foram calculados pelo modelo logístico de dois parâmetros. Em suma, tais resultados, em consonância com resultados os obtidos pela análise via TCT, apontam para uma avaliação positiva do teste. Com a exceção de três itens que apresentaram problemas de discriminação, o teste apresentou, em geral, itens com um bom poder discriminativo e com variados níveis de dificuldade em ambas as teorias analisadas. A mesma análise pode ser realizada para outros testes e espera-se que este trabalho sirva como um instrumento de difusão destas teorias na análise de testes.

ABSTRACT

SOARES, Denilson Junio Marques, M.Sc., Universidade Federal de Viçosa, February, 2018. **Classical test theory and item response theory applied to basic mathematics evaluation.** Advisor: Paulo César Emiliano. Co-advisors: Sebastião Martins Filho and Lana Mara Rodrigues dos Santos.

The mathematics department of the Federal University of Viçosa has been working at actions to improve the success rate of the students in the Calculus course, whose failure increasing indexes has surprised the professors. One of these actions is focused on the elaboration of a mathematics basic test composed of 50 multiple choice items that aim to analyze the prior knowledge of the enrolled student, as well as to estimate their ability in basic mathematics and, with this, to identify students who are prone to flunk in order to provide a pedagogical intervention capable of reversing this situation. Considering their importance of this evaluation, which happens at the beginning of academic semester, for the teaching-learning process of the discipline, this study aims to provide a statistical analysis of one of these tests based on the classical test theory (CTT) and on item response theory (IRT). The main difference between these two theories lies on the fact that the first one uses the total score in the test as a reference in the analysis, while the second one focuses mainly on the item and not on the whole test itself. The methodology was initially developed as a bibliography through a theoretical framework that encompassed both theories used and, subsequently, with a case study, in which one of these evaluations was used in data collection and analysis. The results point to a positive evaluation of the test. Except three items presenting discrimination problems, the test presented, in general, items with a good discriminative power and with varied levels of difficulty in both theories that were analyzed, besides indicating insignificant probability of getting correct answer just by guessing, which contributed to a good estimation of the students ability. The same analysis can be performed for other tests and it is expected that this work serves as an instrument for the diffusion of these theories in the test analysis.

1 INTRODUÇÃO

O aumento no índice de reprovação dos estudantes é uma realidade em todos os níveis de ensino e tem sido objeto de reflexões e estudos de pesquisadores e educadores das mais diversas áreas. No ensino superior, esse aumento tem sido um obstáculo enfrentado por diversas instituições por estar diretamente relacionado à retenção e à evasão, gerando grande desperdício de recursos financeiros, carência de profissionais capacitados no mercado de trabalho e pondo fim ao sonho de muitos estudantes.

Veloso & Almeida (2013), por exemplo, trazem algumas reflexões acerca do problema da evasão que a Universidade Federal de Mato Grosso (UFMT) tem enfrentado, em que sugerem que o problema é um fenômeno institucional, reflexo da ausência de uma política de permanência do aluno no curso de sua opção. Piacentini (2012) traz um estudo de caso referente aos índices de reprovação, abandono e evasão de um curso da Universidade Tecnológica Federal do Paraná (UTFPR), sugerindo a criação de programas de tutoria e nivelamento, além de auxílios financeiros para estudantes carentes, visando estacionar o aumento destes índices. Vasconcelos; Almeida & Monteiro (2009) fazem esta mesma análise tendo como objeto de pesquisa estudantes dos cursos de engenharia da Universidade do Minho, Portugal e concluem que, entre outros fatores, a distância familiar e a matrícula em curso de graduação que não corresponde à primeira opção de escolha também colaboram para este aumento.

Embora se configure como uma disciplina aplicável em diversas áreas do conhecimento, o cálculo tem um incômodo índice de reprovação. Em particular, na Universidade Federal de Viçosa (UFV), em que diversos cursos de graduação possuem em sua matriz curricular a disciplina em questão, este índice tem chamado a atenção dos professores da disciplina que estão se mobilizando para contornar esta situação.

Na UFV, a disciplina de cálculo se apresenta em 2 vertentes: cálculo I (MAT 140 e MAT 146) e cálculo diferencial e integral I (MAT 141). O que difere as turmas de cálculo I são os cursos para os quais elas são ofertadas. Enquanto que a MAT 140 é ofertada para as engenharias química, de alimentos, civil, agrícola e ambiental, ambiental, mecânica, de produção, além dos cursos de bioquímica, ciência da computação e química, a MAT 146 é ofertada para os cursos de agronomia, ciência e tecnologia de laticínios, administração, engenharia florestal e ciências contábeis. Já a MAT 141 é ofertada exclusivamente para estudantes dos cursos de matemática, física e engenharia elétrica.

Um levantamento simples dos dados passados referente ao índice de reprovação dessas turmas entre os anos de 2011 e 2016, nos permitiram identificar que em cálculo I este problema é mais agudo. As Tabelas 1 e 2 nos mostram como estes índices se apresentaram para esta disciplina em MAT 140 e em MAT 146, respectivamente. Cabe ressaltar que no ano de 2012 devido a uma greve dos servidores da universidade, ao retornar as atividades acadêmicas o estudante pôde efetuar o trancamento da disciplina, o que pode responder pelo diferente índice de reprovação para este ano.

Tabela 1 Índice de reprovação em cálculo I na UFV - MAT 140

Ano	Número de estudantes matriculados	Número de estudantes aprovados	Índice de reprovação
2011	535	233	0,56
2012*	511	291	0,43
2013	594	239	0,60
2014	539	195	0,66
2015	428	169	0,61
2016	516	197	0,62

Tabela 2 Índice de reprovação em cálculo I na UFV - MAT 146

Ano	Número de estudantes matriculados	Número de estudantes aprovados	Índice de reprovação
2011	661	328	0,50
2012*	550	329	0,40
2013	658	313	0,52
2014	607	232	0,62
2015	368	100	0,73
2016	476	177	0,63

Lopes (1999) e Wisland; Freitas & Ishida (2014), apresentam estudos acerca dos crescentes índices de reprovação de cálculo em diferentes cenários e propõem soluções, como a criação de um curso de nivelamento, cujo objetivo seria o de revisar os principais conteúdos de matemática básica que se fazem necessários nos cursos de Cálculo, além de uma revisão acerca do número de estudantes matriculados por turma e um aumento no número de vagas em turmas de tutorias o que seriam possíveis soluções paliativas.

Medidas como estas já foram tomadas na UFV, cujos docentes também criaram turmas especiais voltadas para estudantes propensos à reprovação, denominadas turmas de estudo dirigido.

As turmas de estudo dirigido contam com uma maior carga horária semanal e com um número reduzido de estudantes. Nelas, o professor tem um tempo maior para se dedicar à resolução de exercícios e o aluno é, a todo momento, incentivado a estabelecer conexões entre os conceitos trabalhados e a desenvolver métodos para solucionar os problemas propostos.

Para selecionar os estudantes da disciplina em questão a serem matriculados nas turmas de estudo dirigido, tem-se aplicado desde o primeiro semestre do ano de 2016, no início de cada período letivo, um teste de matemática básica, composto por 50 itens, com conteúdos de matemática básica necessários ao desenvolvimento de disciplinas de nível superior tais como cálculo I, cada um com 5 alternativas de resposta. Estes testes têm a finalidade de estimar o nível de conhecimento em matemática do estudante de cálculo antes de cursar a disciplina e utilizar algumas das informações obtidas nesse processo para identificar os estudantes que necessitam de um maior acompanhamento no decorrer da mesma, além de servir como um instrumento de apoio aos professores no desenvolvimento dos seus planos de ensino.

A formulação dos itens destes testes tem sido de responsabilidade dos professores do departamento de matemática da UFV que pretendem desenvolver, futuramente, um banco de itens para que a formulação destes testes ocorra de maneira sistemática nos próximos anos, mas para isto é necessário que se faça uma análise aprofundada de cada um dos itens já aplicados, visando aproveitar o máximo de informação possível que eles possam oferecer.

Seguindo esta vertente, o objetivo geral deste trabalho é oferecer uma análise de um destes testes com base em duas teorias de avaliação: a teoria clássica dos testes (TCT) e a teoria de resposta ao item (TRI) e deste modo, avaliar a sua qualidade métrica, considerando uma análise psicométrica dos itens que o compõe. Dessa forma, os objetivos específicos estão na análise do nível de dificuldade e o poder discriminativo de cada item, além da precisão do instrumento de medida e da informação que eles fornecem para o cálculo das habilidades dos estudantes em matemática, visando também contribuir para a difusão tanto da TCT, quanto da TRI, como alternativas aos atuais métodos utilizados na análise destes testes. Estudos como este, cujo foco está na análise psicométrica de avaliações, como o exame nacional de desempenho de estudantes (ENADE) podem

ser encontrados em Lopes & Vendramini (2015), Oliveira (2017) e Primi; Hutz & Silva (2011), por exemplo.

Com a realização destas análises, pretende-se fornecer aos coordenadores da disciplina de cálculo da UFV informações referentes ao teste, bem como as rotinas utilizadas no software estatístico R (R, 2017) para que esta mesma análise possa ser repetida para os demais testes. A utilização deste software nas análises se justifica pelo fato de ser livre e de código aberto de fácil manipulação, fazendo com que seus comandos possam ser replicados por qualquer leigo em estatística interessado em suas potencialidades.

Dessa forma, a seção 2 apresenta o referencial teórico para a utilização dessas teorias, trazendo um breve histórico sobre testes e uma discussão acerca dos parâmetros psicométricos dos itens e suas funcionalidades. A seção 3 apresenta os materiais e métodos utilizados, destacando o uso do software estatístico livre R (R, 2017) na obtenção dos resultados, discutidos na seção 4, e a seção 5 apresenta algumas conclusões acerca do trabalho realizado.

2 REFERENCIAL TEÓRICO

A teoria clássica dos testes (TCT) e a teoria de resposta ao item (TRI) são provenientes da Psicometria, área que relaciona à Psicologia e a Estatística, cuja proposta está na elaboração de instrumentos de medidas para o conhecimento e o comportamento humano.

Enquanto a TCT investiga as propriedades do conjunto de itens que constituem um teste, a TRI se preocupa em investigar individualmente as propriedades de cada um deles. Cada uma dessas teorias apresenta vantagens e desvantagens em relação a outra, e juntas oferecem um leque de possibilidades de análises que podem ser realizadas em testes avaliativos.

Na seção 2.1 é apresentado um referencial teórico dos principais tópicos que cercam a teoria clássica dos testes enquanto que, na seção 2.2, é apresentado um referencial teórico acerca dos principais tópicos que cercam a teoria de resposta ao item.

2.1 Teoria clássica dos testes (TCT)

Segundo Pasquali (2003), no século XIX, nas pesquisas clínicas, iniciou-se uma revolução no desenvolvimento de instrumentos de avaliação psicológica. Tal revolução, teve início por volta de 1880, com o inglês Francis Galton (1822-1911), que fundou o primeiro laboratório voltado às medições antropométricas em 1884, em Londres, mas também teve contribuições significativas do psicólogo norte-americano James Mckeen Cattell (1860-1944) e do francês Alfred Binet (1857-1911), pioneiro no teste de inteligência, a base dos atuais testes de quociente intelectual (QI). Nessa época, desenvolveu-se a teoria clássica dos testes, a partir dos trabalhos do psicólogo inglês Charles Spearman relacionados à análise fatorial e ao coeficiente de correlação.

O modelo da teoria clássica dos testes, considera que o escore de um teste, que é a quantidade de itens respondidos corretamente pelo respondente, é de natureza falha e, por conseguinte, contém uma parcela de erro. Segundo Primi (2012) temos:

$$X_i = V_i + \epsilon_i$$

em que X_i é o escore observado do i -ésimo indivíduo, V_i é o escore verdadeiro do i -ésimo indivíduo e ϵ_i é o erro aleatório.

É importante salientar que, segundo Fletcher (2010), o erro aleatório ocorre simplesmente ao acaso, é independente dos escores verdadeiros e sem tendenciosidade. Na teoria, ele surge da impossibilidade de termos infinitos itens em um teste, ou do fato de

não conseguirmos aplicar um teste infinitas vezes a um mesmo indivíduo ou de não conseguirmos submeter as respostas a infinitos examinadores.

O escore de erro (ϵ_i) pode ser visto como uma variável aleatória associada a eventuais erros relacionados às condições particulares de aplicação. Ele assume valores positivos e negativos, fazendo, portanto, com que os escores observados sejam ora maiores e ora menores do que os escores verdadeiros. Assume-se que o erro seja assistemático, aleatório, ou seja, não mostra tendência sistemática de assumir valores positivos ou negativos. Nota-se que o valor do escore verdadeiro é fixo entre diferentes aplicações, enquanto o erro tende a variar (PRIMI, 2012).

O escore verdadeiro de um indivíduo poderia ser calculado tomando-se a média dos escores obtidos em infinitas repetições de testes semelhantes. Neste caso, a média do erro aleatório, seria nula, o que faria com que o escore observado médio \bar{X}_M fosse igual ao escore real médio, V_M . Como não é possível tal artifício, segundo Bussab & Morettin (2004), considera-se que quanto maior o número de repetições do teste, mais próxima fica a média dos escores obtidos do verdadeiro escore do respondente, conforme nos diz a lei dos grandes números.

2.1.1 Postulados da teoria clássica dos testes

De acordo com Pasquali (2003), se tivermos um número suficientemente grande de testes, do fato de os erros serem assistemáticos temos que a média dos erros é zero ($E[\epsilon] = 0$) e espera-se que não exista correlação entre os escores verdadeiros e os escores de erro ($\rho_{V\epsilon} = 0$), visto que indivíduos com um bom rendimento tendem a errar tanto positivamente quanto negativamente da mesma maneira. Além disso, supõe-se que não há correlação entre os escores de erro de dois testes paralelos a e b ($\rho_{\epsilon_a\epsilon_b} = 0$).

Segundo Pasquali (2003), várias relações para um conjunto muito grande de escores observados podem ser deduzidas:

- i) A média dos escores observados é igual à média dos escores verdadeiros.

Demonstração. De fato,

$$E[X] = E[V] + E[\epsilon] = E[V], \text{ visto que } E[\epsilon] = 0. \quad \blacksquare$$

- ii) A variância dos escores observados é igual a soma das variâncias dos escores verdadeiros e dos erros.

Demonstração. De fato,

$$\begin{aligned}
\sigma_X^2 &= \sigma_V^2 + \sigma_\epsilon^2 + 2\rho_{V\epsilon}\sigma_V\sigma_\epsilon = \sigma_V^2 + \sigma_\epsilon^2 + 2(0)\sigma_V\sigma_\epsilon \\
&= \sigma_V^2 + \sigma_\epsilon^2.
\end{aligned} \tag{1} \quad \blacksquare$$

iii) A covariância entre o escore observado e o escore verdadeiro é igual à variância do escore verdadeiro.

Demonstração. De fato,

$$\begin{aligned}
\text{cov}[V, X] &= E[(V - E[V])(X - E[X])] \\
&= E[(V - E[V])((V + \epsilon) - E[V])] \\
&= E[(V - E[V])(V - E[V] + \epsilon)] \\
&= E[V^2 - VE[V] + V\epsilon - E[V]V + (E[V])^2 - E[V]\epsilon] \\
&= E[(V^2 - 2VE[V] + (E[V])^2)] + E[V\epsilon] - E[V]E[\epsilon] \\
&= E[(V^2 - 2VE[V] + (E[V])^2)] + 0 - 0 = E[(V - E[V])^2] \\
&= \sigma_V^2.
\end{aligned} \quad \blacksquare$$

iv) A correlação entre o escore verdadeiro e o escore observado é a razão entre os desvios-padrão do escore verdadeiro e os desvios-padrão do escore observado.

Demonstração. De fato,

$$\begin{aligned}
\rho_{VX} &= \frac{\text{cov}[V, X]}{\sigma_V\sigma_X} = \sigma_V^2 \cdot \frac{1}{\sigma_V\sigma_X} \\
&= \frac{\sigma_V}{\sigma_X}.
\end{aligned} \quad \blacksquare$$

Em resumo, os postulados da TCT assumem que os erros aleatórios da medida devem ser independentes, em qualquer circunstância (COUTO & PRIMI, 2011).

2.1.2 Precisão de um teste

A precisão de um teste está associada ao erro de medida, ou seja, a diferença entre os escores observado e verdadeiro em um teste e, de acordo com Primi (2012), é definida como sendo:

Definição 1. O índice de precisão de um teste (em inglês: *index of reliability* - ρ_{VV}) é a razão entre as variâncias do escore verdadeiro e observado do teste. Assim,

$$\rho_{VV} = \frac{\sigma_V^2}{\sigma_X^2} = \left(\frac{\sigma_V}{\sigma_X}\right)^2 = (\rho_{VX})^2. \tag{2}$$

Quando a diferença entre as variâncias dos escores observado e real é relativamente baixa, a confiabilidade do teste será alta (por exemplo, $50/60 = 0,83$). Caso contrário, ou seja, a variância dos escores reais é alta em relação à variância dos escores observados, a confiabilidade do teste será baixa (por exemplo, $10/60 = 0,167$).

Outra maneira de expressar o índice de precisão de um teste é através de uma substituição do resultado obtido em (1) na fórmula (2) acima. Dessa forma,

$$\rho_{VV} = \frac{\sigma_V^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_\epsilon^2}{\sigma_X^2} = 1 - \frac{\sigma_\epsilon^2}{\sigma_X^2}.$$

Assim, podemos observar que o índice de precisão de um teste assume valores entre zero e um. O seu valor indica qual parcela da variância dos escores observados é a variância verdadeira. Quanto menor for o coeficiente, menor será a parcela verdadeira e maior a parcela de erro de medida (FLETCHER, 2010).

Caso conheça-se o índice de precisão e a variância dos escores observados, pode-se determinar a variância dos escores verdadeiros da seguinte forma:

$$\sigma_V^2 = \sigma_X^2 \times \rho_{VV}.$$

Por outro lado, conhecendo-se as variâncias dos escores verdadeiros e observados, pode-se calcular o índice de precisão. Contudo, como o escore verdadeiro não é observável diretamente, sua variância é desconhecida. Um método de estimação desse índice decorre do conceito de formas paralelas de um teste, que equivalem à medidas idênticas, independentes, de uma mesma variável psicológica (PRIMI, 2012). Dessa forma, dois testes com itens distintos são ditos paralelos se possuem uma mesma distribuição dos erros e tiverem os mesmos escores verdadeiros para o mesmo indivíduo em cada teste, o que é uma tarefa com um alto grau de dificuldade.

Para o caso em que temos dois testes paralelos, de acordo com Fletcher (2010), define-se o índice de precisão conforme na definição 2 a seguir.

Definição 2. Como os desvios-padrão σ_{t_1} e σ_{t_2} são iguais para duas formas paralelas t_1 e t_2 de um teste ($\sigma_{t_1} = \sigma_{t_2} = \sigma_X$), a correlação entre elas é dada por:

$$r_{t_1 t_2} = \frac{\sigma_V^2}{\sigma_{t_1} \sigma_{t_2}} = \frac{\sigma_V^2}{\sigma_X^2} = (\rho_{XV})^2 = \rho_{VV}.$$

Nota-se, então, que a correlação entre medidas paralelas é igual ao índice de precisão, podendo ser interpretado como um índice da eficácia de um instrumento de medidas,

revelando até onde os resultados obtidos com determinado método de mensuração são replicáveis.

A precisão de um teste pode ser usada para estimar o erro padrão dos escores e para estabelecer intervalos de confiança em torno dos valores observados. Caso queiramos produzir testes mais precisos, basta aumentarmos o número de itens de nosso teste. A fórmula de Spearman-Brown, a ser definida na seção 2.1.2.3 utiliza o atual índice de precisão para estimar o número de itens que seria necessário para alcançar um novo índice de precisão (FLETCHER, 2010).

Uma estimativa usual da precisão é dada pelo estimador alpha de Cronbach, a ser visto na seção 2.1.2.1, que no caso em que todos os itens são de múltipla escolha, reduz-se a fórmula 20 de Kuder-Richardson (KR_{20}), a qual será vista na seção 2.1.2.2.

2.1.2.1 Coeficiente alfa de Cronbach

O coeficiente alfa de Cronbach, que recebe este nome por ter sido introduzido por Lee J. Cronbach, em 1951, mede a correlação entre respostas em um teste através da análise das respostas dadas pelos respondentes, apresentando uma correlação média entre as perguntas. Dessa forma, ele pode ser utilizado para medir a consistência interna do instrumento de medida. De acordo com Leontitsis & Pagge (2007), o coeficiente alfa de Cronbach é definido por:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_T^2} \right),$$

em que n é o número total de itens, σ_i^2 é a variância relacionada à cada item do teste e σ_T^2 é a variância do escore total dos respondentes.

Observe que, caso haja consistência nas respostas quantificadas, então σ_T^2 será relativamente grande, fazendo com que α tenda a 1. Em contrapartida, caso as respostas sejam aleatórias, σ_T^2 será comparável a soma das variâncias individuais (σ_i^2), fazendo com que α tenda a 0 (ALMEIDA; SANTOS & COSTA, 2010).

Esse coeficiente varia de zero a um, sendo o teste mais consistente a medida que se aproxima de um. De acordo com Hair Júnior et. al (2010), valores acima de 0,7 são considerados satisfatórios.

2.1.2.2 Fórmula vinte de Kuder-Richardson

Segundo Pasquali (2003) a fórmula 20 de Kuder-Richardson (KR_{20}) é utilizada quando os itens são dicotômicos. Neste caso, o estimador alpha de Cronbach se reduz a:

$$KR_{20} = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)}{\sigma_T^2} \right),$$

sendo n e σ_T^2 definidos como anteriormente e $\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)$ o somatório do produto da proporção de pessoas que acertam o item i (\hat{p}_i) e da proporção dos que erraram ($1-\hat{p}_i$).

Caso os itens, além de dicotômicos, tenham o mesmo nível de dificuldade, podemos utilizar a fórmula 21 Kuder-Richardson KR_{21} , a saber:

$$KR_{21} = \frac{n}{n-1} \left(1 - \frac{\bar{T} - \frac{\bar{T}^2}{n}}{\sigma_T^2} \right),$$

sendo n e σ_T^2 como definidos anteriormente e \bar{T} a média da soma dos itens acertados pelo indivíduo (média do escore total).

Observe que $\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i) = n\hat{p}(1-\hat{p}) = n\hat{p} - n\hat{p}^2 = n\hat{p} - \frac{n\hat{p}^2n}{n} = \bar{T} - \frac{\bar{T}^2}{n}$, sendo $\bar{T} = n\hat{p}$.

2.1.2.3 Fórmula de predição de Spearman-Brown

Segundo Sternick (1976), a fórmula de predição de Spearman-Brown permite calcular a precisão de um teste quando ele é aumentado ou diminuído um certo número de vezes. Sua expressão geral é:

$$\rho_{nn} = \frac{n\rho_{XX}}{1 + (n-1)\rho_{XX}}, \quad (3)$$

em que,

n é o número de vezes em que o teste é aumentado;

ρ_{nn} é a precisão para o teste aumentado n vezes;

ρ_{XX} é a precisão do teste.

Observe que com o aumento do teste obtemos uma maior precisão. Se isolarmos n na fórmula (3) acima, podemos calcular o tamanho que o teste deverá ter para possuir determinado índice de precisão. Neste caso,

$$n = \frac{\rho_{nn}(1 - \rho_{XX})}{\rho_{XX}(1 - \rho_{nn})}. \quad (4)$$

Assim, se temos um teste em que a precisão (ρ_{XX}) é 0,50 e queremos aumentá-la para (ρ_{nn}) igual a 0,75, aplicando na fórmula (4), obteremos o número de vezes em que o teste deve ser ampliado:

$$n = \frac{0,75(1 - 0,50)}{0,50(1 - 0,75)} = \frac{0,375}{0,125} = 3.$$

Por estar associada ao erro da medida, tem-se que o aumento no índice de precisão faz com que a medida seja menos errônea, ou seja, quanto mais preciso o teste, mais próximo o modelo estimado pela TCT se aproxima do modelo real.

2.1.3 Análise de itens pela teoria clássica dos testes

A análise clássica dos itens de um teste faz uso de técnicas estatísticas com o intuito de avaliar diversos aspectos dos itens. Existem várias informações que podem ser usadas para determinar se um item é útil como instrumento do que se propõe a medir e sobre como ele funciona em relação aos outros itens de um teste. Nesta seção definiremos algumas destas informações.

2.1.3.1 Estatísticas descritivas

É comum começar uma análise de um conjunto de dados pelas estatísticas descritivas. O mesmo ocorre com os itens de um teste. Cálculos simples como o de médias e desvios-padrão podem fornecer informações sobre a qualidade dos mesmos. Por exemplo, se um item possuir desvio-padrão baixo significa que há pouca variabilidade e isto pode ser um indício da inutilidade do item, pois a capacidade de discriminar os respondentes de um teste é um dos principais objetivos de um teste bem elaborado.

Também podemos realizar uma análise descritiva com base na média de acerto de um determinado item. Caso esta média seja alta, o item pode estar negativamente distorcido e, conseqüentemente, não fornecer o tipo de informação necessária. Dessa forma pode-se observar que, embora não sejam comuns essas análises descritivas dos itens de um teste, essa pode se fazer essencial numa primeira leitura de avaliação. Geralmente, quanto maior a variabilidade do item e quanto mais a distribuição da média de acertos do item for simétrica, melhor será o item (KLINE, 2005).

Enquanto a média p de um item dicotômico é obtida através da proporção de indivíduos que o acertaram, a variância do mesmo pode ser encontrada através do produto $p \times (1 - p)$. Assim, o desvio-padrão pode ser obtido extraíndo a raiz quadrada deste produto. Como exemplo, podemos considerar que se dos 100 respondentes de um item, 75 o acertaram, a média de acertos do item será 0,75 e o desvio-padrão $\sqrt{0,75 \times 0,25} \approx 0,433$. Essa média também é definida como o índice de dificuldade do item, a ser apresentado na seção 2.1.3.2.

Outras análises descritivas como amplitudes, percentis, quartis e decis, e medidas de distribuição (achatamento e assimetria de distribuição), bem como análises gráficas (histogramas, blox-plots, etc.) também podem ser realizadas como instrumentos na avaliação de itens. Estas medidas podem ser encontradas em Bussab & Morettin (2004).

2.1.3.2 Índice de dificuldade

Conforme Borgatto & Andrade (2012), a dificuldade D_i de um determinado item i é definida em termos da proporção \hat{p} de respondentes que acertam tal item. Matematicamente,

$$D_i = \frac{C_i}{N_i}$$

em que, D_i representa o nível de dificuldade do item i , C_i representa o número de indivíduos que responderam corretamente ao item i e N_i representa o número total de indivíduos que foram submetidos ao item i . Por exemplo, se um item i for respondido corretamente por 60% dos indivíduos e um item j , por 40%, afirma-se que o item i é mais fácil do que o item j , pois apresenta uma maior proporção de acertos.

Pasquali (2003) sugere que a dificuldade de um item pode ser afetada pelo chute aleatório certo do respondente e apresenta uma correção para controlar este tipo de resposta, definida como a razão entre o número de indivíduos que erraram o item (E_i) pelo número de alternativas de respostas ao item (K_i) menos um, ou seja,

$$D_i = \frac{C_i - \frac{E_i}{K_i - 1}}{N_i},$$

sendo D_i , C_i e N_i definidos como anteriormente.

Segundo Condé (2001), considera-se o item fácil quando o seu índice de dificuldade for superior a 0,70, moderado quando estiver entre 0,30 e 0,70, inclusive, e difícil quando for inferior ou igual a 0,30.

É importante ressaltar que itens com índice de dificuldade nulo ou igual a um

devem ser removidos da análise do teste, visto que esses não colaboram com qualquer informação para a escala.

Segundo Pasquali (2003), itens com proporção de acerto igual a 0,5 fornecem os maiores níveis de diferenciação entre indivíduos de um grupo. Como exemplo, podemos observar que se houver 100 indivíduos respondendo um item de nível de dificuldade igual a 0,5, ou seja, em que 50 dos 100 indivíduos o responderão corretamente, em média, então haverão $50 \times 50 = 2500$ diferenciações feitas por este item. Já um item com índice 0,3 terá “apenas” $30 \times 70 = 2100$ diferenciações entre os 100 respondentes. Dessa forma, quanto mais próximo o índice de dificuldade estiver de 0,5, mais útil será o item para discriminar (entende-se por diferenciar) os respondentes do teste.

Observe, porém que, ainda segundo Pasquali (2003), caso haja uma grande correlação entre os itens, essa conclusão será falsa. De fato, caso os itens sejam homogêneos, estes não trarão a informação que se busca em termos de discriminação, visto que ao acertar um item, os indivíduos tendem a acertar os demais.

2.1.3.3 Discriminação do item

A discriminação do item avalia a capacidade de diferenciar indivíduos com bom desempenho daqueles indivíduos com baixo rendimento, no mesmo teste. Para isto, podemos analisar estatisticamente este índice através da criação de grupos-critério, de uma análise gráfica ou da correlação de cada item com o escore total observado.

a) Índice de discriminação

Para determinarmos o índice de discriminação (ID_i) separamos os indivíduos respondentes ao teste em dois grupos de igual tamanho, formado pelos indivíduos com maior pontuação (grupo superior) e pelos indivíduos com menor pontuação (grupo inferior). Kelley (1939) sugere que a distribuição dos indivíduos deve ser tratada como normal e que a porcentagem referente aos grupos superior e inferior deve ser de 27%, conforme indicado na Figura 1, em que l_i é a nota máxima obtida para estar entre os 27% alunos com piores notas e l_s é a nota mínima para estar entre os 27% melhores alunos.

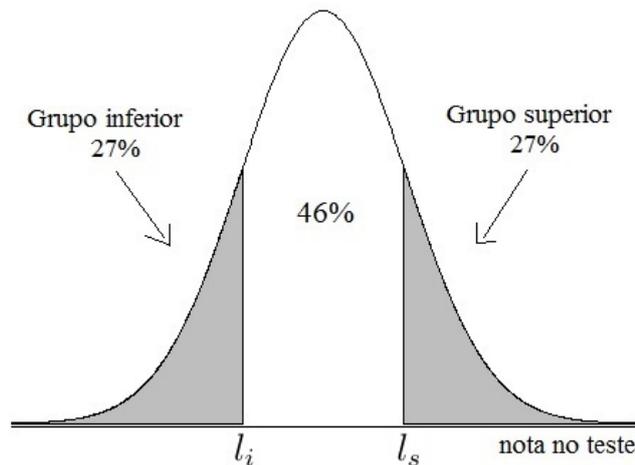


Figura 1 Representação gráfica do método de Kelley.

Pasquali (2003) define o índice de discriminação (ID_i) de um item i como a diferença entre o índice de dificuldade calculado para o grupo superior e inferior.

A Tabela 3 mostra um exemplo para um conjunto de quatro itens.

Tabela 3 Exemplos de índices de discriminação

Item	Índice de dificuldade calculado para o grupo superior	Índice de dificuldade calculado para o grupo inferior	ID_i
1	0,20	0,60	-0,40
2	0,60	0,50	0,10
3	0,75	0,20	0,55
4	0,90	0,05	0,85

Segundo Arias; Lloreda & Lloreda (2006), uma boa referência para a classificação da qualidade discriminativa de um item é a descrita na Tabela 4.

Tabela 4 Classificação do item do teste, de acordo com o índice de discriminação

Índice de discriminação	Classificação do item
$ID \leq 0,20$	Ineficiente. Sugere-se eliminá-lo ou revisá-lo totalmente
$0,20 < ID \leq 0,30$	Necessita ser revisado
$0,30 < ID \leq 0,40$	Aceitável, não sendo necessária uma revisão.
$ID > 0,40$	Satisfatório. Deve permanecer no teste.

Com base na Tabela 4, podemos perceber que os itens 3 e 4 da Tabela 3 apresentam boa discriminação, enquanto que os itens 1 e 2 devem ser eliminados de um teste em

que se procura discriminar grupos de respondentes. É interessante observar que, para o item 1, o valor ID calculado foi negativo. Isso indicaria que aqueles que obtiveram melhor resultado no teste global eram mais suscetíveis a errarem o item do que aqueles que obtiveram um pior resultado.

Também pode-se, através da aplicação do teste t de Student para duas médias de amostras independentes (BUSSAB & MORETTIN, 2004), averiguar se um item é considerado não discriminativo ou inadequado, ao nível de significância α . Isto ocorrerá caso o valor da estatística de teste estiver na região de não rejeição da hipótese nula, indicando que as médias dos grupos inferior e superior são estatisticamente iguais. Lembrando que, para a aplicação do teste t , as pressuposições de normalidade e homocedasticidade devem ser previamente verificadas. Caso contrário, ainda segundo Pasquali (2003), é recomendável a escolha de um teste não paramétrico, como o teste de Wilcoxon ou teste de Mann-Whitney.

b) Curvas baseadas na proporção de acertos

Outra maneira de medir o poder de discriminação de um item é através das curvas baseadas na proporção de acerto de diferentes grupos de respondentes, separados através dos percentis. Observe que o 10º percentil representa os 10% dos estudantes com menores notas em relação aos demais e assim sucessivamente. Este método de análise foi desenvolvido por Rasch (1960), que traçou os escores totais de um teste em função das taxas de aprovação em itens cognitivos. A Figura 2 representa um exemplo desta análise para quatro itens de um determinado teste.

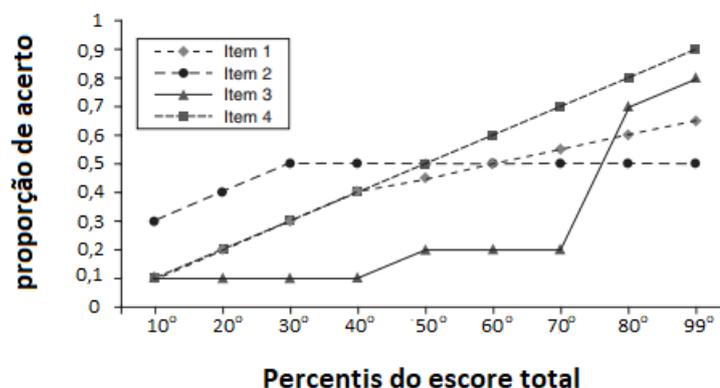


Figura 2 Curvas para itens baseadas na proporção de acertos.

Fonte: Kline (2005, p. 99).

Observe que o item 4 apresenta um aumento retilíneo da porcentagem de acerto

a medida que o desempenho dos respondentes aumenta, isto é, quando eles se encaixam em percentis superiores. Neste caso, a taxa de aumento parece ser constante. O item 1 também apresenta um aumento contínuo. Entretanto, para este item, essa taxa não é uma constante. A partir do 40º percentil a inclinação da reta diminui, indicando que houve uma redução no poder discriminativo deste grupo. O item 2 não consegue discriminar bem os respondentes, a partir do 30º percentil, e o item 3 tem um alto poder discriminativo entre os 70º e 80º percentis.

c) Coeficientes de correlação

Os coeficientes de correlação linear, segundo Pasquali (2003), indicam o grau de associação linear entre duas variáveis. Em outras palavras, tratam-se de medidas utilizadas na verificação da relação linear entre duas variáveis. O coeficiente de correlação mais conhecido e utilizado é o coeficiente de correlação de Pearson, normalmente representado por ρ e assumindo valores de -1 a 1 . O sinal indica a direção, se a correlação é positiva ou negativa. Seu estimador, segundo Bussab & Morettin (2004), é dado por:

$$\hat{\rho} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{S_{XY}}{S_X S_Y}, \quad (5)$$

em que S_{XY} representa a covariância entre X e Y e S_X e S_Y representam os desvios-padrão das variáveis X e Y , respectivamente.

Ainda segundo Pasquali (2003), se desejamos estudar a correlação entre uma variável contínua e uma variável dicotômica, devemos usar o coeficiente de correlação ponto-bisserial (ρ_{pb}). Trata-se de uma estimativa derivada do coeficiente de correlação de Pearson, utilizada na situação na qual as variáveis correlacionadas são contínuas, mas uma delas (o item) foi artificialmente dicotomizada. Sendo assim, consideremos a variável contínua X como a pontuação de um aluno no teste e a variável dicotômica Y como uma indicação sobre o desempenho do mesmo em um item, ou seja, se ele respondeu incorretamente (pontuada com 0) ou corretamente (pontuada com 1) ao item. Observe que a média de Y é a proporção de acerto do item. Assim, com algumas manipulações algébricas podemos reescrever a equação (5) da seguinte maneira:

$$\hat{\rho}_{pb} = \sqrt{\hat{p}\hat{q}} \times \frac{\hat{\mu}_1 - \hat{\mu}_0}{\hat{\sigma}}, \quad (6)$$

em que $\hat{\mu}_1$ e $\hat{\mu}_0$ são as pontuações médias dos alunos que responderam correta e incorretamente ao item, respectivamente, $\hat{\sigma}$ é o desvio padrão do escore total na amostra e \hat{q} é o complemento de \hat{p} , isto é, $\hat{q} = 1 - \hat{p}$.

Embora tenhamos um valor máximo de $\sqrt{\hat{p}\hat{q}}$ para $\hat{p} = \hat{q} = 0,5$, se montássemos um teste apenas com questões com esta proporção de acerto, teríamos uma média grupal de 50%.

Para manter uma média de classe razoável e um alto valor de ρ_{pb} , Leblanca & Coxb (2017) recomendam a seleção de itens em que $0,2 \leq \hat{p} \leq 0,8$. Esses valores resultam em $\sqrt{\hat{p}\hat{q}} \leq 0,4$, o que não está muito longe do valor máximo de 0,5. Obviamente, podemos ter itens com valores de \hat{p} fora deste intervalo e estas podem ser relevantes em alguns contextos. Entretanto, estes itens terão um baixo valor de $\hat{\rho}_{pb}$. Observe que, caso $\hat{p} = 0$ ou $\hat{p} = 1$ teríamos $\hat{\rho}_{pb} = 0$.

A segunda parte da equação (6), $\frac{\hat{\mu}_1 - \hat{\mu}_0}{\hat{\sigma}}$, representa a distância normalizada entre o escore médio dos alunos que acertaram o item e o escore médio dos que erraram. Esta parte irá indicar se o $\hat{\rho}_{pb}$ é positivo ou negativo. Caso seja negativo, estaríamos concluindo que alunos com menores escores são mais propícios a acertar determinado item do que alunos com escores mais alto, o que pode ser um indício de algum problema com a questão, como mal-formulação ou mesmo algum tipo de erro nos dados. Também podemos observar que quanto maior a diferença entre esses grupos, ou seja, quanto mais discriminativo for o item, maior será o $\hat{\rho}_{pb}$.

De acordo com Tôrres (2015), de maneira geral, espera-se que o coeficiente de correlação ponto-bisserial assumam valores positivos e superiores a 0,30 para que sejam considerados de boa discriminação.

Como exemplo prático, consideremos um teste dicotomizado realizado por 10 estudantes e cujos resultados para um determinado item encontram-se na Tabela 5.

Tabela 5 Correlação ponto-bisserial para um teste de dez itens

Respondente	Resposta dicotomizada do item	Escore total do teste
1	1	6
2	0	2
3	0	4
4	1	9
5	0	7
6	0	1
7	1	7
8	1	8
9	0	2
10	0	1

Primeiramente, observemos que a proporção de acertos do item analisado na Tabela 5 é 0,4, ou seja, $\hat{p} = 0,4 \Rightarrow \hat{q} = 0,6$. Observe também que a média dos estudantes que acertaram ao item é 7,5 e que a média dos estudantes que erraram o item é aproximadamente 2,83. O desvio-padrão dos 10 escores apresentados é 2,9. Assim, substituindo na equação (6), temos:

$$\hat{\rho}_{pb} = \sqrt{0,4 \times 0,6} \times \frac{7,5 - 2,83}{2,9} \approx 0,7889. \quad (7)$$

Em nosso exemplo, e na maioria das análises clássicas de testes, a variável dicotômica analisada é naturalmente dicotômica, ou seja, sua classificação já é determinada de forma natural: acerto *versus* erro. Entretanto, existirão casos em que uma variável numérica é artificialmente dicotomizada. Por exemplo, caso queiramos codificar a idade de um grupo de pessoas, podemos fazê-la comparando com um valor específico: se maior que este valor, recebe 1, caso contrário 0.

Nestes casos, segundo Pasquali (2003), utilizamos um outro coeficiente de correlação, também oriundo do coeficiente de correlação de Pearson, denominado coeficiente de correlação bisserial e representado por ρ_b . Como pressuposto, para o cálculo do ρ_b é necessário que a variável contínua a ser dicotomizada siga uma distribuição normal.

Segundo Anjos & Andrade (2012) o coeficiente de correlação bisserial (ρ_b) pode ser determinado em função do coeficiente de correlação ponto-bisserial, através da fór-

mula:

$$\rho_b = \rho_{pb} \frac{\sqrt{p(1-p)}}{\phi(p)},$$

em que, ρ_{pb} é a correlação ponto-bisserial, p é a proporção dos respondentes que responderam corretamente ao item e $\phi(p)$ é o valor da densidade da distribuição normal padrão no ponto em que a área da curva à esquerda deste ponto é igual a p .

Em suma, espera-se de uma resposta a um item discriminativo que os estudantes que saem-se bem na prova como um todo, acertem-no, e por sua vez, aqueles que não vão bem, errem-no. Quanto maior forem os coeficientes de correlação bisserial e ponto-bisserial, maior a capacidade do item de discriminar grupos de indivíduos que construíram determinada competência e habilidade, daqueles que não as construíram. Além disso, os itens com coeficiente de correlação baixo não diferenciam o indivíduo que construiu determinada competência e habilidade, daquele que não a construiu (SANTOS, 2008).

2.1.4 Alguns problemas enfrentados pela TCT

De acordo com Pasquali & Primi (2003), uma das grandes limitações da TCT está no modo em que ela calcula os índices de dificuldade e discriminação, visto que eles dependem da amostra de indivíduos pela qual ela é calculada. Para o índice de dificuldade, por exemplo, um mesmo item pode apresentar um alto valor, caso a amostra seja dominada por indivíduos de baixa habilidade, e um baixo valor, caso a amostra seja dominada por indivíduos de alta habilidade. Dessa forma, um mesmo item pode apresentar diferentes classificações em diferentes cenários. Este problema poderia ser resolvido facilmente se, ao invés de uma amostra, trabalhássemos com toda a população. Contudo, esse trabalho pode ser árduo dependendo do número de respondentes do teste.

Para o índice de discriminação também temos um impasse. Segundo Pasquali (2003), itens muito fáceis ou itens muito difíceis tendem a ter índice de discriminação nulo. Isto ocorre porque, no primeiro caso, todos os respondentes tendem a acertar o item e, no segundo, a errá-lo. Dessa forma, a interpretação da discriminação através da TCT não é confiável para estes tipos de item. Além disso, ambos os parâmetros (dificuldade e discriminação) são calculados com base no escore total do teste. Entretanto, para o cálculo do escore total é necessário que os itens estejam medindo um mesmo constructo, o que ainda não se conclui com as informações obtidas.

Uma saída para este problema está na análise fatorial dos itens para saber se eles

estão medindo um mesmo fator. Segundo Dalmoro; Vieira & Venturini (2008), a análise fatorial aborda o problema de analisar a estrutura das correlações entre um grande número de variáveis (por exemplo, escores de testes, itens de testes, respostas de questionários) definindo um conjunto de dimensões latentes comuns, chamadas de fatores. Dessa forma, é possível identificar as dimensões separadas da estrutura e, então, determinar o grau em que cada variável é explicada por cada dimensão (HAIR JÚNIOR et. al., 2010). Em suma, podemos utilizar a análise fatorial para testar a dimensionalidade de um teste, como pode ser visto na seção 2.2.1.

Embora o desenvolvimento das técnicas da TRI estejam ocorrendo em um ritmo acelerado e seus modelos se apresentarem mais robustos e completos, a TCT ainda tem seu espaço no âmbito das avaliações. A fácil manipulação e compreensão dos seus parâmetros e o vasto leque de opções que eles ainda apresentam faz com que esta teoria ainda seja a mais utilizada por professores e pesquisadores relacionados a educação. Em suma, esta teoria tem se apresentado como a mais democrática.

2.2 Teoria de resposta ao item (TRI)

Segundo Andrade; Tavares & Valle (2000), a Teoria de Resposta ao Item (TRI) envolve um conjunto de modelos matemáticos para mensuração de variáveis latentes, como a inteligência ou o grau de proficiência em determinada área do conhecimento, cuja medição não ocorre de forma direta e cujo objetivo principal está na análise de seus itens separadamente.

Embora tenha surgido na década de 1950, devido à complexidade dos cálculos matemáticos envolvidos na estimação dos parâmetros dos seus modelos, foi apenas a partir da década de 1980, com o desenvolvimento de softwares apropriados, que a TRI começou a ser frequentemente utilizada em avaliações em larga escala.

Atualmente, a aplicação da TRI ocorre em diversas avaliações de habilidades e conhecimentos em todo o mundo, como no *Test of English as a Foreign Language* (Toefl) e no *Programme for International Student Assessment* (PISA). No Brasil, a TRI foi usada pela primeira vez em 1995 na análise dos dados do Sistema Nacional de Ensino Básico (SAEB) e, a partir daí, outras avaliações em larga escala também foram planejadas e implementadas de modo a serem analisadas por meio da TRI, como o Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA) e o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP). Também faz uso dessa teoria o principal exame de acesso às universidades públicas, o Exame Nacional do Ensino Médio (ENEM), o qual a tornou fruto de discussão aberta e coletiva, tanto na mídia, quanto nas escolas e universidades brasileiras.

2.2.1 Pressupostos da TRI

Existem duas pressuposições básicas dos modelos mais correntes da teoria de resposta ao item: a unidimensionalidade e a independência local. Segundo Andrade; Tavares & Valle (2000), um teste é unidimensional se houver apenas uma habilidade responsável pela realização de todos os itens da prova. Obviamente, há uma série de fatores responsáveis pela execução de uma determinada tarefa. Entretanto, o que este postulado propõe é que haja uma habilidade dominante responsável pelo conjunto de itens, a qual está sendo medida pelo teste.

Ainda segundo Andrade; Tavares & Valle (2000), um outro fator que deve ser levado em consideração para a aplicação do modelo é a independência local. Tal pressuposto assume que, para uma dada habilidade, as respostas aos diferentes itens da prova são independentes, ou seja, o desempenho em um item não interfere no desempenho em outro

item. Assim, pode-se considerar que a probabilidade de acerto entre dois itens distintos de um teste se dá pelo produto das probabilidades de acerto dos itens, separadamente.

Em outras palavras, a independência local indica que a probabilidade de resposta a um conjunto de itens, dada a habilidade de um respondente, é igual ao produto das probabilidades das respostas do respondente a cada item separadamente. Matematicamente,

$$P(x_1, x_2, \dots, x_n | \theta) = P(x_1 | \theta) \times P(x_2 | \theta) \times \dots \times P(x_n | \theta) = \prod_{i=1}^n P(x_i | \theta),$$

em que θ é o parâmetro que representa o conjunto de habilidades que afetam os itens em teste, x_i representa a resposta de um indivíduo ao item i , $i \in 1, \dots, n$ e $P(x_i | \theta)$ representa a probabilidade de resposta do indivíduo com habilidade θ ao item i .

Segundo Pasquali & Primi (2003), tais pressuposições não podem ser empiricamente demonstradas e nem possuem bases lógicas; elas são simplesmente aceitas ou não aceitas, isto é, elas são axiomas.

Segundo Hambleton; Swaminathan & Rogers (1991), a unidimensionalidade implica independência local. Além disso, segundo Lord & Novick (1968), a independência local implica também a unidimensionalidade. Assim, tem-se somente uma e não duas suposições a serem verificadas.

Pode-se verificar o pressuposto da unidimensionalidade através de uma análise fatorial para dados dicotomizados, como sugerido por Embretson & Reise (2013). Segundo Brown (2006), a análise fatorial compreende uma série de técnicas multivariadas que se preocupa em verificar se há apenas um fator dominante, responsável por responder por todos os demais.

Tal fator pode ser denominado como um constructo, podendo ser uma variável não observada, escalas, itens, ou uma medida de qualquer espécie. Em suma, ele representa o que se propõe a medir através do teste. Caso haja mais do que um fator, deve-se calcular um escore total para cada fator e as análises estatísticas devem ocorrer independentemente deles. Caso seja único, ele explicaria toda a variância comum observada.

Entretanto, segundo Arias; Lloreda & Lloreda (2006), antes de aplicarmos uma técnica como a de análise fatorial, devemos verificar se a mesma se adequa aos dados analisados. O critério ou teste de Kaiser-Meyer-Olkin (KMO) consiste em verificar o grau de intercorrelações das variáveis e a adequação ao uso da Análise Fatorial. Sua

expressão é dada pela equação:

$$KMO = \frac{\sum_{j \neq k} \sum r_{jk}^2}{\sum_{j \neq k} \sum r_{jk}^2 + \sum_{j \neq k} \sum p_{jk}^2}$$

em que r_{jk} é o coeficiente de correlação ponto-bisserial entre as variáveis X_j e X_k , com $j, k \in 1, \dots, n$ e p_{jk} é o coeficiente de correlação parcial entre X_j e X_k , ou seja, o coeficiente de correlação quando se exclui o efeito de uma terceira variável dados os outros X 's.

Em suma, esta estatística, cujos valores variam entre 0 e 1, indica a proporção da variância dos dados que pode ser considerada comum a todas as variáveis. A classificação quanto à qualidade da análise fatorial, pela estatística KMO, sugerida por Favero et al. (2009), pode ser vista na Tabela 6.

Tabela 6 Estatística Kaiser-Meyer-Olkin (KMO)

Intervalo do teste	Análise fatorial
$KMO \geq 0,9$	muito boa
$0,8 \leq KMO < 0,9$	boa
$0,7 \leq KMO < 0,8$	média
$0,6 \leq KMO < 0,7$	razoável
$0,5 \leq KMO < 0,6$	má
$KMO < 0,5$	inaceitável

Além disso, de acordo com Damásio (2012), a análise fatorial é inapropriada se a matriz de correlação entre as variáveis é uma matriz identidade, ou seja, se as variáveis não estiverem correlacionadas na população.

Segundo Moreira (2007), o teste de esfericidade de Bartlett avalia em que medida essas matrizes são similares, testando a hipótese de que as variáveis não sejam correlacionadas na população. A estatística de teste definida por Bartlett tem distribuição assintótica de χ^2 com $\frac{p(p-1)}{2}$ graus de liberdade, em que p representa o número de variáveis do modelo. Dessa forma, caso haja significância no teste (valor- $p < \alpha$), conclui-se que a matriz de correlações não apresenta, aparentemente, problemas para o procedimento de extração de fatores.

Segundo Hair Júnior et. al (2010), a carga fatorial de um item, calculada para um fator, indica o percentual de covariância existente entre o item e o fator. Espera-se que

cada item obtenha um alto valor para carga fatorial, pois isto indicaria que ele estaria altamente associado com outros itens e que eles representam um mesmo constructo.

A Tabela 7, presente em Hair Júnior et. al (2010, p. 107), indica orientações para identificarmos cargas fatoriais significativas, baseadas no tamanho da amostra analisada. Tabachnick & Fidell (1996) propõem 0,32 como carga mínima para qualificar o item como um representante do construto.

Tabela 7 Carga fatorial: orientações para o tamanho da amostra

Carga fatorial	Tamanho necessário da amostra para significância de 5%
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Também podemos realizar uma análise dos componentes principais para testar a unidimensionalidade dos dados. Trata-se de uma análise da matriz de covariância residual, depois de extraído o primeiro fator. Nesta análise, busca-se comparar os autovalores da primeira e da segunda dimensão, para ver se a segunda tem impacto na explicação da covariância das respostas.

De acordo com Vicini & Souza (2005), a análise de componentes principais e a análise fatorial são técnicas da análise multivariada, aplicadas a um conjunto de variáveis, com o intuito de descobrir as mais relevantes na decomposição de cada fator, que são não correlacionados. Ambas se assemelham por serem usadas para simplificar a estrutura de um conjunto de variáveis.

Modelos multidimensionais, que consideram mais de um traço latente para explicar a probabilidade de resposta ao item, também têm sido alvo de estudos, e podem ser encontrados em Hambleton & Cook (1977) e Hambleton; Swaminathan & Rogers (1991).

Modelos de resposta gradual, ou seja, em que não se consideram a resposta na forma dicotômica podem ser encontrados em Samejima (1974) e modelos em que mais de uma população é analisada podem ser encontrados em Bock; Thissem & Zimowski (1997).

2.2.2 Modelos unidimensionais

Pasquali (2003) define a teoria de resposta ao item como um conjunto de técnicas estatísticas que relacionam a probabilidade de um indivíduo acertar a resposta de um determinado item, com sua proficiência e características do item.

De acordo com Andrade; Tavares & Valle (2000), grande parte dos modelos de teoria de resposta ao item propostos na literatura se referem à natureza do item (se são dicotômicos ou não); da quantidade de populações ou grupos envolvidos e da quantidade de traços latentes a serem mensurados. No caso de uma avaliação psicológica, trataremos esses traços latentes mensuráveis como o nível de conhecimento do respondente.

Os modelos de TRI mais utilizados na prática são os modelos logísticos que, segundo Muñiz (1990), permitem melhor tratamento matemático e também são mais frequentes na literatura especializada. Entre os modelos logísticos unidimensionais da TRI se destacam os modelos de um, dois ou três parâmetros para itens dicotômicos, os quais consideram a dificuldade do item; a dificuldade e a discriminação; a dificuldade, a discriminação e a resposta correta dada ao acaso, respectivamente. O modelo logístico de um parâmetro é também denominado como modelo de Rasch, por ter sido desenvolvido pelo matemático dinamarquês Georg Willian Rasch (1901 - 1980).

Pode-se supor que o modelo logístico de três parâmetros seja melhor por apresentar mais parâmetros que os demais. Entretanto, em alguns casos, este modelo não apresenta diferenças significativas em relação aos demais que acabam se tornando mais vantajosos, devido à menor complexibilidade dos cálculos matemáticos envolvidos.

Esses modelos fornecem uma equação matemática que relaciona a probabilidade $P(\theta)$ do respondente acertar um item com sua respectiva habilidade θ . Dessa forma, $P(\theta)$ produz uma curva, denominada curva característica do item (CCI), cujo aspecto geral é mostrado na Figura 3.

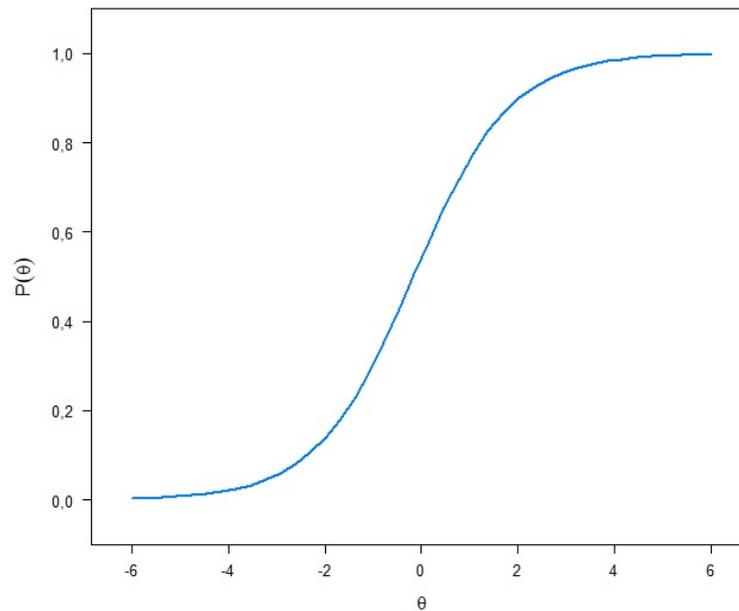


Figura 3 Exemplo de uma curva característica do item.

Se somarmos as CCI's de todos os itens de um teste, obteremos uma curva denominada curva característica do teste (CCT). Esta curva representa a probabilidade de obtenção de um escore total em função de θ .

Os modelos unidimensionais para respostas dicotômicas distinguem-se, basicamente, pela quantidade de parâmetros que consideram. As seções 2.2.2.1, 2.2.2.2 e 2.2.2.3 tratam dos principais modelos unidimensionais utilizados na análise de itens via TRI.

2.2.2.1 Modelo logístico de um parâmetro

Este modelo considera apenas o parâmetro de dificuldade para distinguir os itens de um teste e foi desenvolvido inicialmente por Rasch (1960), sendo apresentado na forma de uma ogiva normal, e adaptado por Wright (1968), que substituiu a função ogiva normal pela função logística (MUÑIZ, 1990).

Formalmente, define-se a dificuldade de um item, neste modelo, como sendo o valor de θ que faz com que $P(\theta)$ seja exatamente 0,5. A Figura 4 ilustra a CCI de dois itens que apresentam diferentes valores de dificuldade:

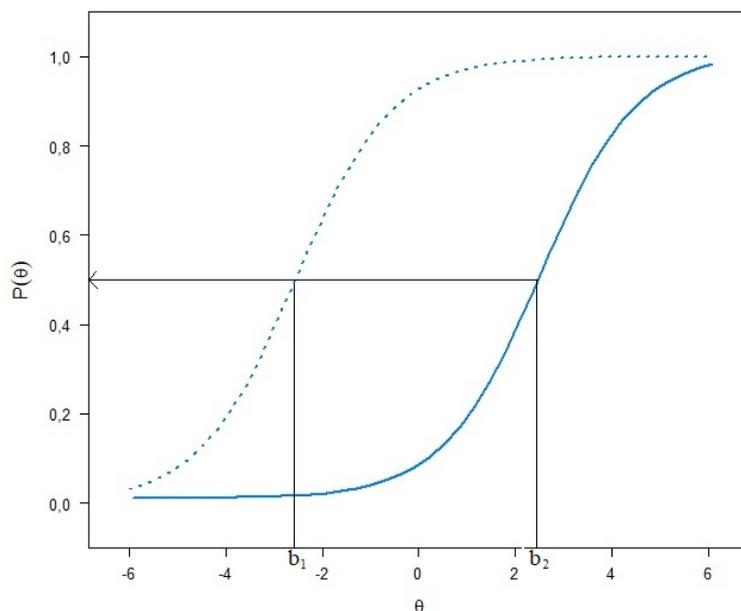


Figura 4 Índice de dificuldade de dois itens de um teste.

Observe que, na Figura 4, a curva tracejada possui índice de dificuldade menor do que a curva contínua, pois é necessário que se tenha uma menor habilidade para se ter 50% de probabilidade de acerto ao item. Segundo Baker (2001), embora o intervalo de variação do índice de dificuldade seja $(-\infty, +\infty)$, é comum encontrarmos valores entre -3 e 3 ($-3 \leq b \leq 3$) para ele.

Segundo Andrade; Tavares & Valle (2000), a expressão matemática utilizada para o modelo de Rasch é:

$$P(U_{ij} = 1|\theta_j, b_i) = \frac{e^{D(\theta_j - b_i)}}{1 + e^{D(\theta_j - b_i)}}, \quad (8)$$

em que $P(U_{ij} = 1|\theta_j, b_i)$ é a probabilidade do indivíduo j com habilidade θ_j acertar o item i , b_i é o parâmetro de dificuldade do item i , medido na mesma escala de habilidade e D é uma constante de aproximação aos valores da curva normal acumulada, que no caso da curva logística vale 1.

De acordo com Pasquali (2003), a constante D foi incluída na fórmula para tornar a curva logística, atualmente utilizada, igual à curva ogiva normal acumulada utilizada nos estudos pioneiros da TRI. Para isto, basta substituir o valor D por 1,7.

Como exemplo prático, suponhamos que em um item previamente testado, obtivemos o valor de $b = 1,1$ para o índice de dificuldade. Encontremos a probabilidade de um indivíduo de habilidade $\theta = 2,5$ acertar a este item.

$$P(U = 1|2, 5; 1, 1) = \frac{e^{2,5-1,1}}{1 + e^{2,5-1,1}} \approx 0,8022. \quad (9)$$

Embora, ao contrário dos demais modelos logísticos, tenha a vantagem de não necessitar de grandes amostras para seu ajuste, o modelo de Rasch supõe que o índice de discriminação seja o mesmo para todos os itens, o que, na prática é muito raro acontecer.

2.2.2.2 Modelo logístico de dois parâmetros

Desenvolvido por Birnbaum (1968), este modelo considera dois parâmetros para a análise de itens: a dificuldade e a discriminação. A expressão matemática utilizada neste modelo é:

$$P(U_{ij} = 1|\theta_j, a_i, b_i) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}, \quad (10)$$

em que $P(U_{ij} = 1|\theta_j, b_i)$ é a probabilidade do indivíduo j com habilidade θ_j acertar o item i , a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da curva característica do item no ponto de abscissa b_i , que representa o parâmetro de dificuldade do item i , medido na mesma escala de habilidade e D é uma constante de aproximação aos valores da curva normal acumulada, que no caso da curva logística vale 1.

Observe que o modelo logístico de um parâmetro é obtido substituindo o valor de a_i por um na equação (10) acima.

Como visto na seção 2.2.2.1, suponhamos que um item previamente testado obteve o valor de $b = 1,1$ para o índice de dificuldade e o valor de $a = 0,6$ para o índice de discriminação. Encontremos a probabilidade de um indivíduo de habilidade $\theta = 2,5$ acertar este item.

$$P(U = 1|\theta = 2, 5; a = 0, 6; b = 1, 1) = \frac{e^{0,6(2,5-1,1)}}{1 + e^{0,6(2,5-1,1)}} \approx 0,6985. \quad (11)$$

Segundo Andrade; Tavares & Valle (2000), a discriminação de um item representa a capacidade que ele tem de diferenciar pessoas com diferentes habilidades, principalmente no ponto de inflexão da CCI, isto é, no ponto em que a probabilidade de acerto ao item é de 50%. Em suma, trata-se do ângulo de incidência da curva no momento da inflexão. Dessa forma, quanto mais inclinada a curva, mais discriminativo o item.

De acordo com Hambleton; Swaminathan & Rogers (1991) o valor deste parâmetro, frequentemente representado pela letra a , normalmente não extrapola o intervalo de 0 e 4 ($0 \leq a \leq 4$).

Um valor negativo para este índice indicaria que a habilidade do respondente e a probabilidade dele acertar o item seriam inversamente proporcionais, ou seja, indivíduos com baixa habilidade acertariam o item com maior frequência do que indivíduos com alta habilidade. Isto poderia ser um indicativo de que o item encontra-se com algum tipo de erro no gabarito ou na formulação, sendo necessária uma revisão. A Figura 5 representa uma curva característica do item que retrata esta situação.

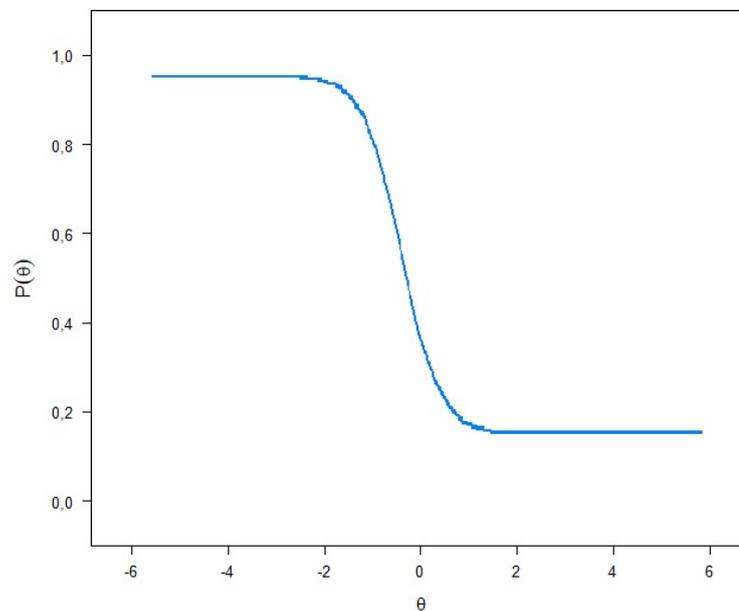


Figura 5 Índice de discriminação negativo.

A Figura 6 representa três itens de mesma dificuldade ($b = 0$) que apresentam índices de discriminação distintos. Observe que o item 1 é o mais discriminativo e o item 3, o menos discriminativo, estando o item 2 com uma discriminação intermediária em relação aos outros.

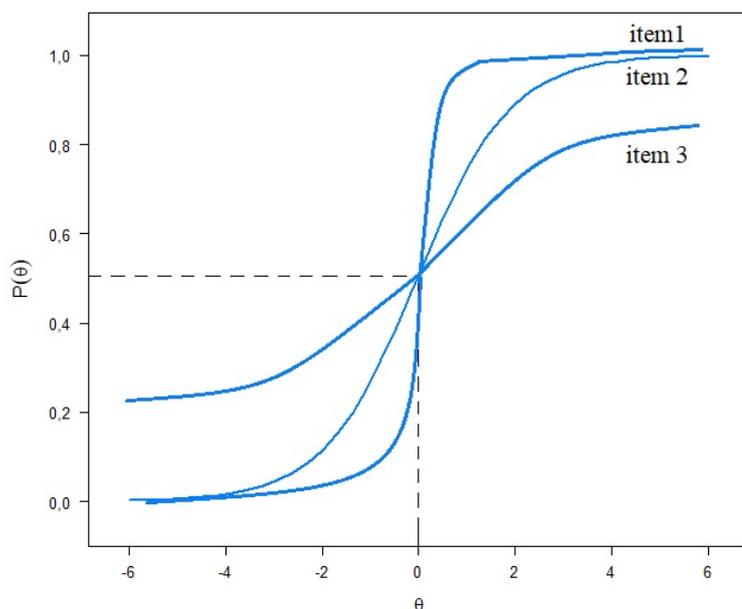


Figura 6 Curvas características dos itens para diferentes valores do coeficiente de discriminação a .

Cabe ressaltar que itens com elevados valores para o índice de discriminação tendem a colaborar de maneira mais significativa para uma maior precisão do instrumento de medida. A Tabela 8 indica uma classificação, proposta por Baker (2001) para o índice de discriminação. Se quisermos uma classificação sobre o modelo ogiva normal, basta dividirmos os valores da Tabela 8 por 1,7.

Tabela 8 Classificação para o índice de discriminação

Classificação	Parâmetro de discriminação
nenhuma	0
muito baixa	0,01 – 0,34
baixa	0,35 – 0,64
moderada	0,65 – 1,34
alta	1,35 – 1,69
muito alta	(1,70; $+\infty$)
perfeita	$+\infty$

A construção de uma tabela de classificação quanto ao nível de dificuldade do item requer alguns cuidados. Classificar um item como fácil, moderado ou difícil, por exemplo, vai de acordo com o grupo em que ele foi examinado. Dessa forma, um mesmo item pode ser considerado fácil para um grupo cuja habilidade média é mais alta do que um outro grupo em que o mesmo item foi considerado difícil.

Para contornar esta situação podemos construir uma tabela de classificação que leva em consideração os quantis da curva logística sob estudo. Por exemplo, caso queiramos classificar um item em cinco grupos (muito fácil, fácil, moderado, difícil e muito difícil), podemos utilizar o 1º, 2º, 3º e 4º quintis correspondem ao 20º, 40º, 60º e 80º percentis, respectivamente.

2.2.2.3 Modelo logístico de três parâmetros

O modelo logístico de três parâmetros mantém todas as características dos modelos de um e dois parâmetros, porém acrescenta a possibilidade de se avaliar a probabilidade do respondente acertar o item ao acaso, ou seja, sem que tenha habilidade suficiente (COUTO & PRIMI, 2011).

Segundo Andrade; Tavares & Valle (2000), sua expressão matemática é dada por:

$$P(U_{ij} = 1|\theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}, \quad (12)$$

em que $P(U_{ij} = 1|\theta_j, a_i, b_i, c_i)$ é a probabilidade do indivíduo j com habilidade θ_j acertar o item i , a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da curva característica do item no ponto de abscissa b_i , que representa o parâmetro de dificuldade do item i , medido na mesma escala de habilidade, c_i é o parâmetro do modelo que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i , também conhecido como probabilidade do chute certo ou probabilidade de acerto casual e D é uma constante de aproximação aos valores da curva normal acumulada, que no caso da curva logística vale 1.

Enquanto que, nos modelos logísticos de um e dois parâmetros, o índice de dificuldade de um item é um ponto na escala de habilidade em que a probabilidade de resposta correta é 0,5, para o modelo logístico de três parâmetros, esta probabilidade deverá ser $\frac{1 + c_i}{2}$, em que c_i representa o parâmetro do item i que avalia a resposta correta dada ao item por acaso e é expresso, de acordo com Pasquali (2003), pela assíntota inferior da curva característica do item.

A fórmula, desenvolvida por Birnbaum (1968), para o modelo logístico de dois parâmetros pode ser encontrada substituindo $c_i = 0$ na fórmula (12), dada acima. Analogamente, para encontrarmos a fórmula do modelo logístico de um parâmetro, criado por Rasch (1960), basta substituímos $c_i = 0$, e $a_i = 1$, na fórmula (12).

Com efeito, tomemos agora um item previamente testado que também obteve os valores de $b = 1,1$ para o índice de dificuldade e $a = 0,6$ para o índice de discriminação,

como o exemplo dado no seção 2.2.2.2 e o valor de $c = 0,2$ para a probabilidade de acerto ao acaso. Encontremos a probabilidade de um indivíduo de habilidade $\theta = 2,5$ acertar este item.

$$P(U = 1|\theta = 2,5; a = 0,6; b = 1,1; c = 0,2) = 0,2 + (1 - 0,2) \cdot \frac{e^{0,6(2,5-1,1)}}{1 + e^{0,6(2,5-1,1)}} \approx 0,7588.$$

A Figura 7 representa uma curva característica do item, em que os parâmetros do item encontram-se destacados.

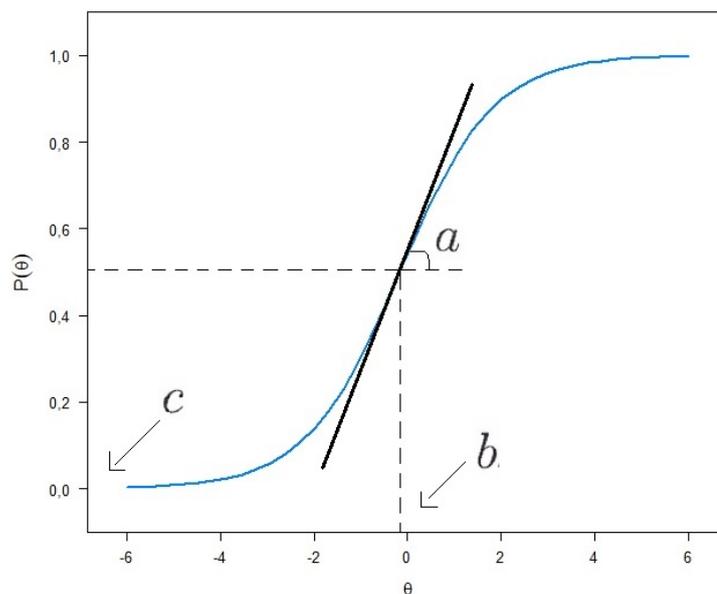


Figura 7 Parâmetros da curva característica do item.

O modelo de quatro parâmetros permite controlar circunstâncias aleatórias relacionadas com falhas do construtor ao elaborar itens para o teste, gerando erros de sujeitos com alta habilidade na resolução desses itens. Entretanto, a inclusão deste quarto parâmetro não parece trazer ganhos significativos em comparação com o modelo de 3 parâmetros (COUTO & PRIMI, 2011), e não será objeto de estudo aqui.

2.2.3 Estimação

Uma das principais etapas da TRI se encontra no ajuste dos modelos, fornecendo informações tanto dos parâmetros dos itens quanto das habilidades dos respondentes que, geralmente, não são previamente conhecidas. O método da Máxima Verossimilhança Marginal (MVM) e o método da média *a posteriori* (EAP), são os mais utilizados para este tipo de estimação. O desenvolvimento desta seção se fez baseado no exposto por Azevedo

(2003), cujas referências se encontram nos trabalhos de Baker (1992) e Andrade; Tavares & Valle (2000).

Vamos considerar, inicialmente, que n indivíduos foram submetidos a um teste de I itens dicotomizados, tal que Y_{ij} representa a seguinte variável aleatória de Bernoulli:

$$Y_{ij} = \begin{cases} 1, & \text{se o indivíduo } j \text{ responde corretamente ao item } i \\ 0, & \text{se o indivíduo } j \text{ responde incorretamente ao item } i, \end{cases}$$

com $j = 1, \dots, n$ e $i = 1, \dots, I$.

Consideremos ainda que a_i, b_i, c_i, θ_j e D são definidas do mesmo modo que nas equações (8), (10) e (12) e as notações:

- $\mathbf{Y}_{.j} = (Y_{1j}, \dots, Y_{Ij})'$, vetor aleatório de respostas do j -ésimo respondente dos I itens;
- $\mathbf{Y}_{..} = (\mathbf{Y}'_{.1}, \dots, \mathbf{Y}'_{.n})'$, o conjunto integral de respostas;
- $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)'$, vetor dos parâmetros do i -ésimo item;
- $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \dots, \boldsymbol{\zeta}'_I)'$ vetor dos parâmetros de todos os itens;
- $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$ vetor de habilidades de todos os indivíduos.

Para o processo de estimação, segundo Azevedo (2003), basicamente considera-se que os n respondentes constituem uma amostra aleatória simples de uma população Π com distribuição de probabilidade latente existente, associada às habilidades. Dessa forma, a densidade $g(\theta)$ pode ser determinada através da realização de um experimento de retirada de um respondente desta população acompanhado da observação de seu traço latente θ .

Considerando-se uma densidade $g(\theta|\boldsymbol{\eta})$ de classe C^2 , ou seja, duplamente diferenciável, cujas componentes de $\boldsymbol{\eta}$ são conhecidas e finitas (vetor de parâmetros associado a $g(\cdot)$) e que θ é normalmente distribuída, temos que $\boldsymbol{\eta} = (\mu, \sigma^2)$, em que μ e σ^2 representam a média e a variância das habilidades dos respondentes pertencentes a Π , respectivamente, temos a probabilidade marginal de um respondente j apresentar um determinado padrão de resposta $\mathbf{y}_{.j}$, dada por:

$$\begin{aligned} P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \boldsymbol{\zeta}, \boldsymbol{\eta}) &\equiv P(\mathbf{Y}_{.j} | \boldsymbol{\zeta}, \boldsymbol{\eta}) = \int_{\mathbb{R}} P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \theta, \boldsymbol{\zeta}, \boldsymbol{\eta}) g(\theta, \boldsymbol{\eta}) d\theta \\ &= \int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\zeta}) g(\theta, \boldsymbol{\eta}) d\theta, \end{aligned}$$

em que nesta última igualdade, tem-se que a distribuição de $\mathbf{Y}_{\cdot j}$ não é função de $\boldsymbol{\eta}$.

Como as respostas de diferentes indivíduos são eventos independentes, a verossimilhança será dada por:

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = P(\mathbf{Y}_{\cdot} | \boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{j=1}^n P(\mathbf{Y}_{\cdot j} | \boldsymbol{\zeta}, \boldsymbol{\eta}),$$

e, assim, a log-verossimilhança será dada por:

$$l(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^n \ln P(\mathbf{Y}_{\cdot j} | \boldsymbol{\zeta}, \boldsymbol{\eta}). \quad (13)$$

Segundo Azevedo (2003), embora a componente $\boldsymbol{\eta}$ tenha sido mantida na log-verossimilhança definida na equação (13), essa componente foi fixada na busca de solucionar o problema da indeterminação. Dessa forma, de acordo com Baker (1992) e Andrade; Tavares & Valle (2000), citados por Azevedo (2003), obtém-se as seguintes equações de estimação:

$$a_i : (1 - c_i) \sum_{j=1}^n \int_{\mathbb{R}} [(y_{ij} - P_i)(\theta - b_i)W_i] g_j^*(\theta) d\theta = 0 \quad (14)$$

$$b_i : -a_i(1 - c_i) \sum_{j=1}^n \int_{\mathbb{R}} [(y_{ij} - P_i)W_i] g_j^*(\theta) d\theta = 0, \quad (15)$$

$$c_i : \sum_{j=1}^n \int_{\mathbb{R}} \left[(y_{ij} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) d\theta = 0 \quad (16)$$

em que,

$$P_i = c_i + (1 - c_i) \{1 + \exp\{-Da_i(\theta - b_i)\}\}^{-1};$$

$$P_i^* = \{1 + \exp\{-Da_i(\theta - b_i)\}\}^{-1};$$

$$Q_i = 1 - P_i;$$

$$Q_i^* = 1 - P_i^*;$$

$$W_i = \frac{P_i^* Q_i^*}{P_i Q_i};$$

$$g_j^*(\theta) = \frac{P(\mathbf{Y}_{\cdot j} | \boldsymbol{\zeta}, \theta) g(\theta | \boldsymbol{\eta})}{\int_{\mathbb{R}} P(\mathbf{Y}_{\cdot j} | \boldsymbol{\zeta}, \theta) g(\theta | \boldsymbol{\eta}) d\theta}.$$

É importante ressaltar que as equações estimadas requerem métodos iterativos, como o de *Newton-Raphson* e “*Scoring*” de *Fisher* para a resolução, que podem ser encontrados em Ruggiero & Lopes (1997).

2.2.4 Critérios para seleção de modelos

Dados dois ou mais modelos ajustados é necessário que se determine o que melhor explica a variável resposta, se ajustando melhor aos dados. Para isto, pode-se utilizar os critérios de informação de Akaike (AIC) e bayesiano (BIC) e o teste da razão de máxima verossimilhanças (TRV).

2.2.4.1 Critérios de informação de Akaike (AIC) e bayesiano (BIC)

O critério de informação de Akaike (AIC), criado por Akaike (1974), e o critério de informação bayesiano (BIC), descrito por Schwarz (1978), são definidos, respectivamente, como:

$$AIC_p = -2 \ln(L_p) + 2[(p + 1) + 1], \quad (17)$$

$$BIC_p = -2 \ln(L_p) + [(p + 1) + 1] \ln n, \quad (18)$$

em que L_p é a função de verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo e n o tamanho da amostra.

Ao compararmos as equações (17) e (18), podemos perceber que o BIC impõe medidas corretivas mais duras que o AIC. Em termos de ajustamento, segundo Gujarati & Porter (2011), o melhor modelo é aquele que apresenta menores valores para ambos critérios de informação.

2.2.4.2 Teste da razão de verossimilhanças

Segundo Gujarati & Porter (2011), podemos também comparar modelos estimados através do teste da razão de verossimilhanças (TRV). Trata-se de um teste de hipóteses que compara a qualidade do ajuste entre dois modelos. Por exemplo, podemos usar o teste da razão de verossimilhanças para comparar a qualidade do ajuste de um modelo logístico de um parâmetro, com um modelo logístico de dois parâmetros. Se o valor- p do TRV for menor do que o nível de significância adotado, concluímos que um dos modelos oferece uma qualidade de ajuste significativamente melhor do que o outro para nossos dados amostrais.

2.2.5 Qualidade do ajuste

Definido e ajustado o modelo, faz-se necessário avaliar a adequação do seu ajuste aos dados empíricos. Segundo Lenhard (2013), o teste qui-quadrado nas suas variações (Bock (1972), Yen (1981) e McKinley & Mills (1985)), é o método mais tradicional para verificar a qualidade do ajuste dos modelos da TRI. Dessa forma, a estatística é obtida comparando as respostas do modelo real com as respostas do modelo esperado, dado que as suposições do modelo já foram previamente atendidas.

Segundo Gomes & Silva (2009), o teste proposto por Bock (1972) é comum para modelos dicotômicos. Sua estatística é dada por:

$$\chi_i^2 = \sum_{g=1}^G \frac{N_g(O_{ig} - E_{ig})^2}{E_{ig}(1 - E_{ig})}$$

em que O_{ig} é a proporção de respostas corretas para o i -ésimo item no intervalo g de proficiências, E_{ig} é a proporção esperada de respostas corretas baseada na curva característica do item ajustada e avaliada na mediana das proficiências estimadas, que se encontram dentro do g -ésimo intervalo, e N_g é o número de indivíduos com proficiência dentro do intervalo g . Esta estatística tem distribuição aproximadamente qui-quadrado com $G - m$ graus de liberdade, em que m é o número de parâmetros estimados para o item e G representa o número de subgrupos para os quais a habilidades são divididas (GOMES & SILVA, 2009).

Através da estatística χ^2 podemos construir a região crítica do teste referente à qualidade do ajuste do modelo. Assim, devemos rejeitar a hipótese H_0 de que o modelo proposto se ajusta bem aos dados, quando a estatística χ_i^2 é maior do que um valor crítico χ_α^2 .

2.2.6 Função de informação

Uma medida bastante utilizada em conjunto com a CCI é a função de informação do item, útil por permitir analisar quanto um item traz de informação para a medida da habilidade do indivíduo, sendo um poderoso método para descrever itens, bem como para selecioná-los.

A função de informação de um item, segundo Andrade; Tavares & Valle (2000) é dada por:

$$I_i(\theta) = \frac{[\frac{d}{d\theta} P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}, \quad (19)$$

em que, $I_i(\theta)$ é a “informação” fornecida pelo item i no nível de habilidade θ , $P_i(\theta) = P(X_{ij} = 1|\theta)$ e $Q_i(\theta) = 1 - P_i(\theta)$.

No caso do modelo logístico unidimensional de 3 parâmetros, segundo Andrade; Tavares & Valle (2000), a equação (19) resulta em:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2. \quad (20)$$

Observe que a informação é maior quando:

- i) o índice de dificuldade b_i se aproxima de θ , visto que, neste caso, $P_i(\theta) = Q_i(\theta) = 0,5$;
- ii) quanto maior for o índice de discriminação a_i ;
- iii) quanto menor for o valor da probabilidade de acerto ao acaso c_i de um item.

No caso do modelo logístico unidimensional de um e dois parâmetros, basta substituímos a_i por um e c_i por zero, respectivamente, na equação (20).

O total de informação do teste, visto a independência entre os itens, é dado pela soma das informações de cada item, calibrados em uma mesma escala. Dessa forma, segundo Anjos & Andrade (2012), a função de informação de um teste com n itens pode ser escrita como:

$$I(\theta) = \sum_{i=1}^n I_i(\theta).$$

Ainda segundo Anjos & Andrade (2012), a informação do teste também pode ser expressa através do erro-padrão de estimação:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

A Figura 8 indica a função de informação para quatro itens. Observe que o item 1 fornece a maior informação quando $\theta = -2$ e o item 4 quando $\theta = 4$, pois estes valores apresentam os pontos de máximo da curva de informação do item. Observe também que o item 2 é mais informativo que o item 3 para qualquer nível de θ . Dessa forma, o item 3 pode ser descartado, visto que ele não produziu nenhuma informação a mais do que a já produzida pelo item 2.

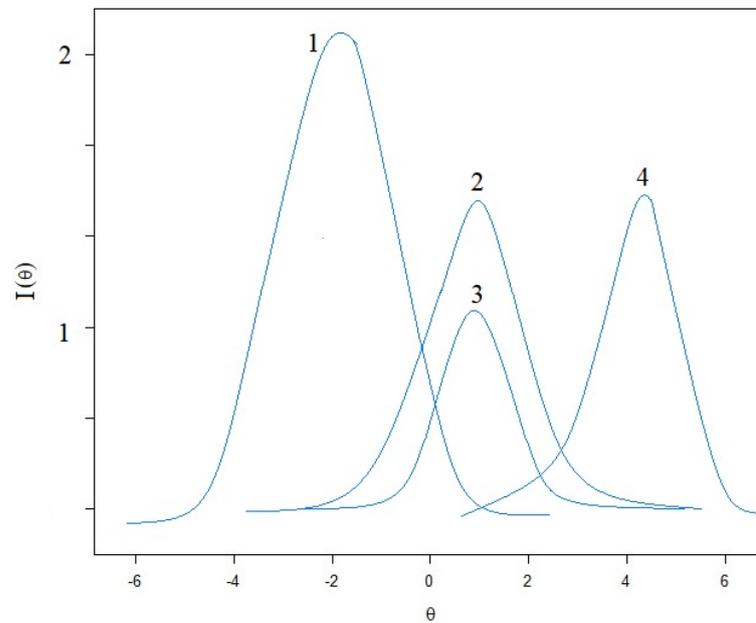


Figura 8 Função de informação para quatro itens.

A principal diferença entre a TCT e a TRI está no fato de que esta última não contabiliza apenas o número total de acertos no teste. Aqui, o item é a unidade básica de análise. O surgimento dessa teoria trouxe um leque de possibilidades para os estudiosos da área de avaliação como a possibilidade de elaboração de provas diferentes para um mesmo exame de seleção, tornando possível a aplicação dessas provas em qualquer período do ano sem trazer prejuízos na comparação dos resultados. Além disso, o fato de não ter um limite inferior ou superior padrão entre as áreas de conhecimento colabora para uma análise mais significativa das proficiências dos respondentes de um teste. Ambas as teorias devem caminhar juntas, uma complementando a outra, para que se possa obter uma avaliação concisa e eficiente.

3 MATERIAL E MÉTODOS

3.1 Material

Os dados sob estudo são provenientes de um teste com conteúdos de matemática básica necessários ao desenvolvimento de disciplinas de nível superior tais como Cálculo I, aplicado no segundo semestre do ano de 2016 a 354 estudantes, matriculados na disciplina.

O teste é composto de 50 itens de matemática básica, cujos tópicos encontram-se na Tabela 9. Uma matriz de referências, elaborada pelos professores da disciplina de cálculo, com os possíveis tópicos que podem ser utilizados na elaboração destes testes encontram-se no Apêndice B. O Apêndice C traz os 50 itens analisados.

Tabela 9 Tópicos que compõem o teste analisado

Itens	Tópico
1, 6 e 7	Teoria dos conjuntos e conjuntos numéricos
2, 3, 4 e 5	Expressões numéricas
8, 9, 10, 11, 12, 13, 14, 15, 16, 17 e 18	Expressões algébricas
20, 21 e 27	Equações e inequações
22, 23, 24, 25, 26, 35, 36, 49 e 50	Funções: introdução, funções do 1º e 2º graus.
19 e 37	Função modular
41, 42, 43 e 44	Função exponencial e logarítmica
28, 38, 39 e 40	Função composta e inversa
29	Transformação de funções
30, 31, 32, 33 e 34	Polinômios e equações polinomiais
45, 46, 47 e 48	Trigonometria

Cada item possui 5 alternativas de resposta, das quais apenas uma era a correta, com exceção do item 18 que não apresentou nenhuma alternativa correta. Além disso, todos os itens apresentavam a opção “*não sei*” como resposta. O objetivo de incluir esta opção entre as alternativas estava na busca do controle ao acerto casual, que poderia prejudicar algumas análises estatísticas, e também para servir como um auxílio aos professores, visto que tópicos que apresentavam um grande índice de resposta para esta alternativa, indicariam que uma revisão sobre o assunto se faria necessária.

3.2 Métodos

Os métodos foram primeiramente desenvolvidos seguindo a linha bibliográfica, através de um referencial teórico que se concentrou nos principais trabalhos relacionados à TCT e à TRI, tanto a nível nacional, com os trabalhos de Pasquali (2003) e Andrade; Tavares & Valle (2000), quanto a nível internacional, com os trabalhos de Baker (2001) e Embretson & Reise (2013).

Na parte prática, como um estudo de caso, pôde-se aplicar os conceitos estudados e a aprendizagem desenvolvida a fim de alcançar os objetivos propostos, cujo centro está na análise de uma avaliação de matemática básica utilizando ambas as teorias. Para isto, selecionou-se a avaliação aplicada no segundo semestre do ano de 2016, que contou com 354 respondentes, número que representa quase o dobro do total de respondentes da edição anterior da avaliação, justificando a seleção desta.

Em seguida, coletou-se os dados (gabaritos) que foram transcritos para uma planilha do Excel, como mostra a Figura 9, visando organizá-los para a leitura e análise por meio do *software* R (R, 2017), utilizado por se tratar de um software livre e de código aberto de fácil manipulação e compreensão. É possível baixar este *software* no site <http://cran.r-project.org/>.

	A	B	C	D	E	F	...	AY
1		Questão 1	Questão 2	Questão 3	Questão 4	Questão 5	...	Questão 50
2	Gabarito	D	D	A	A	B	...	D
3	Respondente 1	A	D	A	A	B	...	D
4	Respondente 2	B	E	A	C	B	...	C
5	Respondente 3	D	D	A	A	B	...	B
6	Respondente 4	D	B	A	A	B	...	C
7	Respondente 5	A	D	E	A	E	...	B
8	Respondente 6	A	C	A	A	E	...	D
9	Respondente 7	C	A	A	A	B	...	B
10	Respondente 8	D	D	A	A	B	...	D
11	Respondente 9	C	D	A	C	B	...	D
12	Respondente 10	A	A	D	A	C	...	B
	...	B	D	C	A	C	...	B
356	Respondente 354	C	A	B	A	B	...	B

Figura 9 Modelo do banco de dados.

Os principais pacotes desse software utilizados no trabalho foram o *ltm* (Rizopoulos (2006)), *mirt* (Chalmers (2012)) e *psych* (Revelle (2014)). Entretanto, outros pacotes como *irtosys* (Partchev (2009)) e *CTT* (Willse & Shu (2014)), também podem ser utilizados na análise de itens.

O pacote *ltm* foi desenvolvido por Rizopoulos (2006) e realiza análises de dados dicotômicos e politômicos usando variáveis latentes sob a abordagem TRI, incluindo os

modelos de Rasch, e logísticos de 2 e 3 parâmetros, além do modelo de resposta gradual e os modelos de crédito parcial generalizado.

O pacote *mirt* foi desenvolvido por Chalmers (2012) e realiza análise de dados de resposta dicotômicos e politômicas usando modelos de traços latentes unidimensionais e multidimensionais sob o paradigma da teoria de resposta do item.

O pacote *psych* foi desenvolvido por Revelle (2014) e, geralmente, é utilizado para análises psicométricas, psicologia experimental e multivariadas. Além de fornecer estatísticas descritivas básicas, fornece modelos de traços latentes unidimensionais e multidimensionais sob o enfoque da teoria de resposta do item.

O próprio *software* conta com funções que dicotomizam os dados, através de uma correção baseada no gabarito inserido, e possui funções que tratam tanto da análise via TCT, quanto via TRI, como pode ser visto nas seções 3.2.1 e 3.2.2.

3.2.1 Análise dos itens pela TCT

Algumas estatísticas básicas foram obtidas, como o cálculo de medidas de posição e dispersão e análise das frequências de resposta. Para isto, utilizou-se as funções *summary()* e *descript()* do pacote *ltm* do *software* R, que calculam estas estatísticas pela forma convencional, conforme pode ser visto em Rizopoulos (2006).

Dentro da abordagem clássica, a primeira análise feita diz respeito à consistência interna do teste a fim de verificar a qualidade do instrumento de medida na mensuração das habilidades dos respondentes. Para isto, utilizou-se o coeficiente alfa de Cronbach, definido como:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_T^2} \right),$$

em que n é o número total de itens, σ_i^2 é a variância relacionada à cada item do teste e σ_T^2 é a variância do escore total dos respondentes.

Em seguida, foi analisado para cada item avaliado o percentual de acertos e erros e, conseqüentemente, o índice de dificuldade, viabilizando a classificação dos itens entre fácil, moderado e difícil, como proposto pro Condé (2001).

Para a análise da discriminação, utilizou-se o coeficiente de correlação ponto-bisserial (ρ_{pb}) que representa a correlação entre uma variável categórica dicotômica (acerto e erro) e uma variável contínua (escore do teste). Como visto na seção 2.1.3.3, itens com ρ_{pb} maiores que 0,30 são considerados satisfatórios, de acordo com Tôrres (2015).

$$\widehat{\rho}_{pb} = \sqrt{\widehat{p}\widehat{q}} \times \frac{\widehat{\mu}_1 - \widehat{\mu}_0}{\widehat{\sigma}}$$

em que $\widehat{\mu}_1$ e $\widehat{\mu}_0$ são as pontuações médias dos alunos que responderam correta e incorretamente ao item, respectivamente, $\widehat{\sigma}$ é o desvio padrão do escore total na amostra e \widehat{q} é o complemento de \widehat{p} , isto é, $\widehat{q} = 1 - \widehat{p}$.

Para estas análises, também utilizou-se a função *descript()* do pacote *ltm* do software estatístico R (2017). Com os resultados desta função pode-se construir um gráfico com o uso das funções *hist()* e *plot()*, do mesmo pacote. Cabe ressaltar que a correlação biserial e o coeficiente alfa de Cronbach também poderiam ser obtidos pelas funções *biserial.cor* e *cronbach.alpha*, respectivamente.

3.2.2 Análise dos itens pela TRI

Sob o enfoque da TRI, primeiramente foram realizados testes para unidimensionalidade, pressuposto necessário de ser atendido para as análises via esta teoria. Dessa forma, através das funções *KMO()* e *cortest.bartlett()*, ambas do pacote *psych*, realizou-se os testes de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett, cuja hipótese de nulidade postula que a matriz de correlações não apresenta problemas para o procedimento de extração de fatores. Assim, procedeu-se à realização da análise fatorial dos dados analisados, através da função *irt.fa()*, do mesmo pacote.

$$KMO = \frac{\sum_{j \neq k} \sum r_{jk}^2}{\sum_{j \neq k} \sum r_{jk}^2 + \sum_{j \neq k} \sum p_{jk}^2}$$

em que r_{jk} é o coeficiente de correlação ponto-biserial entre as variáveis X_j e X_k , com $j, k \in 1, \dots, n$ e p_{jk} é o coeficiente de correlação parcial entre X_j e X_k , ou seja, o coeficiente de correlação quando se exclui o efeito de uma terceira variável dados os outros X 's.

Para estimar os coeficientes dos modelos logísticos, utilizou-se a função *mirt()*, que realiza a estimação pelo método da média a *posteriori* e para a seleção do modelo que melhor se ajusta aos dados, fez-se uso da função *anova()*, ambas do pacote *mirt()*. A função *anova()* fornece os critérios de informação AIC e BIC e executa um teste de razão de verossimilhanças para a comparação dos modelos, cuja hipótese de nulidade postula

que há diferença entre os modelos analisados.

$$\begin{aligned} AIC_p &= -2 \ln(L_p) + 2[(p + 1) + 1] \\ BIC_p &= -2 \ln(L_p) + [(p + 1) + 1] \ln n, \end{aligned}$$

em que L_p é a função de verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo e n o tamanho da amostra.

Com os modelos ajustados, pôde-se realizar a análise fatorial utilizando a função *irt.fa()* e um teste para unidimensionalidade baseado no segundo autovalor da matriz de correlações tetracóricas dos itens dicotomizados, utilizando a função *unidimTest()*. As cargas fatoriais foram obtidas utilizando a função *summary()*. Realizou-se também o teste qui-quadrado de Bock, para verificar a qualidade do ajuste do modelo escolhido. Para isto, utilizou-se a função *item.fit()* do pacote *ltm()*.

Os gráficos das curvas características dos itens e de informação para o modelo ajustado, foram obtidos pela função *plot()*, do pacote *mirt()*, cujas coordenadas foram obtidas pela estimação dos coeficientes já realizada.

A equação do modelo logístico de três parâmetros, relaciona a probabilidade de acerto a um determinado item com os parâmetros deste item e a habilidade do respondente.

$$P(U_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}$$

em que $P(U_{ij} = 1 | \theta_j, a_i, b_i, c_i)$ é a probabilidade do indivíduo j com habilidade θ_j acertar o item i , a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da curva característica do item no ponto de abscissa b_i , que representa o parâmetro de dificuldade do item i , medido na mesma escala de habilidade, c_i é o parâmetro do modelo que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i , também conhecido como probabilidade do chute certo ou probabilidade de acerto casual e D é uma constante de aproximação aos valores da curva normal acumulada, que no caso da curva logística vale 1.

Para determinar a equação do modelo logístico de dois parâmetros, basta substituir $c_i = 0$ na equação do modelo logístico de três parâmetros e para determinar a equação do modelo logístico de um parâmetro, basta substituir $a_i = 1$ na equação do modelo logístico de dois parâmetros.

Finalmente, para o cálculo das proficiências dos estudantes sob o enfoque da TRI, foi utilizada a função *fscores()* também do pacote *mirt()*.

4 RESULTADOS E DISCUSSÃO

O objetivo deste trabalho é o de oferecer uma análise estatística baseada na TCT e na TRI para avaliações de matemática básica que estão sendo oferecidas a estudantes de cálculo I da UFV. Segundo a ótica da TCT, o teste apresentou notas que variaram entre 5 e 44 pontos. A média obtida foi de 20,61 pontos, o desvio-padrão de 6,81 pontos e o coeficiente de variação foi 33,04%. Além disso, o primeiro e o terceiro quartil foram de 16 e 25 pontos, respectivamente.

A Figura 10 representa um gráfico de barras da frequência das notas do teste. Observe que o comportamento desse gráfico lembra o de uma curva normal. É interessante observar que os extremos inferior e superior possuem uma probabilidade ínfima de ocorrência e que a maioria dos estudantes obteve notas que variavam entre 13 e 27 pontos.

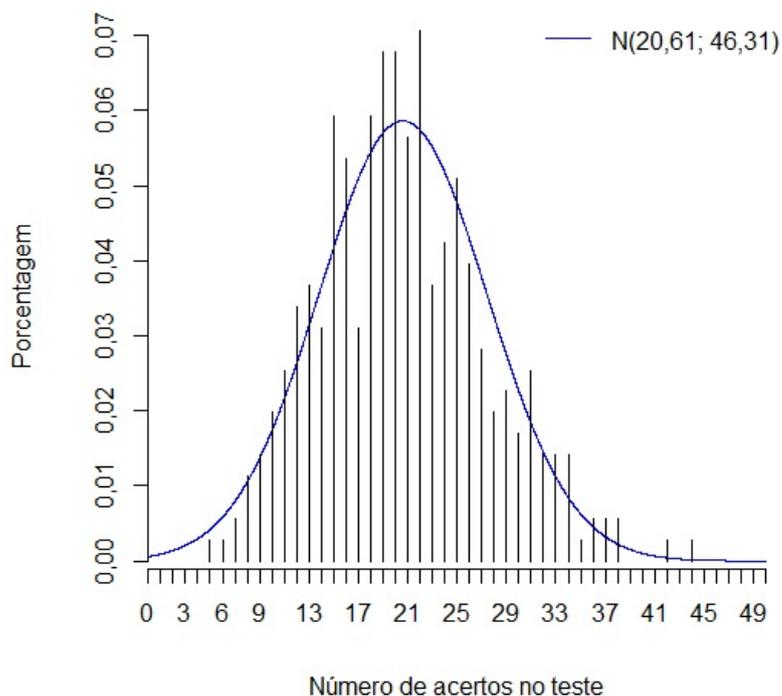


Figura 10 Distribuição do número de acertos dos estudantes.

A Tabela 10 indica algumas análises clássicas básicas na TCT: o índice de dificuldade do item, o coeficiente alpha de Cronbach e o coeficiente de correlação ponto-bisserial.

Primeiramente, observe que os coeficientes de correlação ponto-bisserial dos itens 4, 18, e 46 foram negativos, indicando que esses itens são mais acertados por indivíduos de menor do que os de maior habilidade, indicando que existe algum problema na formulação do item ou em seu gabarito.

Tabela 10 Análise clássica do teste

item	Índice de dificuldade (D_i)	Coefficiente alpha de Cronbach	Correlação pontobisserial ($\hat{\rho}_{pb}$)
1	0,2853	0,7998	0,3810
2	0,5734	0,8038	0,2455
3	0,9153	0,8036	0,2804
4	0,6977	0,8166	-0,2032
5	0,7881	0,8015	0,3298
6	0,7232	0,8052	0,1930
7	0,3531	0,7971	0,4604
8	0,5169	0,8001	0,3601
9	0,8559	0,8041	0,2354
10	0,7175	0,8078	0,1054
11	0,5395	0,8003	0,3562
12	0,1045	0,8039	0,2518
13	0,4718	0,8004	0,3525
14	0,3531	0,7979	0,4353
15	0,2768	0,8044	0,2239
16	0,0960	0,8060	0,1426
17	0,0960	0,8082	0,0247
18	0,3051	0,8136	-0,0951
19	0,5876	0,8031	0,2675
20	0,2429	0,8003	0,3697
21	0,1186	0,8032	0,2844
22	0,5056	0,8001	0,3593
23	0,3333	0,7994	0,3875
24	0,5395	0,8022	0,2957
25	0,3333	0,8004	0,3550
26	0,2429	0,8040	0,2377
27	0,7006	0,8008	0,3454
28	0,6667	0,7984	0,4202
29	0,1215	0,8061	0,1409
30	0,3079	0,8079	0,1026
31	0,3107	0,8020	0,3048
32	0,7090	0,8025	0,2884
33	0,2853	0,8062	0,1603
34	0,2090	0,8042	0,2284
35	0,6328	0,8003	0,3587
36	0,4407	0,7977	0,4350
37	0,6441	0,8026	0,2835
38	0,3333	0,8051	0,2026
39	0,1667	0,8065	0,1301
40	0,1610	0,8080	0,0651
41	0,4266	0,8112	0,0070
42	0,2401	0,8048	0,2077
43	0,4153	0,7989	0,3994
44	0,3588	0,8024	0,2897
45	0,3559	0,7993	0,3886
46	0,0367	0,8086	-0,0671
47	0,1328	0,8078	0,0671
48	0,4350	0,8025	0,2855
49	0,2797	0,8059	0,1718
50	0,6667	0,8019	0,3078

Descartando estes itens temos que, dos 47 itens restantes, de acordo com a classificação sugerida por Condé (2001), 7 são classificados como fáceis, 24 como moderados e 16 como difíceis. O item mais fácil foi o 3, que obteve cerca de 91,53% de acerto, enquanto que os mais difíceis foram os itens 16 e 17, que obtiveram apenas cerca de 9,60% de acerto, cada.

A Figura 11 traz o item mais fácil pela ótica da TCT e o respectivo histograma, cujo eixo das abscissas representa a alternativa marcada pelo estudante e eixo das ordenadas representa a proporção com que esta alternativa foi marcada. Observe que se trata de um item que exigia um conhecimento simples de expressões numéricas, especificamente era exigido do aluno apenas uma escrita na forma decimal de uma fração irredutível. Dos 354 respondentes, 324 acertaram este item.

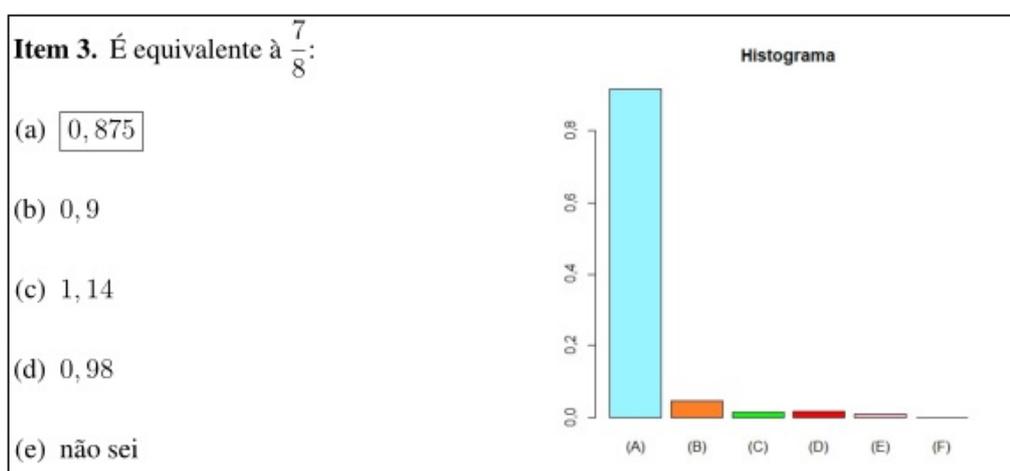


Figura 11 Item 3 e histograma das alternativas marcadas.

As Figuras 12 e 13 trazem os itens mais difíceis do teste pela ótica da TCT e os respectivos histogramas, cujo eixo das abscissas representa a alternativa marcada pelo estudante e eixo das ordenadas representa a proporção com que esta alternativa foi marcada. Apenas 14 dos 354 respondentes acertaram estes itens. Em ambos, exigia-se do aluno um conhecimento prévio em fatoração e simplificação de frações algébricas, indicando que estes tópicos precisam ser retomados e revisados pelos professores da disciplina.

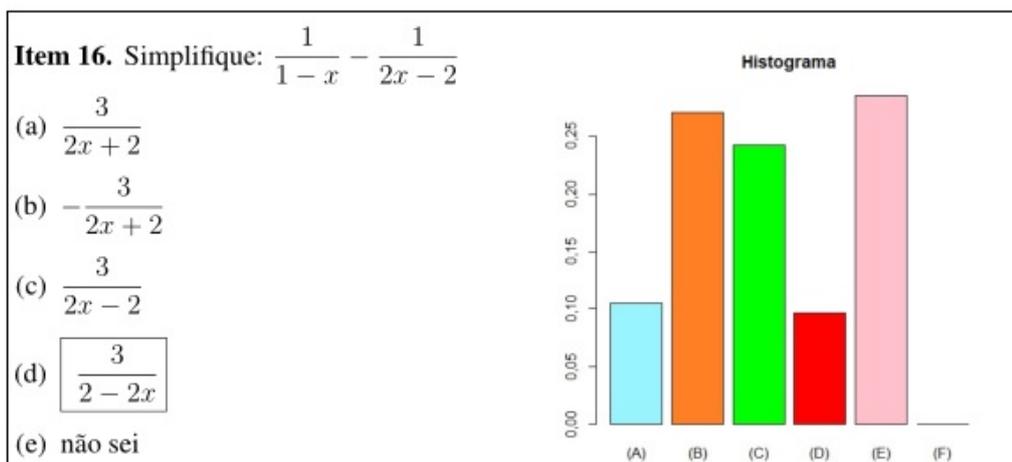


Figura 12 Item 16 e histograma das alternativas marcadas.

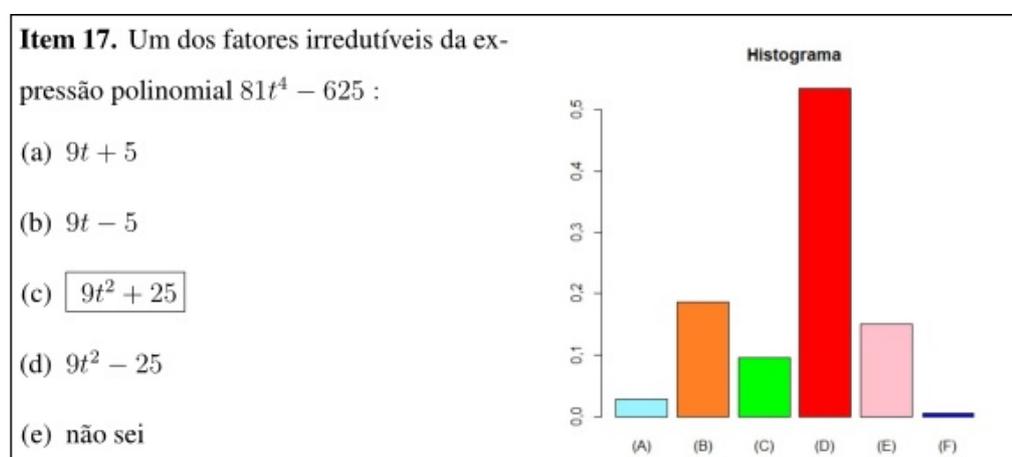


Figura 13 Item 17 e histograma das alternativas marcadas.

Quanto a precisão do instrumento de medida, nenhum item apresentou problemas de consistência interna, visto que todos apresentam coeficiente alpha de Cronbach maior do que 0,7, sugerido por Hair Júnior et. al (2010) como um valor satisfatório.

Para avaliar a dimensionalidade do teste foi realizada uma análise de fatores via componentes principais. Nesta análise, verificou-se KMO de 0,76 e teste de esfericidade de Bartlett significativo, o que indica adequação dos dados para a realização da análise.

Observou-se um fator predominante, indicando evidências de unidimensionalidade, que será melhor discutida depois de selecionarmos o modelo logístico que melhor se ajusta aos dados empíricos.

A Tabela 11 representa a saída da função *anova* do software R, ao se comparar os modelos de um e dois parâmetros:

Tabela 11 Comparação entre os modelos logísticos de 1 e 2 parâmetros

Modelo	AIC	BIC	Verossimilhança	valor- p
Rasch	19386,50	19583,84	-9642,25	
Logístico de 2 parâmetros	19090,40	19477,33	-9445,20	0

Observe que, quando comparados, os modelos logísticos de 1 e 2 parâmetros apresentaram uma diferença significativa à 5% de significância, pelo teste da razão de verossimilhanças (valor- $p < \alpha$), o que indica que um modelo se sobressai ao outro em relação ao ajuste. Analisando a Tabela 11, podemos afirmar que, tanto pelos critérios de informação quanto pelo teste da razão de verossimilhanças que podemos optar pelo modelo logístico de 2 parâmetros. Na Tabela 12 temos a comparação entre os modelos de dois e três parâmetros:

Tabela 12 Comparação entre os modelos logísticos de 2 e 3 parâmetros

Modelo	AIC	BIC	Verossimilhança	valor- p
Logístico de 2 parâmetros	19090,40	19477,33	-9445,20	
Logístico de 3 parâmetros	19129,85	19710,24	-9414,92	0,1457

Observe que os modelos logísticos de 2 e 3 parâmetros não apresentam diferenças significativas à 5% de significância (valor- $p > \alpha$), pelo teste de razão de verossimilhanças, quanto ao ajustamento. Este fato pode ser explicado pela presença da opção “*não sei*” entre as alternativas de resposta dos itens, o que tende a minimizar a marcação aleatória de uma das alternativas, quando o aluno não tem o conhecimento necessário para a solução do item. Assim, optamos pelo modelo logístico de 2 parâmetros devido a menor complexidade do tratamento estatístico dos dados. A Figura 14 representa a curva característica do teste obtida por este modelo.

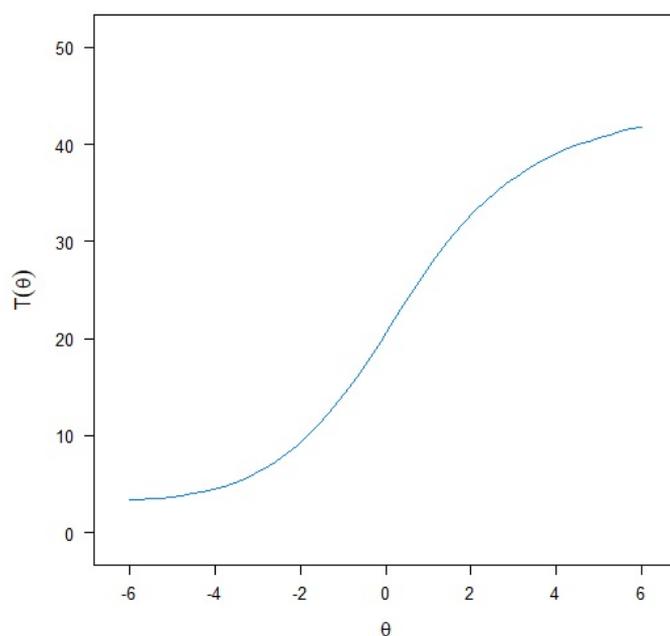


Figura 14 Curva característica do teste para o modelo logístico de dois parâmetros.

A curva característica do teste obtida na Figura 14, nos permite interpretar que para que um indivíduo obtenha escore 20 na avaliação analisada, é necessário, em média, uma habilidade igual a zero. Da mesma forma, um indivíduo com habilidade maior do que dois tende a acertar mais do que 30 itens da avaliação. Subjetivamente, o formato discriminativo da curva e a variação entre os escores obtidos nos permitem fazer uma avaliação inicial positiva do teste, quando o mesmo é analisado através de um modelo logístico de dois parâmetros da TRI.

Com o modelo estimado, verificou-se, empiricamente, o pressuposto da unidimensionalidade. O valor- p retornado pelo comando *unidimTest()* do software R foi de 0,6337, assim inferimos que o segundo autovalor dos dados observados é substancialmente menor do que o segundo autovalor dos dados sob o modelo assumido, o que confirma a unidimensionalidade. Além disso, com exceção dos itens 4, 18 e 46, que apresentaram cargas fatoriais negativas, indicando um impacto inverso no fator, os demais itens apresentaram cargas fatoriais positivas e significativas, com média igual a 0,3977 e variância 0,0293.

Em seguida, visando analisar a qualidade do ajustamento, realizou-se o teste qui-quadrado de Bock, cuja hipótese de nulidade postula que o modelo se ajusta bem aos itens. A Tabela 13 traz os valores χ^2 e os valores- p calculados. Observe que apenas os itens 4 e 42 não estão bem ajustados ao modelo logístico de dois parâmetros, sendo que o primeiro deles já apresentara problemas quanto a discriminação pela análise clássica, como visto na Tabela 10.

Tabela 13 Teste χ^2 de Bock para a qualidade do ajuste

item	χ^2	valor- p	item	χ^2	valor- p
1	7,4971	0,4841	26	9,0438	0,3386
2	13,9848	0,0822	27	4,7980	0,7789
3	13,7720	0,0879	28	8,2244	0,4119
4	21,5465	0,0058	29	4,0019	0,8569
5	11,8790	0,1567	30	9,9587	0,2679
6	3,2847	0,9152	31	10,6502	0,2223
7	8,0858	0,4251	32	5,0670	0,7504
8	7,8704	0,4462	33	5,5539	0,6971
9	9,0720	0,3363	34	14,8208	0,0627
10	9,9734	0,2669	35	8,8112	0,3585
11	8,7022	0,3670	36	12,3224	0,1374
12	4,4941	0,8100	37	10,7124	0,2185
13	5,0813	0,7375	38	4,2839	0,8306
14	14,1988	0,0767	39	6,1166	0,6342
15	8,2905	0,4056	40	6,9462	0,5424
16	8,8568	0,3545	41	6,2562	0,6186
17	5,1865	0,7375	42	16,9541	0,0306
18	14,6937	0,0654	43	4,1678	0,8417
19	7,1006	0,5258	44	5,8573	0,6632
20	10,2230	0,2497	45	10,7028	0,2191
21	8,6842	0,3696	46	9,9262	0,2702
22	7,8477	0,4485	47	15,0701	0,0578
23	6,5422	0,5867	48	8,6915	0,369
24	4,1553	0,8428	49	9,7366	0,284
25	8,0298	0,4306	50	5,5831	0,6938

A Tabela 14 traz os valores estimados para o índice de discriminação e o índice de dificuldade para os itens do teste. É interessante ressaltar que os itens 4, 18 e 46 também apresentaram discriminação negativa pela análise da TRI, o que indica algum erro na elaboração do item ou no gabarito, a mesma conclusão que havíamos obtido ao analisar o coeficiente de correlação ponto bisserial, na TCT.

Tabela 14 Parâmetros de discriminação (a) e dificuldade (b) dos itens

item	<i>a</i>	<i>b</i>	item	<i>a</i>	<i>b</i>
1	1,112	1,017	26	0,651	1,901
2	0,589	-0,546	27	1,083	-0,969
3	1,565	-2,050	28	1,396	-0,681
4	-0,564	1,586	29	0,479	4,311
5	1,227	-1,632	30	0,230	3,569
6	0,559	-1,837	31	0,815	1,110
7	1,508	0,552	32	0,855	-1,202
8	0,986	-0,090	33	0,370	2,557
9	0,927	-2,216	34	0,638	2,259
10	0,241	-3,912	35	1,027	-0,649
11	1,000	-0,199	36	1,276	0,234
12	0,826	2,912	37	0,768	-0,874
13	0,965	0,132	38	0,440	1,645
14	1,380	0,582	39	0,309	5,314
15	0,532	1,915	40	0,143	11,561
16	0,533	4,417	41	0,019	15,355
17	0,057	39,2935	42	0,499	2,430
18	-0,093	-8,858	43	1,134	0,370
19	0,703	-0,563	44	0,716	0,898
20	1,116	1,254	45	0,986	0,714
21	1,071	2,242	46	-0,460	-7,316
22	1,035	-0,035	47	0,218	8,695
23	1,052	0,797	48	0,697	0,411
24	0,711	-0,253	49	0,402	2,438
25	1,069	0,788	50	0,856	-0,938

Descartando estes itens, observe que dos 47 itens restantes, 16 apresentam uma discriminação muito baixa ou baixa, 27 apresentam discriminação média e 4 discriminação alta. Em média, obtemos discriminação igual a 0,7131 com variância de 0,2189. O item mais fácil, pela ótica da TRI, foi o item 10, e o mais difícil, em conformidade com a análise pela TCT, foi o item 17. O item 3 que pela TCT foi o item mais fácil, obteve um baixo valor para o parâmetro de dificuldade pela análise via TRI, confirmando a conclusão

obtida sobre o baixo nível de dificuldade do mesmo.

A Figura 15 traz o item mais fácil pela ótica da TRI e o respectivo histograma, cujo eixo das abscissas representa a alternativa marcada pelo estudante e eixo das ordenadas representa a proporção com que esta alternativa foi marcada. Observe que este item também exigia do estudante conhecimentos prévios no desenvolvimento de equações numéricas, como o item 3, mais fácil sob o enfoque da TCT. O percentual de acerto para este item foi de 71,75%, sendo classificado também como fácil, tomando como base a classificação proposta por Condé (2001), na TCT.

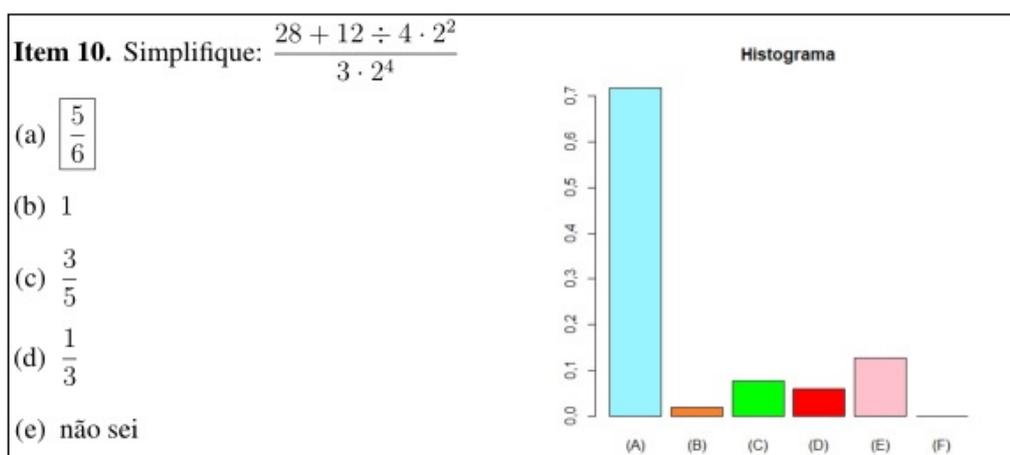


Figura 15 Item 15 e histograma das alternativas marcadas.

A Figura 16 mostra as curvas características dos 50 itens do teste. Observe que, embora a maioria dos itens se comportem como uma função logística, alguns como os itens 4, 18 e 46, indicam algum tipo de falha, que já esperávamos pela interpretação feita através do coeficiente de correlação ponto-bisserial, na análise clássica, e também pela interpretação do índice de discriminação, na análise via TRI.

Observe também que o itens classificados como mais difíceis tendem a obter curvas mais voltadas para a direita e itens mais fáceis, para a esquerda, tomando como referencial o ponto em que a habilidade é zero. Itens como o 3, 7, 14 e 28, cuja discriminação é considerada alta, apresentam curvas bem definidas, o que indica qualidade na estrutura dos mesmos.

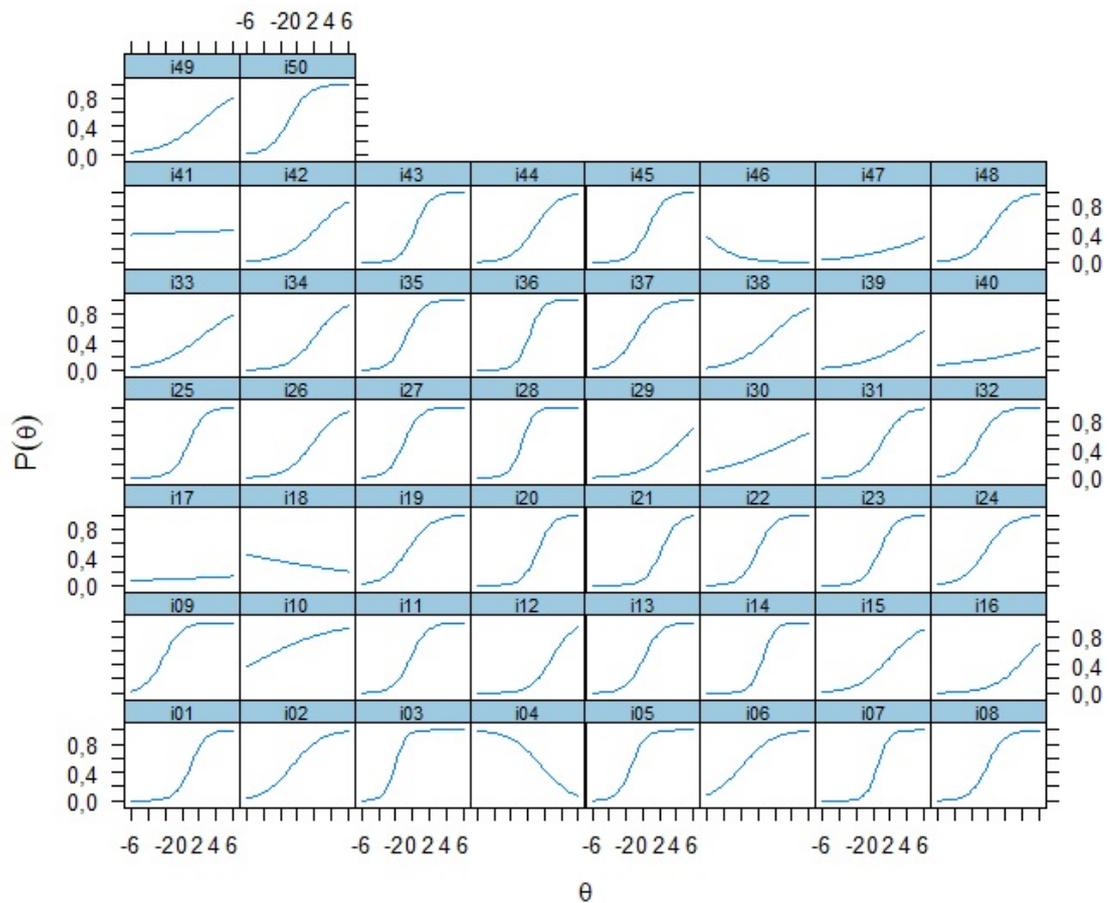


Figura 16 Curvas características dos itens.

As Figuras 17, 18 e 19 trazem os itens que apresentaram falha e os respectivos histogramas, cujo eixo das abscissas representa a alternativa marcada pelo estudante e eixo das ordenadas representa a proporção com que cada alternativa foi marcada.

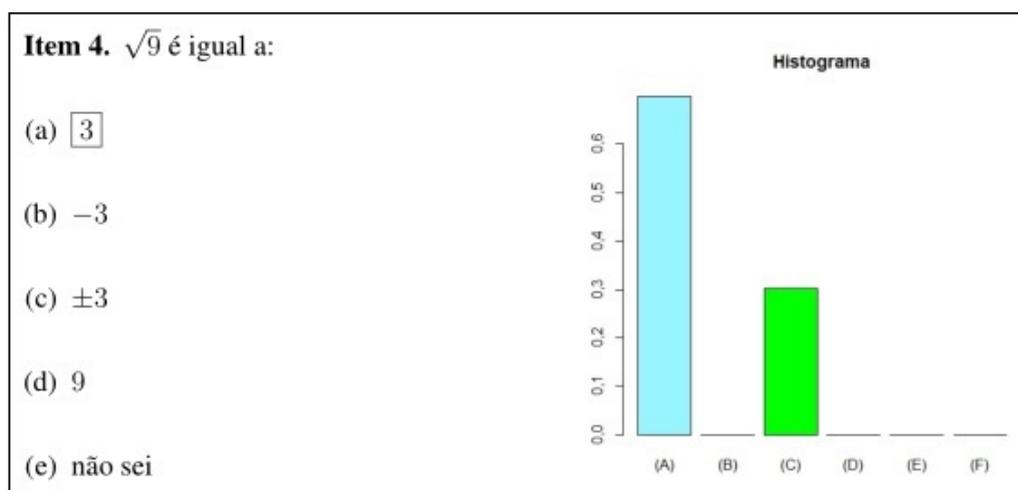


Figura 17 Item 4 e histograma das alternativas marcadas.

Acredita-se que o problema evidenciado no item 4 tenha relação com uma confusão com o sinal de suas alternativas que pode ter ocorrido entre os estudantes com um

maior grau de conhecimento e que entendem que existe uma diferença entre as soluções da equação $x^2 = 9$ e da radiciação $\sqrt{9}$.

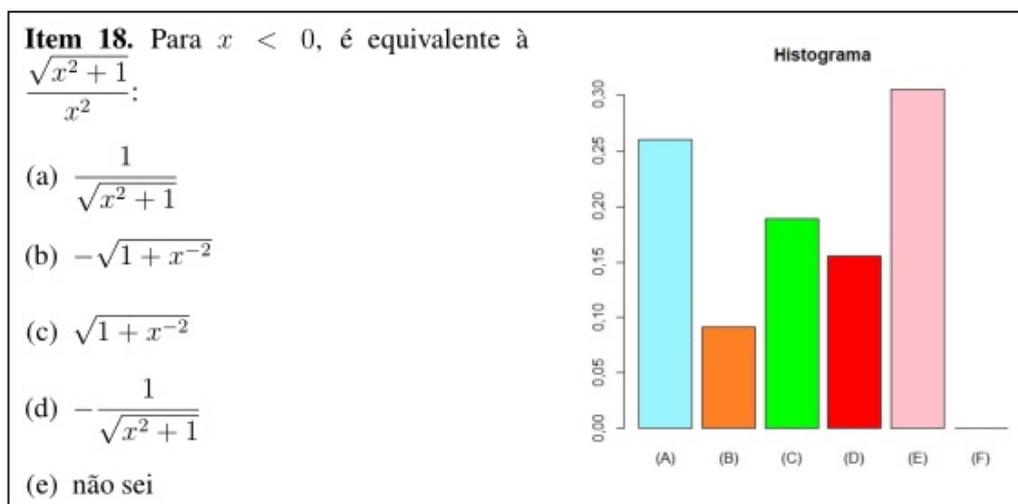


Figura 18 Item 18 e histograma das alternativas marcadas.

O item 18 não apresenta alternativa correta, o que pode ser um indicativo do erro indicado pelas análises.

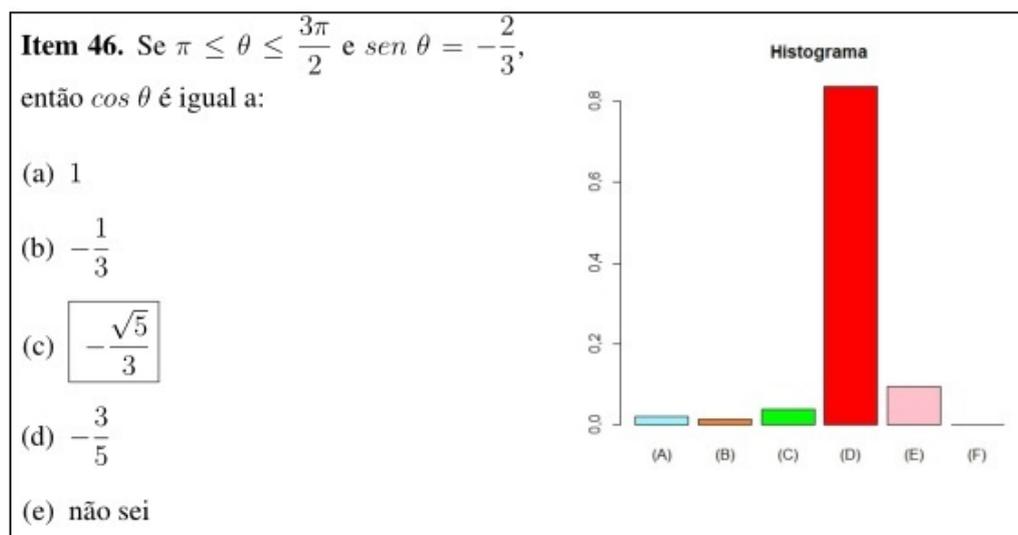


Figura 19 Item 46 e histograma das alternativas marcadas.

o item 46 teve um baixo índice de acerto, como pode ser visualizado no histograma. Apenas 13 dos 354 estudantes apresentaram resposta correta a este item e acredita-se que este fato tenha sido o responsável pelos índices de discriminação terem sido negativos, pois pouca inferência pode ser realizada entre os respondentes que acertaram a este item devido ao tamanho desta população.

A Figura 20 traz as curvas de informação dos itens. Nelas, pode-se perceber que alguns itens trazem mais informações para o cálculo das habilidades do que outros. Por

exemplo, os itens 3 e 7 são bastante informativos. O primeiro, para indivíduos com habilidades negativas. O segundo, para indivíduos com habilidades positivas. Entretanto, alguns itens como o 10, 17, 18, 30, 40, 41 e 47 não acrescentam informações em nenhum grupo de habilidades. Sendo assim, se o objetivo do teste é avaliar habilidades estes item também podem ser descartados.

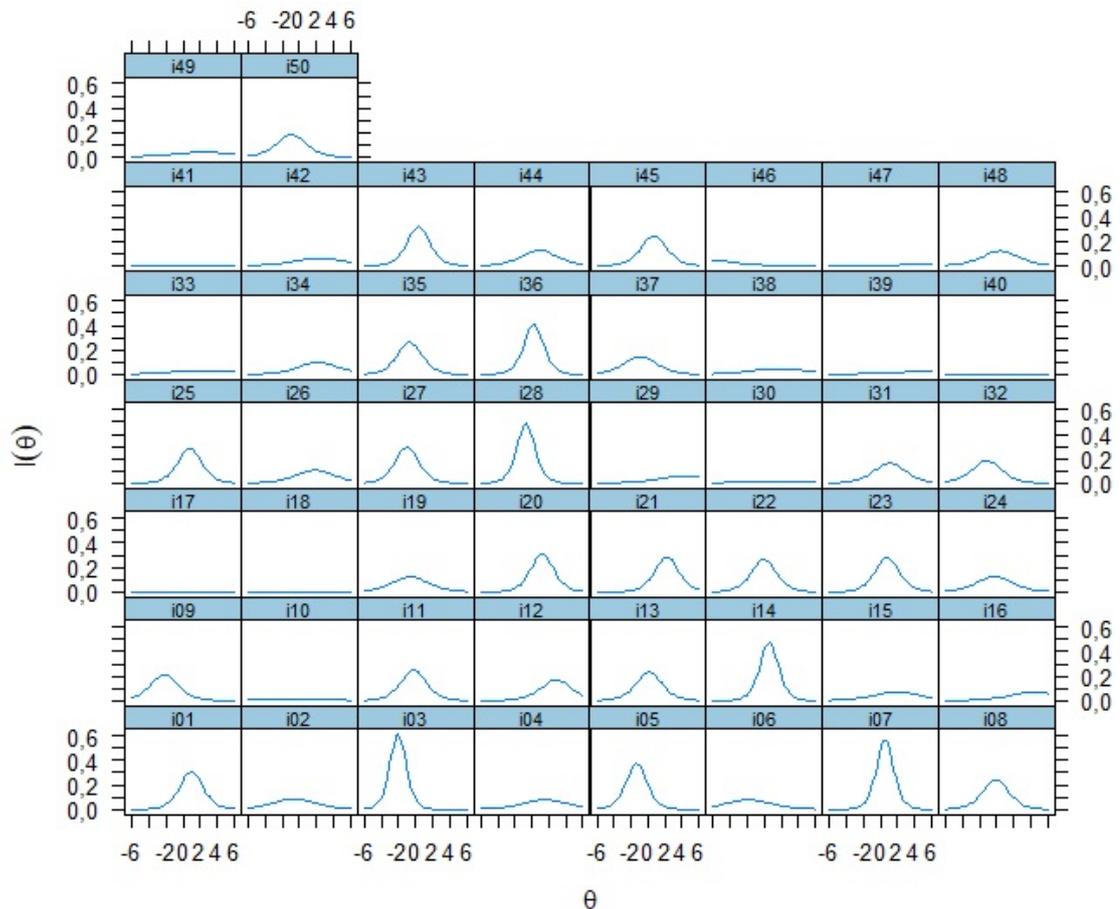


Figura 20 Curvas de informação.

Ao se estimar as habilidades dos estudantes em matemática pelo método da TRI, observa-se que os valores das estatísticas clássicas obtidas são bem próximos dos valores obtidos via TCT. A menor e a maior nota obtida no teste foram 4,555 e 43,039 pontos, respectivamente, a média obtida foi de 20,614 pontos, o desvio-padrão de 6,301 e o coeficiente de variação foi 30,57%. Além disso, o primeiro e o terceiro quartil foram de 16,125 e 24,450 pontos, respectivamente.

Os comandos utilizados nessas análises encontram-se no Apêndice A e os tópicos que serviram como base para a elaboração dos itens encontram-se no Apêndice B. No Apêndice C, temos uma análise item-a-item. Note que, no histograma apresentado para cada item, aparece uma alternativa (F), que representa os estudantes que deixaram o item sem nenhuma marcação. Nestes histogramas, o eixo das abscissas representa a alternativa

marcada, enquanto que o eixo das ordenadas representa em que proporção essa alternativa foi marcada. Além disso, as curvas características dos itens e de informação também são apresentadas.

Como pôde-se perceber, a TCT e a TRI caminham juntas e muitas das interpretações obtidas através de uma análise sob enfoque de uma destas teorias acabam sendo confirmadas pela outra. Por exemplo, os mesmos itens que apresentaram problemas na discriminação em uma teoria, apresentaram o mesmo problema na outra e a classificação quanto ao nível de dificuldade em ambas também foram harmoniosas. Isto nos leva a perceber que caso queira-se utilizar apenas uma teoria na análise de uma informação, tanto a TCT quanto a TRI oferecerão parâmetros confiáveis e significativos. Cabe ao pesquisador definir o tipo de análise específica a ser feita e a complexibilidade das interpretações.

5 CONCLUSÕES

Com a popularização do uso de computadores e o desenvolvimento de *softwares* específicos, as análises estatísticas e psicométricas se tornaram acessíveis e fáceis de serem executadas. Especialmente nas teorias de avaliação, este desenvolvimento proporcionou a difusão de conhecimentos até então estacionados e a propagação de métodos ainda desconhecidos. A teoria de resposta ao item, por exemplo, embora tenha sido desenvolvida na década de 1950, se tornou popular apenas a partir da década de 1980, devido a complexibilidade das análises que a envolvem, que necessitavam de recursos computacionais para serem melhor exploradas.

As tendências e perspectivas para a TRI são bastante animadoras. Além do desenvolvimento de um modelo logístico de quatro parâmetros, que considera que as circunstâncias aleatórias relacionadas com falhas do construtor ao elaborar itens para o teste podem ser significativas na análise de testes, modelos para itens politômicos e discursivos tem ganhado bastante atenção dos pesquisadores da área. Dessa forma, é possível enxergar que há um caminho muito grande a percorrer e que muitos trabalhos no campo desta teoria ainda serão desenvolvidos.

Neste trabalho, procurou-se analisar uma avaliação composta por 50 itens de múltipla escolha, por meio da TCT e da TRI. Em geral, os itens desta avaliação se apresentaram com um bom poder discriminativo e índice de dificuldade variados, em ambas as teorias utilizadas. Além disso, não houve probabilidade significativa de acerto aos itens por indivíduos com baixa habilidade, visto que não houve diferenças significativas entre os modelos logísticos de 2 e 3 parâmetros. Isto contribui para uma boa estimação das habilidades dos respondentes, e indica que o processo de seleção dos itens tem ocorrido, em geral, de maneira significativa.

Vale ressaltar que é necessário também a revisão de alguns itens que apresentaram problemas quanto a discriminação tanto pela TCT quanto pela TRI e que caso haja interesse dos responsáveis pela elaboração do teste em diminuir o número de questões, estes podem optar por aquelas que trouxeram maiores informações no cálculo das habilidades dos estudantes.

Também pôde-se perceber algumas vantagens e desvantagens da aplicação da TRI em relação à TCT. Entre as muitas vantagens, ressaltamos a análise item a item, respondente a respondente, obtendo um resultado mais minucioso e representativo na análise do teste. Por outro lado, dentre as desvantagens, destaca-se a complexidade da manipulação

dos dados e interpretação dos resultados.

Com a análise realizada, os professores da disciplina de cálculo podem obter um retorno das habilidades dos estudantes nos principais tópicos trabalhados e, com isto, terem o conhecimento acerca dos tópicos que precisam ser melhor desenvolvidos. A partir daí, pode-se criar um plano de ações, almejando um trabalho minucioso com o objetivo de diminuir, ou pelo menos estacionar, o crescente índice de reprovação na disciplina, por meio de um maior acompanhamento frente aos estudantes que apresentarem maior dificuldade.

O domínio dos tópicos de matemática básica que compuseram o teste, descritos na Tabela 9, são essenciais para um bom desenvolvimento na disciplina de cálculo, embora muitos estudantes apresentem falhas de aprendizagem oriundas da má formação proveniente da educação básica. Sendo assim, o baixo rendimento no teste proposto pode ser indício de que uma intervenção pedagógica deve ser urgentemente realizada, como o encaminhamento à turmas especiais de acompanhamento individualizado ou outras ações que já são realizadas pelo departamento de matemática da UFV. Dessa forma, o teste pode também ser utilizado na seleção de estudantes para participarem destas atividades especiais. Cabe ressaltar que o campo da educação matemática oferece muitos recursos que visam uma melhoria no ensino de cálculo, como a modelagem matemática, a resolução de problemas e as tecnologias de desenvolvimento da aprendizagem e estes recursos também podem contribuir para reverter esta situação.

Outros estudos podem ser realizados a partir deste. Por exemplo, analisar a relação do curso do estudante com os saberes básicos necessários para um bom desenvolvimento da disciplina, avaliar a matriz curricular e debater a criação de uma disciplina de pré-cálculo, inserir a TRI como meio alternativo para o processo de avaliação dentro da universidade, entre outros.

Ações como estas podem contribuir, consideravelmente, para a qualidade das aulas e para uma melhor formação dos envolvidos no processo ensino-aprendizagem, promovendo ganhos de conhecimento e melhorando a produtividade tanto do aluno, quanto do professor.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Notre Dame, v. 19, n. 6, p. 716-723, 1974.
- ALMEIDA, D.; SANTOS, M.; COSTA, A. F. B. Aplicação do coeficiente alfa de Cronbach nos resultados de um questionário para avaliação de desempenho da saúde pública. In: Encontro Nacional de Engenharia de Produção, 30., 2010, São Carlos. **Anais...** São Carlos: Associação Brasileira de Engenharia de Produção, p. 10-15, 2010.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE), 14., 2000, Caxambu. **Anais...** Caxambu: Associação Brasileira de Estatística, 2000. 164 p.
- ANJOS, A. dos; ANDRADE, D. F. Teoria da resposta ao item com uso do R. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE), 20., 2012, João Pessoa. **Anais...** João Pessoa: Associação Brasileira de Estatística, 2012. 110 p.
- ARIAS, M. R. M.; LLOREDA, M. V. H.; LLOREDA, M. J. H. **Psicometria**. [S.1.]: Alianza Editorial, 2006. 488 p.
- AZEVEDO, C. L. N. **Métodos de estimação na teoria de resposta ao item**. 2003. 133 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo. 2003.
- BAKER, F. B. **Item Response Theory: Parameter Estimation Techniques**. New York: University of Wisconsin-Madison, 1992. 528 p.
- BAKER, F. B. **The basics of item response theory**. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001. 185 p.
- BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick. **Statistical Theories of Mental Test Scores**. Reading, MA: Addison-Wesley, Boston, p. 397-479, 1968.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, New York, v. 37, n. 1, p. 29-51, 1972.
- BOCK, R. D.; THISSEN, D.; ZIMOWSKI, M. F. IRT estimation of domain scores. **Journal of educational measurement**, New York, v. 34, n. 3, p. 197-211, 1997.
- BORGATTO, A. F.; ANDRADE, D. F. Análise clássica de testes com diferentes graus de dificuldade. **Estudos em Avaliação Educacional**, São Paulo, v. 23, n. 52, p. 146-156, 2012.

BROWN, T. A. **Confirmatory factor analysis for applied research**. New York: The Guilford Press, 2006. 462 p.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 5 ed. São Paulo: Saraiva, 2004. 526 p.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, Wien, v. 48, n. 6, p. 1-29, 2012.

CONDÉ, F. N. **Análise empírica de itens**. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais-DAEB/INEP/MEC, Brasília, 2001. 193 p.

COUTO, G.; PRIMI, R. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. **Boletim de Psicologia**, São Paulo, v. 61, n. 134, p. 1-15, 2011.

DALMORO, M.; VIEIRA, K.; VENTURINI, J. Percepção dos fatores de sucesso e insucesso de redes interorganizacionais de cooperação. In: V Encontro de Estudos Organizacionais, 2008, Belo Horizonte. **Anais...** Belo Horizonte: ANPAD, p. 1-15, 2008.

DAMÁSIO, B. F. Uso da análise fatorial exploratória em psicologia. **Avaliação psicológica**, Campinas, v. 11, n. 2, p. 213-228, 2012.

EMBRETSON, S. E.; REISE, S. P. **Item response theory**. Psychology Press, Hove, United Kingdom, 2013. 384 p.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Campus/Elsevier, 2009. 646 p.

FLETCHER, P. R. **Da Teoria Clássica dos Testes para os Modelos de Resposta ao Item**. [S.l.]: Escola Nacional de Ciências Estatísticas, Rio de Janeiro: ENCE, 2010. 37 p.

GOMES, A. E.; SILVA, C. Q. Um teste para ajuste de modelos de teoria de resposta ao item sob a suposição de monotonicidade. In: I CONBRATRI - Congresso Brasileiro de Resposta ao Item, 2009, Florianópolis-SC. **Anais...** São Paulo-SP: ABE - Associação Brasileira de Estatística, 2009. p. 1-4.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5 ed. Porto Alegre: McGraw Hill, 2011. 924 p.

HAIR JÚNIOR, J.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate data analysis**. 7th ed. Upper Saddle River: Prentice Hall, 2010. 785 p.

HAMBLETON, R. K.; COOK, L. L. Latent trait models and their use in the analysis of educational test data. **Journal of educational measurement**, New York, v. 14, n. 2, p. 75-96, 1977.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. **Fundamentals of item response theory**. Newbury Park: Sage Publications, 1991. 184 p.

KELLEY, T. L. The selection of upper and lower groups for the validation of test items. **Journal of educational psychology**, Warwick & York, v. 30, n. 1, p. 17-24, 1939.

KLINE, T. **Psychological testing: A practical approach to design and evaluation**. Thousand Oaks, CA: Sage, 2005. 363 p.

LEBLANCA, V.; COXB, M. A. Interpretation of the point-biserial correlation coefficient in the context of a school examination. **Tutorials in quantitative methods for psychology**, UNIV, Montreal, Dept Psychologie Pavillon Marie-Victorin, 90, Ave Vincent Indy, Montreal, QC H2V 2S9, Canada, v. 13, n. 1, p. 46-56, 2017.

LENHARD, T. H. **Métodos de verificação das suposições e da qualidade de ajuste dos modelos TRI cumulativos unidimensionais** [trabalho de conclusão de curso]. Porto Alegre: Universidade Federal do Rio Grande do Sul, Curso de Estatística, Departamento de Estatística; 2013. 30 p.

LEONTITSIS, A.; PAGGE, J. A simulation approach on Cronbach's alpha statistical significance. **Mathematics and Computers in Simulation**, New York, v. 73, n. 5, p. 336-340, 2007.

LOPES, A. Algumas reflexões sobre a questão do alto índice de reprovação nos cursos de cálculo da UFRGS. **Sociedade Brasileira de Matemática**, Rio de Janeiro, v. 27, n. 26, p. 123-146, 1999.

LOPES, F. L.; VENDRAMINI, C. M. M. Propriedades psicométricas das provas de pedagogia do ENADE via TRI. **Avaliação: Revista da Avaliação da Educação Superior**. Sorocaba, v. 20, n. 1, p. 27-47, 2015.

LORD, F. M. A theory of test scores (No. 7). **Psychometric Monograph**. Iowa City, IA: Psychometric Society, v. 35, n. 1, 1952. 84 p.

LORD, F. M.; NOVICK, M. R. **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley, 1968. 568 p.

MCKINLEY, R. L.; MILLS, C. N. A comparison of several goodness-of-fit statistics. **Applied Psychological Measurement**, New York, v. 9, n. 1, p. 49-57, 1985.

MOREIRA, A. C. Comparação da Análise de Componentes Principais e da CATPCA na Avaliação da Satisfação do Passageiro de uma Transportadora Aérea. **Investigação Operacional**, Lisboa, v. 27, n. 2, p. 165-178, 2007.

MUÑIZ, J. **Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa**. Madri: Ediciones Pirámide, S. A., 1990. 158 p.

OLIVEIRA, A. L. S. **Avaliação psicométrica da medida do componente de formação geral da prova do exame nacional de desempenho de estudantes (ENADE) de 2010, 2011 e 2012**. 2017. 107 p. Dissertação (Mestrado em Métodos e Gestão em Avaliação.) - Universidade Federal de Santa Catarina, Florianópolis. 2017.

PARTCHEV, I. **irtoys: Simple Interface to the Estimation and Plotting of IRT Models**. R package version 0.1.3, v. 2, 2009. Disponível em: <http://CRAN.R-project.org/package=irtoys>. Acesso em 04/12/2017.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2003. 397 p.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item TRI. **Avaliação Psicológica**, Campinas, v. 2, n. 2, p. 99-110, 2003.

PIACENTINI, C. C. **Reprovação, abandono e evasão: um estudo de caso no curso de bacharelado em zootecnia da Universidade Tecnológica Federal do Paraná–Campus Dois Vizinhos**. 2012. 115 p. Dissertação (Mestrado em Ciências) - Universidade Federal Rural do Rio de Janeiro, Seropédica. 2012.

PRIMI, R. Psicometria: fundamentos matemáticos da teoria clássica dos testes. **Avaliação Psicológica**, Instituto Brasileiro de Avaliação Psicológica. UFRGS, v. 11, n. 2, p. 297-307, 2012.

PRIMI, R.; HUTZ, C. S.; SILVA, M. C. R. A prova do ENADE de psicologia 2006: concepção, construção e análise psicométrica da prova. **Avaliação Psicológica**, Campinas, v. 10, n. 3, p. 271-294, 2011.

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation on Statistical Computing, 2017. Disponível em: <https://www.r-project.org>. Acesso: 04/12/2017.

RASCH, G. **Probabilistic models for some intelligence and achievement tests**. Copenhagen: Danish Institute for Education Research, 1960. 184 p.

REVELLE, W. **psych: Procedures for personality and psychological research**. Northwestern University, Evanston. Illinois, USA, 2014. Disponível em: <https://cran.r-project.org/web/packages/psych/index.html>. Acesso em 04/12/2017.

RIZOPOULOS, D. ltm: An R package for latent variable modeling and item response analysis. **Journal of statistical software**, New York, v. 17, n. 5, p. 1-25, 2006.

RUGGIERO, M. A. G.; LOPES, V. L. R. **Cálculo numérico: aspectos teóricos e computacionais**. Makron Books do Brasil, 1997. 395 p.

SAMEJIMA, F. Normal ogive model on the continuous response level in the multidimensional latent space. **Psychometrika**, New York, v. 39, n. 1, p. 111-121, 1974.

SANTOS, L. M. **Desempenho escolar em Pernambuco: análise dos itens e das habilidades usando teoria clássica e tri**. 2008. 104 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife. 2008.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, New York, v. 6, n. 2, p. 461-464, 1978.

STERNICK, C. Características técnicas de um instrumento de medida. **Arquivos Brasileiros de Psicologia Aplicada**, Rio de Janeiro, v. 28, n. 1, p. 48-67, 1976.

TABACHNICK, B.; FIDELL, L. Analysis of covariance. **Using multivariate statistics**, New York, v. 8, n. 1, p. 321-374, 1996.

TÔRRES, F. C. **Uma aplicação da teoria de resposta ao item em um simulado de matemática no modelo enem**. 2015. 116 p. Dissertação (Mestrado em Matemática) - Programa de Mestrado Profissional em Matemática em Rede Nacional, Universidade de Brasília, Brasília, 2015.

VASCONCELOS, R.; ALMEIDA, L. S.; MONTEIRO, S. O insucesso e abandono acadêmico na universidade: Uma análise sobre os cursos de engenharia. In: **International Conference on Engineering and Computer Education**. Buenos Aires: ICECE, cop. 2009. p. 457-461.

VELOSO, T. C.; ALMEIDA, E. P. Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, Campus Universitário de Cuiabá—um processo de exclusão. **Série-Estudos-Periódico do Programa de Pós-Graduação em Educação da UCDB**, v. 1, n. 13, p. 133-148, 2013.

VICINI, L.; SOUZA, A. M. **Análise multivariada da teoria à prática**. Santa Maria: UFSM, CCNE, 2005. 215 p.

WILLSE, J. T.; SHU, Z. **CTT: Classical test theory functions**. R package version, v. 2, 2014. Disponível em: <http://CRAN.R-project.org/package=CTT>. Acesso em: 04/12/2017.

WISLAND, B.; FREITAS, M. D. C. D.; ISHIDA, C. Y. Desempenho acadêmico dos alunos em curso de engenharia e licenciatura na disciplina de cálculo i. **Iberoamerican Journal of Industrial Engineering**, Florianópolis, v. 6, n. 11, p. 94-112, 2014.

WRIGHT, B. D. Sample-free test calibration and person measurement. In: **Proceedings of the 1967 invitational conference on testing problems**. Princeton, NJ: Educational Testing Service, 1968. p. 85-101.

YEN, W. M. Using simulation results to choose a latent trait model. **Applied Psychological Measurement**, New York, v. 5, n. 2, p. 245-262, 1981.

APÊNDICE

	Página
APÊNDICE A: Rotina para análise de itens.	66
APÊNDICE B: Matriz de referências para a elaboração do teste.	68
APÊNDICE C: Análise individual dos itens do teste.	72

APÊNDICE A: Rotina para análise de itens.

```
rm(list=ls(all=TRUE))
options(OutDec=",")

#install.packages("mirt, ltm, psych")
library(mirt)
library(ltm)
library(psych)

#Carregando os dados
dados=read.table("C:\\Users\\Usuario\\Dropbox\\Mestrado\\DISSERTACAO DENILSON\\gabaritos 2016 ii.txt")
head(dados)
prova=as.matrix(dados[-1,])

#Carregando o gabarito
gab=as.character(as.matrix(dados[1,]))

#Categorizando os dados
prova.c<-mult.choice(prova,gab)

#Analise descritiva
descript(prova.c)

#Calculando as notas dos respondentes
notas=rowSums(prova.c)

#Resumo dos resultados classicos obtidos
summary(notas)

#Construindo o grafico de frequencias
AcerIt=colSums(prova.c)
ex=as.numeric(names(table(rowSums(prova.c))))
ey=prop.table(table(rowSums(prova.c)))[1:length(names(table(rowSums(prova.c)))]
mean(notas)
sd(notas)
max(notas)
min(notas)

x=seq(0,50,0.01)
plot(x, dnorm(x, mean(notas), sd(notas)), type="l", xlim=c(0,50),
ylim=c(0, max(max(dnorm(x, mean(notas), sd(notas)), max(ey))),
axes=F, ylab="Porcentagem", xlab="Numero de acertos no teste", col="blue")
axis(1, pos=-0.001, at=seq(0,50,1))
axis(2, pos=0, at=seq(0, max(max(dnorm(x, mean(notas), sd(notas)), max(ey))), .01))
lines(ex, ey, type="h")
legend("topright", paste("N(", round(mean(notas), 2), "; ", round(var(notas), 2),
"),", sep=""), box.lty=0, lty=1, col=4)

#Analise fatorial
cortest.bartlett(prova.c, n = NULL, diag=TRUE)
KMO(prova.c)
irt.fa(prova.c)

#Estimacao dos modelos logisticos
mod1<-mirt(prova.c,1, itemtype='Rasch')
mod2<-mirt(prova.c,1, itemtype='2PL')
mod3<-mirt(prova.c,1, itemtype='3PL')

#Testes de comparacao entre os modelos
```

```

anova(mod1,mod2)
anova(mod2,mod3)

#Informacoes basicas do modelo escolhido
summary(mod2)

#Teste para a unidimensionalidade do modelo
unidimltm=unidimTest(ltm(prova.c~z1))

#Testes para a qualidade do ajuste
item.fit(modelo2,G=10,FUN=mean)
item.fit(modelo2,G=5,FUN=median)

#Coeficientes do modelo e analise descritiva dos mesmos
x=coef(mod2,simplify=TRUE,IRTpars=TRUE)
a=x$items[,1]
b=x$items[,2]
mean(a)
sd(a)
mean(b)
sd(b)

#Curvas caracteristicas dos itens:
plot(mod2, type='trace')

#Curvas de informacao dos itens:
plot(mod2, type='infotrace')

#Plot de curvas individuais dos itens:
itemplot(mod2, 1,col=1,lty=1,lwd=2)

#Calculo das notas em TRI pelo metodo EAP e algumas analises descritivas:
z=fscores(mod2,met="EAP")
ntri=z*sd(notas)+mean(notas)
mean(ntri)
sd(ntri)
max(ntri)
min(ntri)

#Resumo das notas em TRI:
notas=ntri
summary(notas)

#Fim

```

APÊNDICE B: Matriz de referências para a elaboração do teste.

Tópicos tomados como base para a elaboração das questões:

- (1) Lógica matemática.
- (2) Raciocínio lógico.
- (3) Conjuntos: elementos, pertinência e descrição de um conjunto.
- (4) Operações entre conjuntos.
- (5) Conjuntos numéricos: definições e propriedades.
- (6) Potenciação e radiciação. Racionalização.
- (7) Expressão numérica.
- (8) Conjuntos dos números reais e propriedades.
- (9) Intervalos.
- (10) Produtos notáveis.
- (11) Expressões algébricas: operações e valor numérico.
- (12) Fatoração e simplificação de expressões algébricas.
- (13) Valor absoluto e propriedades.
- (14) Equações e inequações modulares.
- (15) Equações e inequações racionais.
- (16) Razão e proporção.
- (17) Taxa e porcentagem.
- (18) Somatório e produtório.
- (19) Equações e inequações irracionais.
- (20) Produto cartesiano e sistema de coordenadas cartesianas.
- (21) Distância entre dois pontos.

- (22) Relações e gráficos de curvas.
- (23) Retas no plano e coeficiente angular.
- (24) Solução gráfica de equações e inequações em uma variável.
- (25) Sistema de equações de duas variáveis.
- (26) Funções: definição. Domínio e imagem.
- (27) Funções: gráficos. Interseção com os eixos coordenados.
- (28) Funções de primeiro e segundo graus.
- (29) Máximos e mínimos locais e globais.
- (30) Simetria. Função par e função ímpar.
- (31) Intervalos de crescimento e decrescimento.
- (32) Raízes de funções. Interseção de funções.
- (33) Operações com funções: adição, multiplicação e divisão.
- (34) Composição de funções.
- (35) Transformação de funções: translação, rotação, reflexão, contração e expansão.
- (36) Polinômios e funções polinomiais: definição e propriedades.
- (37) Divisão de polinômios.
- (38) Raízes de polinômios. Teorema das raízes racionais.
- (39) Teorema do resto e teorema de D'Alembert.
- (40) Algoritmo de Briot Ruffini.
- (41) Equações polinomiais.
- (42) Inequações polinomiais.
- (43) Mínimo múltiplo comum e máximo divisor comum de polinômios. Decomposição em frações parciais.
- (44) Função racional.

- (45) Funções definidas por várias sentenças.
- (46) Função raiz.
- (47) Função modular.
- (48) Taxa de variação média.
- (49) Função sobrejetora, injetora e bijetora.
- (50) Imagem direta e imagem inversa.
- (51) Função inversa: propriedades.
- (52) Função exponencial: definição e propriedades.
- (53) Logaritmos: definição e propriedades. Mudança de base.
- (54) Função logarítmica. Propriedades.
- (55) Equações exponenciais e logarítmicas.
- (56) Inequações exponenciais e logarítmicas.
- (57) Trigonometria e funções trigonométricas.
- (58) Circunferência, ângulos e arcos.
- (59) Relações trigonométricas no triângulo retângulo.
- (60) O círculo trigonométrico.
- (61) Identidade trigonométrica.
- (62) Função trigonométrica.
- (63) Equação trigonométrica.
- (64) Inequação trigonométrica.
- (65) Função trigonométrica inversa.
- (66) Função definida implicitamente.
- (67) Área entre curvas.
- (68) Semelhança de triângulos.

(69) Intervalos em que a função é positiva e negativa.

(70) Intervalos em que a função é côncava e convexa.

(71) Áreas e volumes.

APÊNDICE C: Análise individual dos itens do teste.

Análise do item 1

Item 1. Sejam os intervalos $A = [2; 7]$ e $B = [5; 9]$. Então:

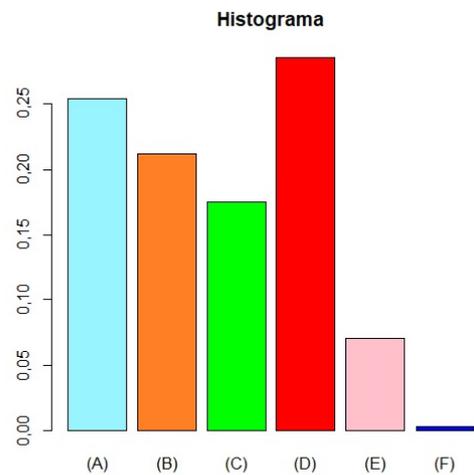
(a) $A \cup B = [2; 9]$

(b) $A \cap B = [5; 7]$

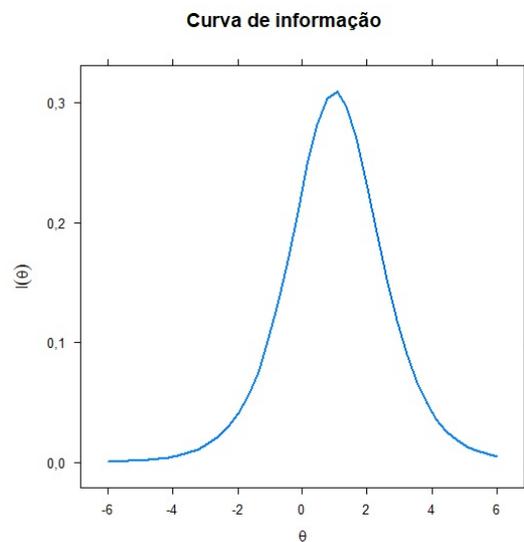
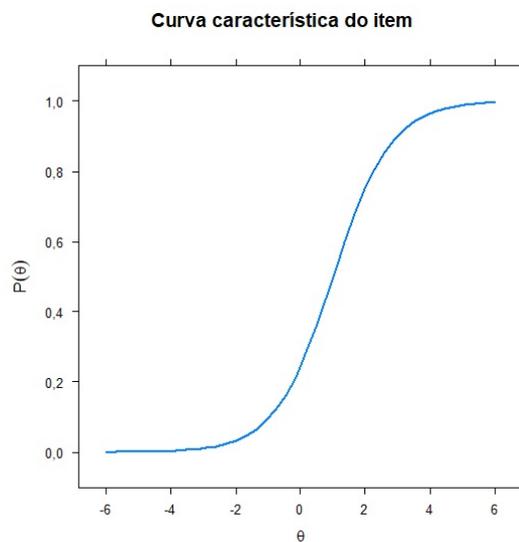
(c) $A - B = [2; 5]$

(d) $B - A = (7; 9]$

(e) não sei



Tópico: Operações entre conjuntos.

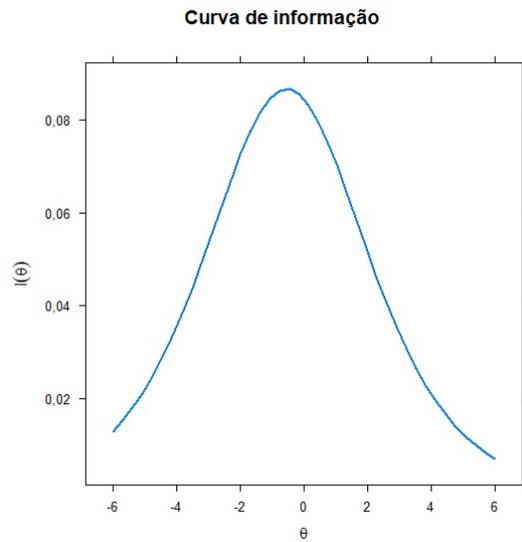
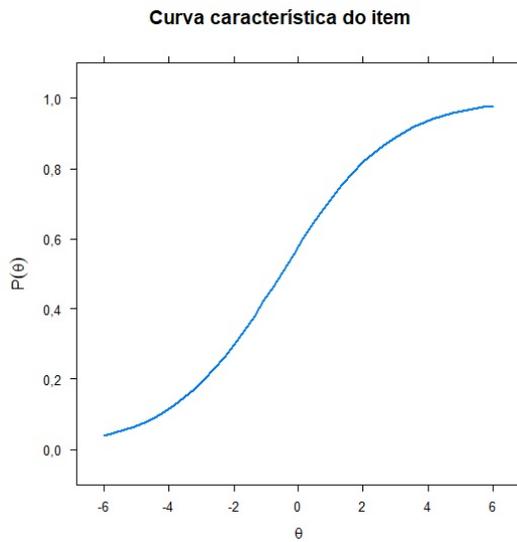
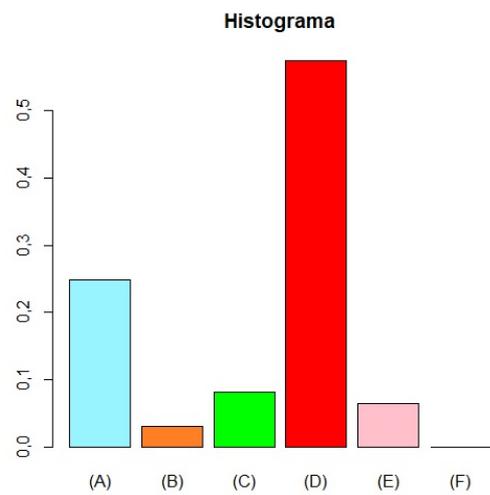


Análise do item 2

Item 2. O máximo divisor comum (MDC) de 198 e 154 é:

- (a) 11
- (b) 33
- (c) 44
- (d)
- (e) não sei

Tópico: Potenciação e radiciação. Racionalização.



Análise do item 3

Item 3. É equivalente à $\frac{7}{8}$:

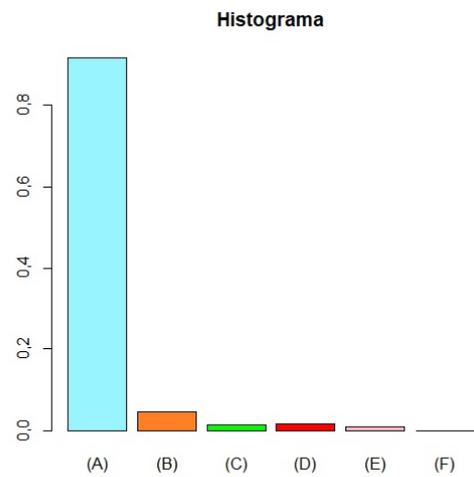
(a)

(b) 0,9

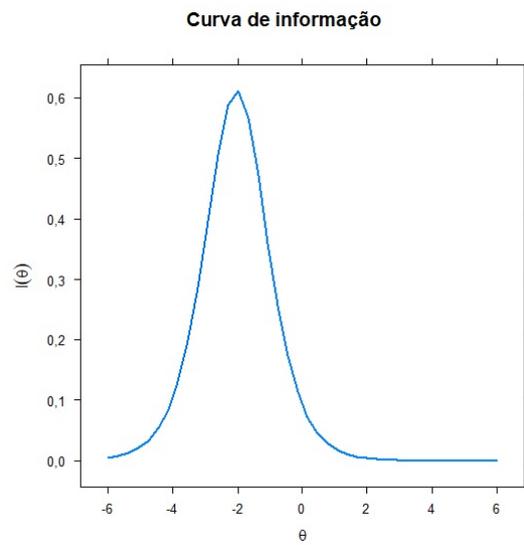
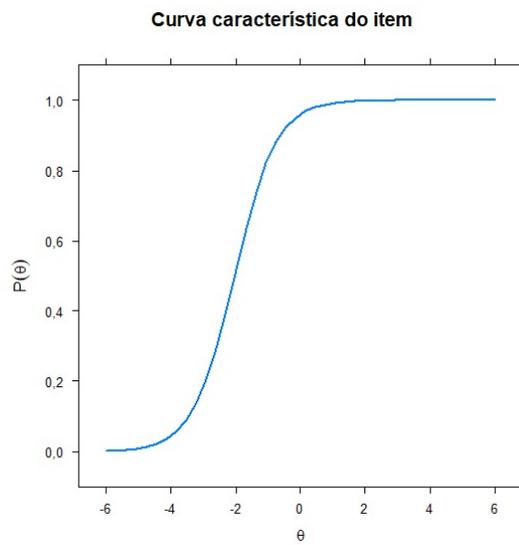
(c) 1,14

(d) 0,98

(e) não sei



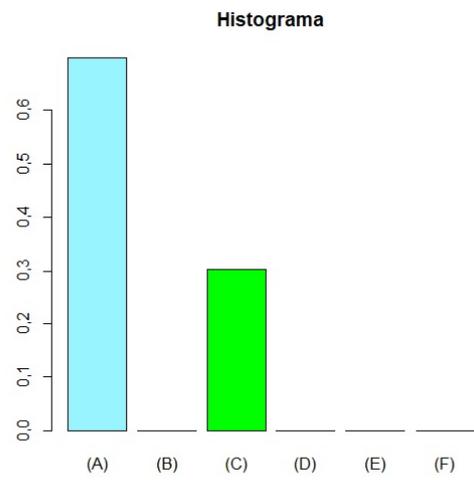
Tópico: Expressão numérica.



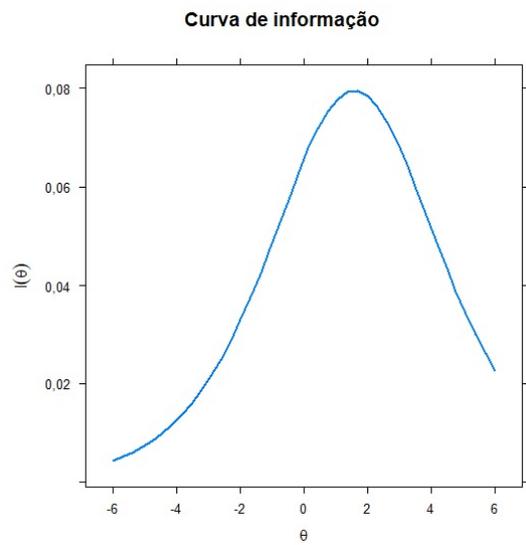
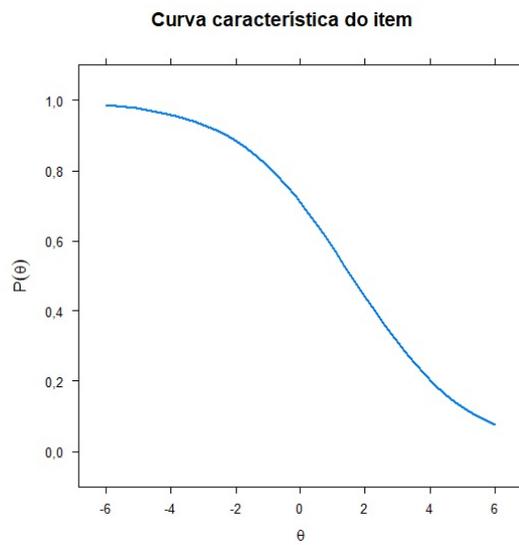
Análise do item 4

Item 4. $\sqrt{9}$ é igual a:

- (a) 3
- (b) -3
- (c) ± 3
- (d) 9
- (e) não sei



Tópico: Expressão numérica.



Análise do item 5

Item 5. Calcule: $\left(\frac{1}{2}\right) : \left(-\frac{3}{8}\right)$

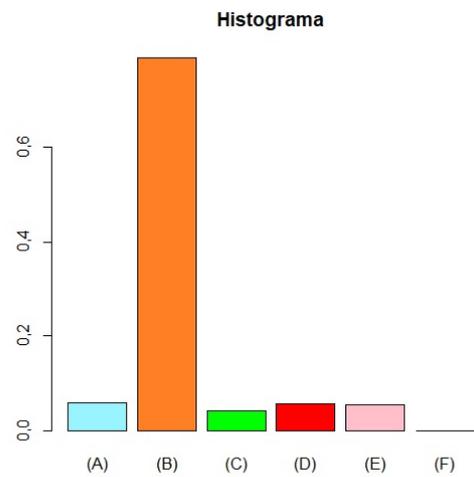
(a) $\frac{4}{3}$

(b) $\frac{4}{3}$

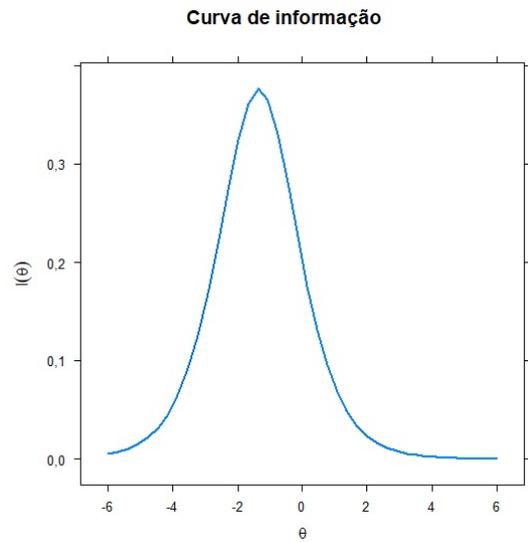
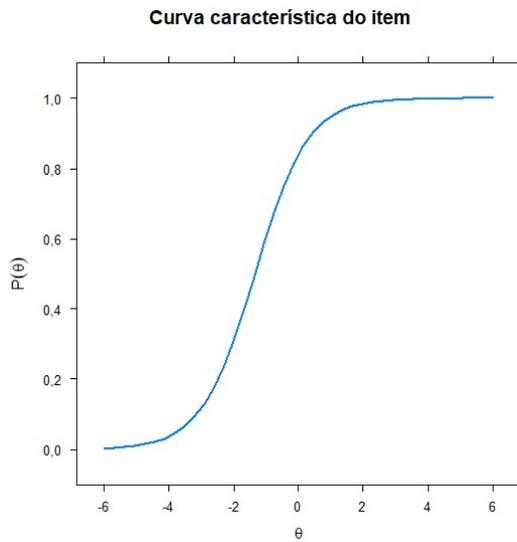
(c) $\frac{4}{3}$

(d) $-\frac{7}{6}$

(e) não sei



Tópico: Expressão numérica.



Análise do item 6

Item 6. Marque a afirmação falsa:

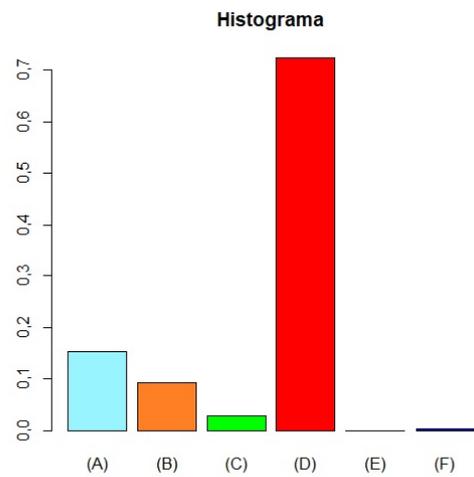
(a) $2 \leq 2$

(b) $\frac{8}{3} > \frac{9}{4}$

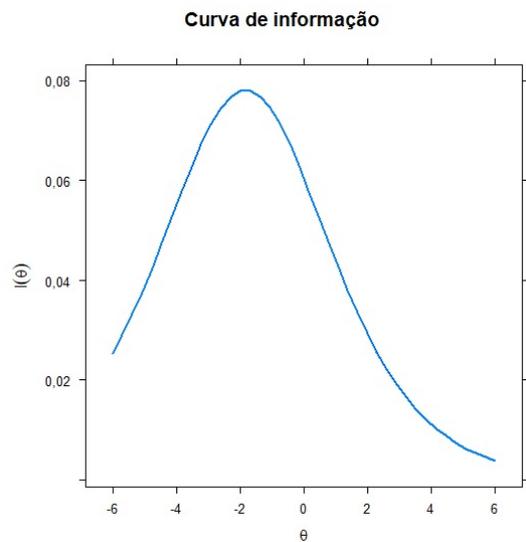
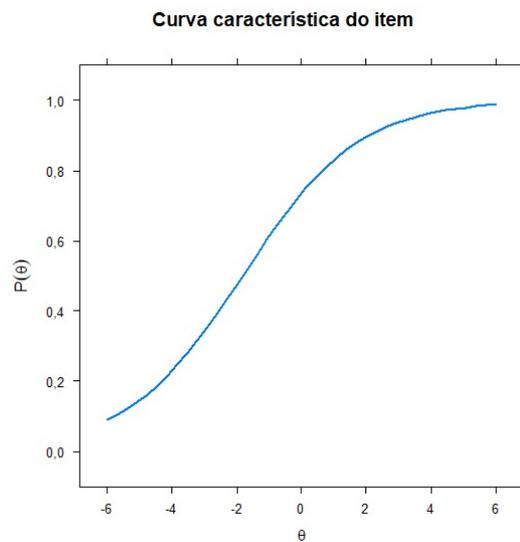
(c) $\sqrt{2} > 1,14$

(d) $-4 > -3$

(e) não sei



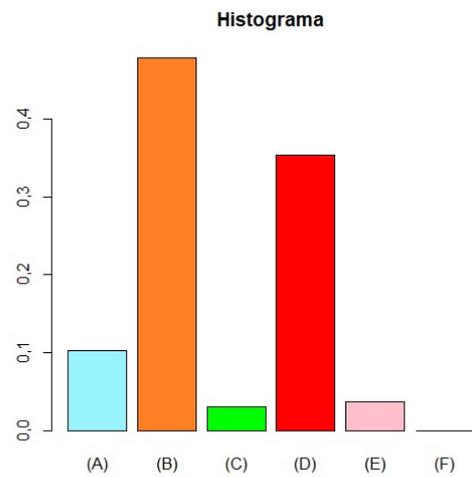
Tópico: Conjuntos dos números reais e propriedades.



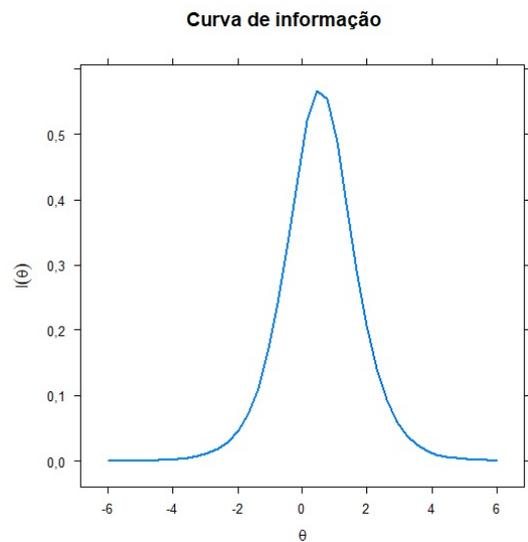
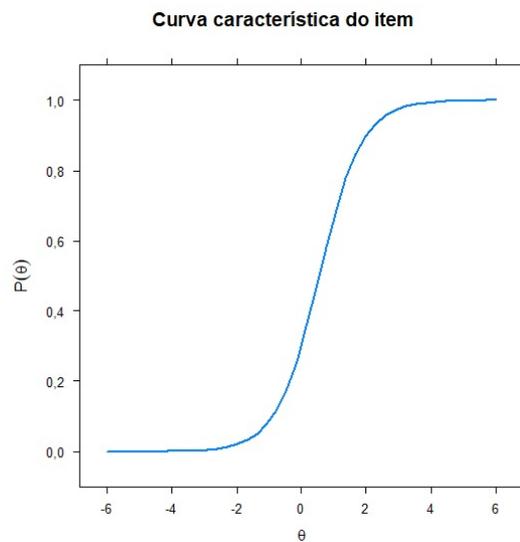
Análise do item 7

Item 7. A expressão $-18 < -6x < 12$ é equivalente à:

- (a) $-3 < x < -2$
- (b) $-3 < x < 2$
- (c) $2 < x < 3$
- (d) $-2 < x < 3$
- (e) não sei



Tópico: Intervalos.



Análise do item 8

Item 8. A fração $\frac{a+b}{c+d}$ é equivalente à:

(a) $\frac{a}{c} + \frac{b}{d}$

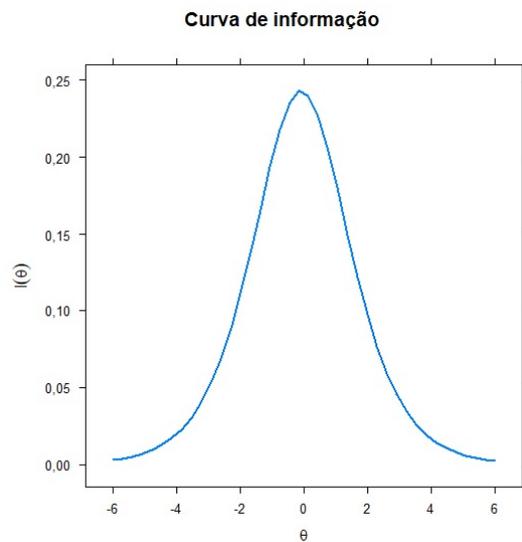
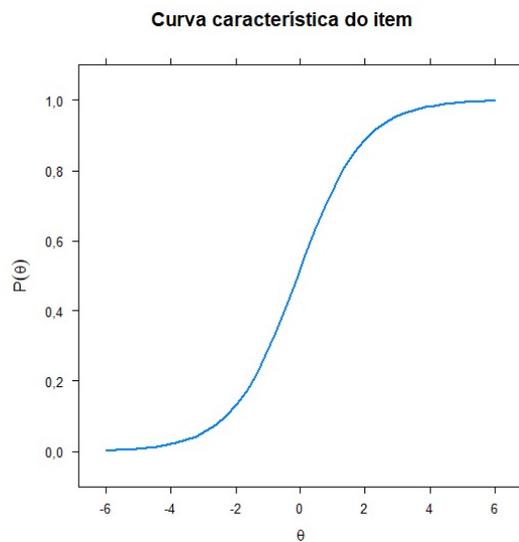
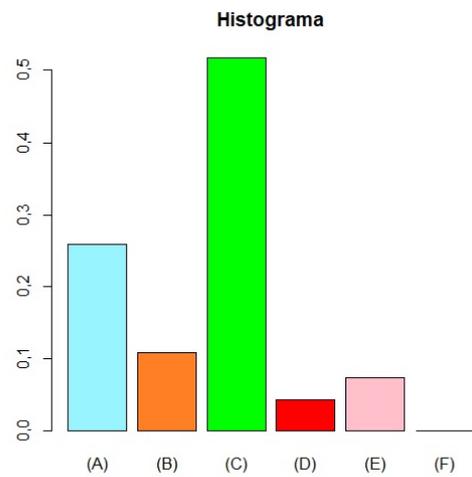
(b) $\frac{a+b}{c} + \frac{a+b}{d}$

(c) $\frac{a}{c+d} + \frac{b}{c+d}$

(d) $\frac{a}{d} + \frac{b}{c}$

(e) não sei

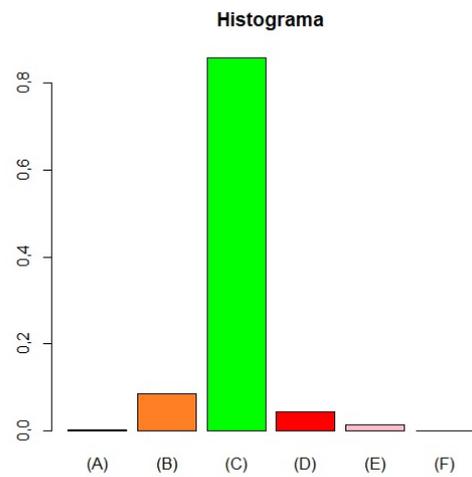
Tópico: Expressões algébricas: operações e valor numérico.



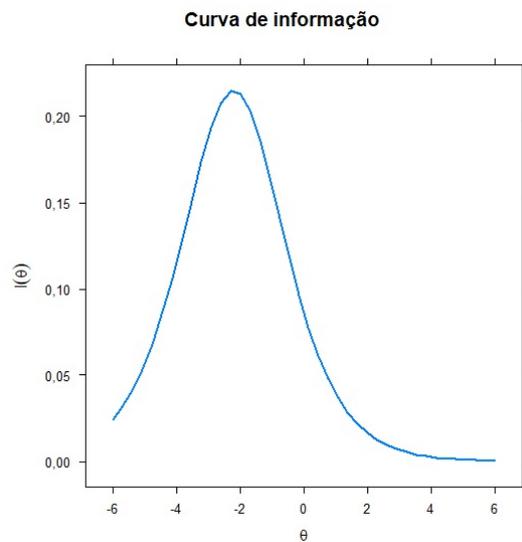
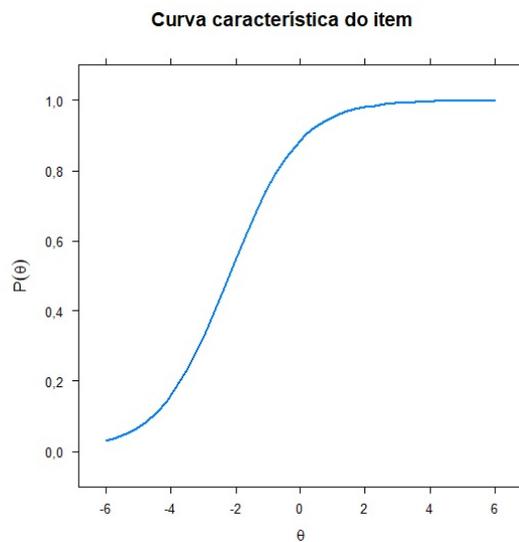
Análise do item 9

Item 9. É equivalente à $(2x + y)(y - 3x)$:

- (a) $-xy$
- (b) $6x^2 + y^2 - xy$
- (c) $-6x^2 + y^2 - xy$
- (d) $-6x^2 + y^2$
- (e) não sei



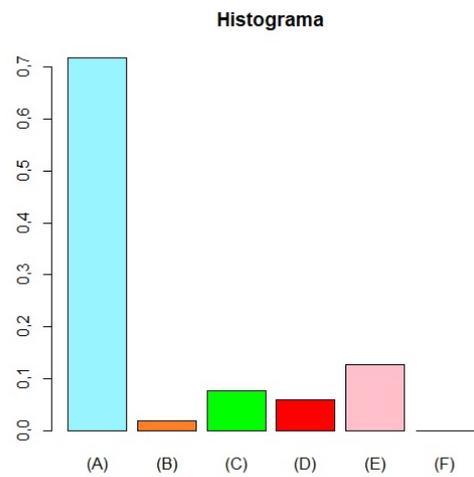
Tópico: Expressões algébricas: operações e valor numérico.



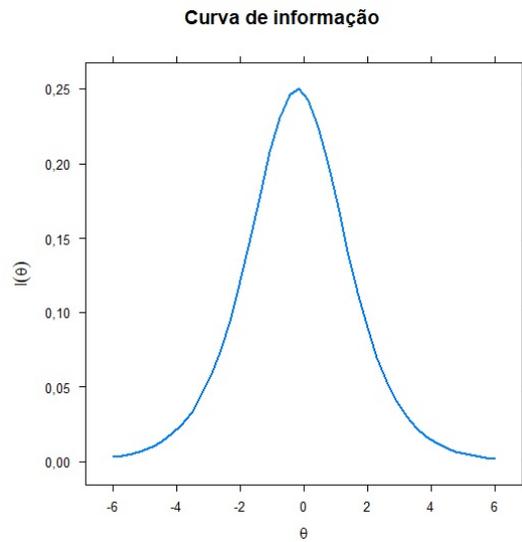
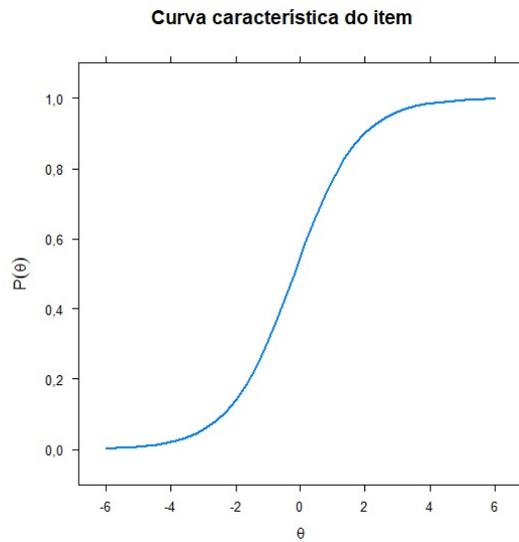
Análise do item 10

Item 10. Simplifique: $\frac{28 + 12 \div 4 \cdot 2^2}{3 \cdot 2^4}$

- (a) $\frac{5}{6}$
- (b) 1
- (c) $\frac{3}{5}$
- (d) $\frac{1}{3}$
- (e) não sei



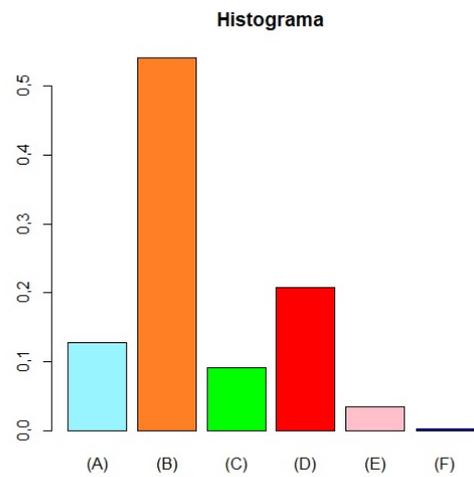
Tópico: Expressões algébricas: operações e valor numérico.



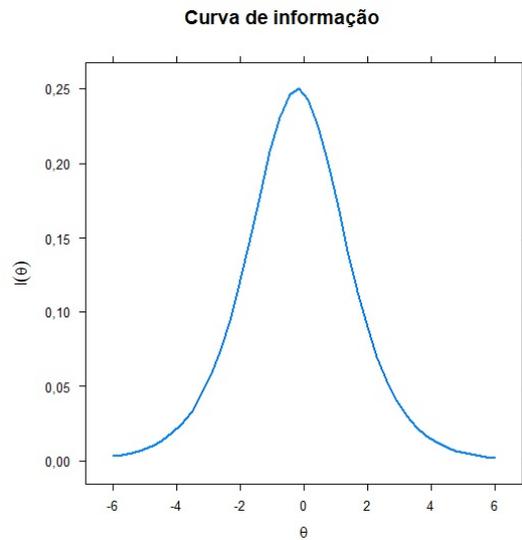
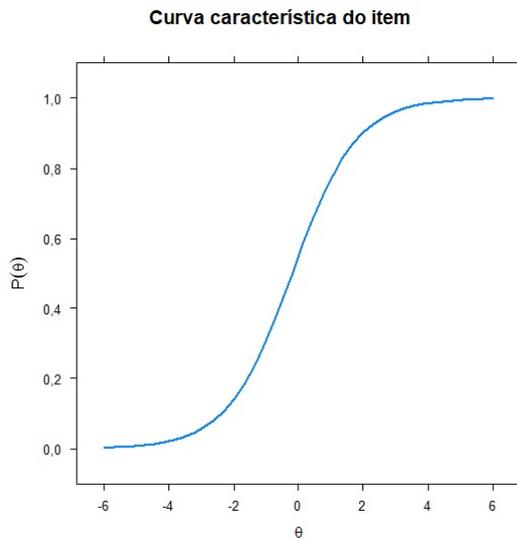
Análise do item 11

Item 11. Simplifique: $\frac{0,04 \cdot 0,01}{0,004}$

- (a) 0,01
- (b)
- (c) 0,001
- (d) 1
- (e) não sei



Tópico: Expressões algébricas: operações e valor numérico.



Análise do item 12

Item 12. Fatore: $2(3x+1)^7 - 16x(3x+1)^6$

(a) $2(3x+1)^6(1+11x)$

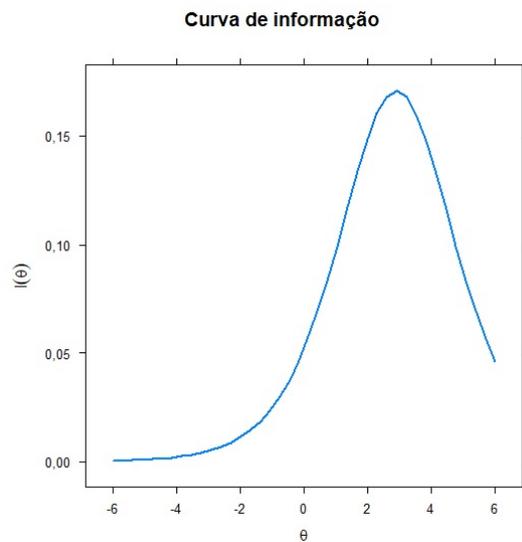
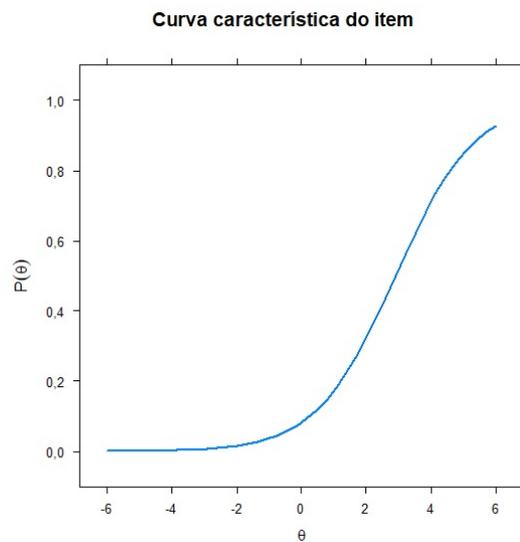
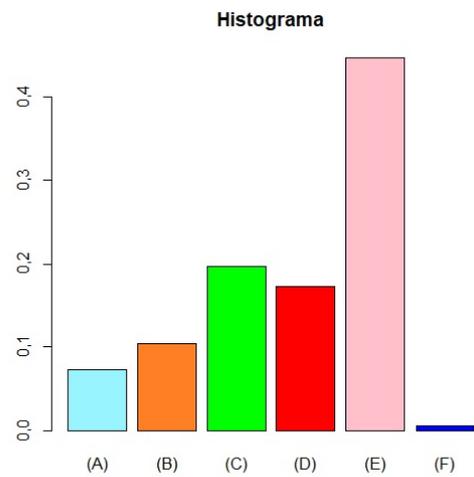
(b) $2(3x+1)^6(1-5x)$

(c) $2(3x+1)^6(1-4x)$

(d) $2(3x+1)^6(1-11x)$

(e) não sei

Tópico: Expressões algébricas: operações e valor numérico.



Análise do item 13

Item 13. Simplifique: $\frac{-3xy^2 + 15x^2 - 12y^2}{-3xy^2}$

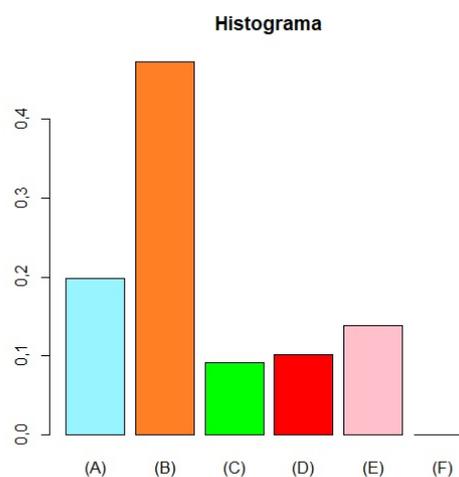
(a) $1 + 15x^2y - 12y^2$

(b) $1 - \frac{5x}{y^2} + \frac{4}{x}$

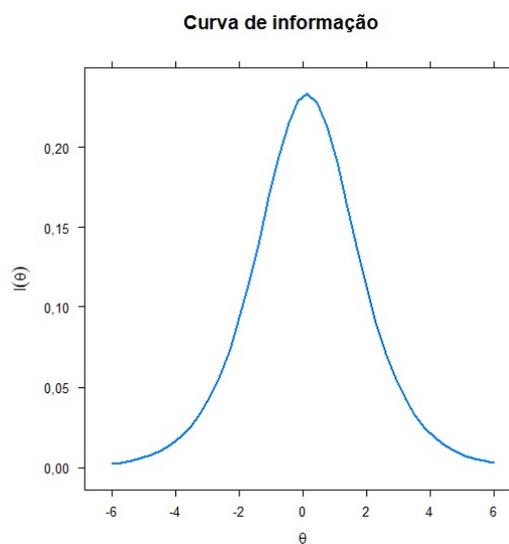
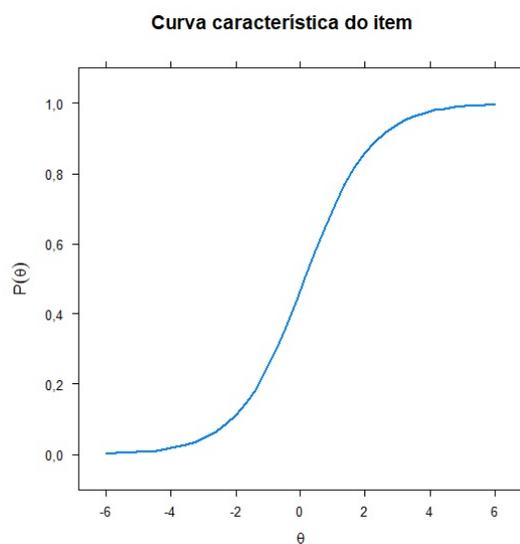
(c) $1 + \frac{5x}{y} + \frac{4}{x}$

(d) $1 + \frac{5x}{y} - \frac{4}{x}$

(e) não sei



Tópico: Fatoração e simplificação de expressões algébricas.



Análise do item 14

Item 14. Simplifique: $\frac{x^2 + 2x - 8}{x^2 - 4}$

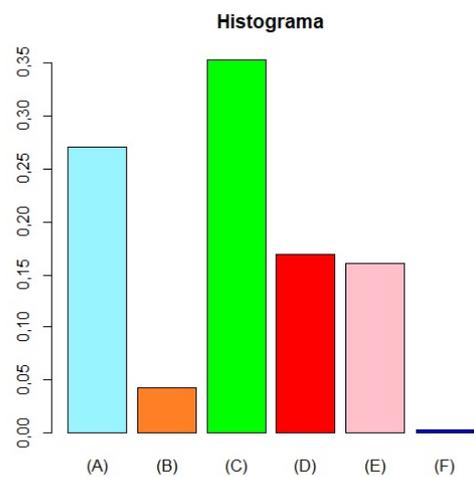
(a) $\frac{x - 4}{x - 2}$

(b) $\frac{x + 6}{x + 2}$

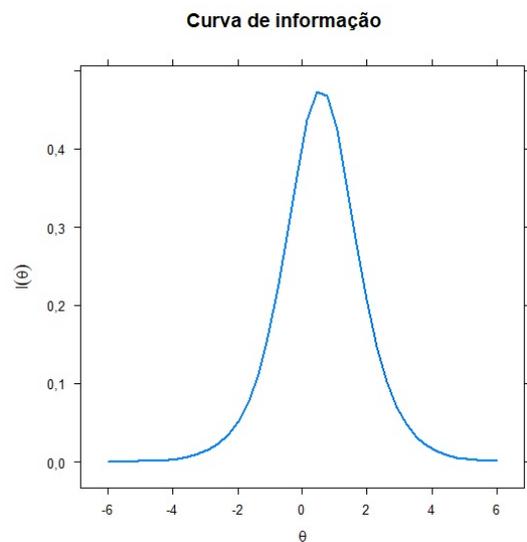
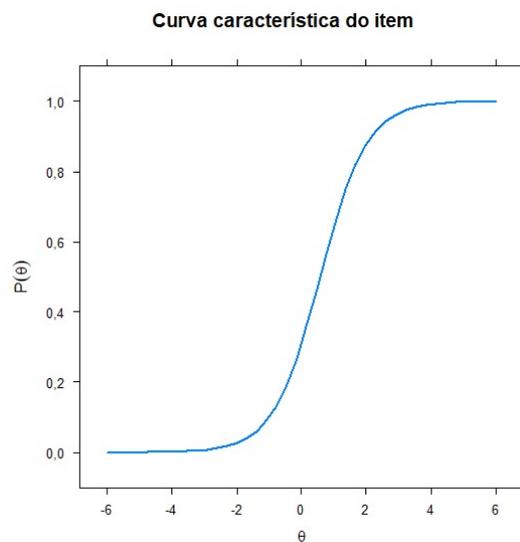
(c) $\frac{x + 4}{x + 2}$

(d) $1 + 2x + 2$

(e) não sei



Tópico: Fatoração e simplificação de expressões algébricas.



Análise do item 15

Item 15. Se $x \neq 9$, uma expressão equivalente à $\frac{\sqrt{x}-3}{x-9}$ é:

(a) $\frac{1}{\sqrt{x}+3}$

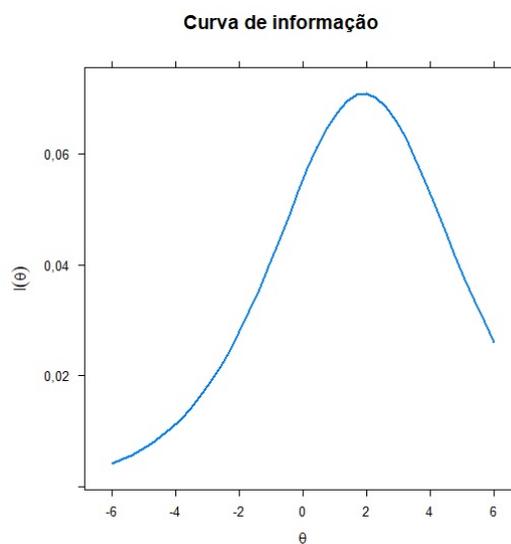
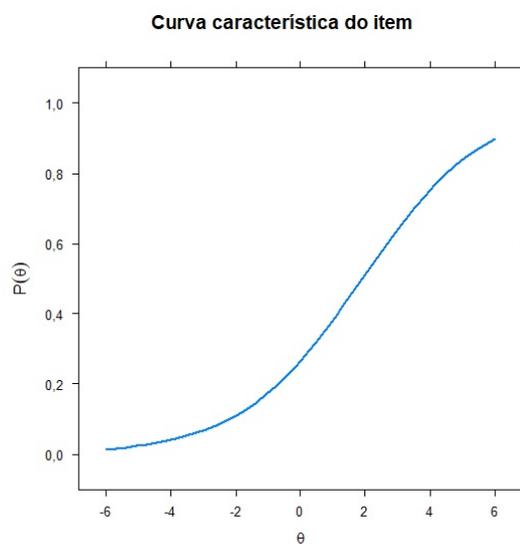
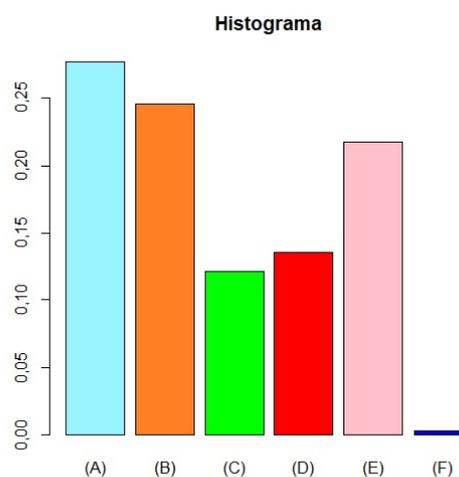
(b) $\frac{1}{\sqrt{x}-3}$

(c) $\frac{1}{x+3}$

(d) $\frac{1}{x-3}$

(e) não sei

Tópico: Fatoração e simplificação de expressões algébricas.



Análise do item 16

Item 16. Simplifique: $\frac{1}{1-x} - \frac{1}{2x-2}$

(a) $\frac{3}{2x+2}$

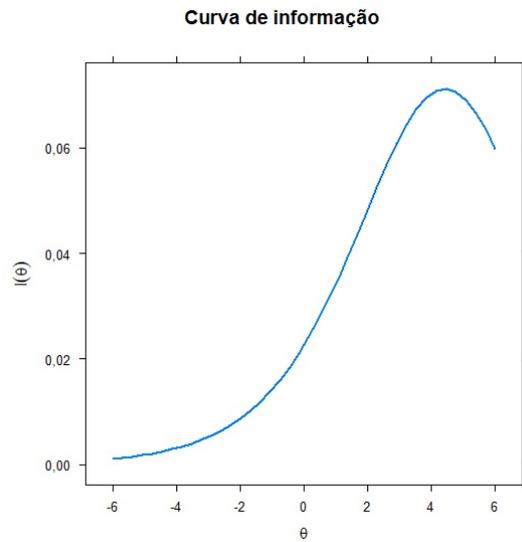
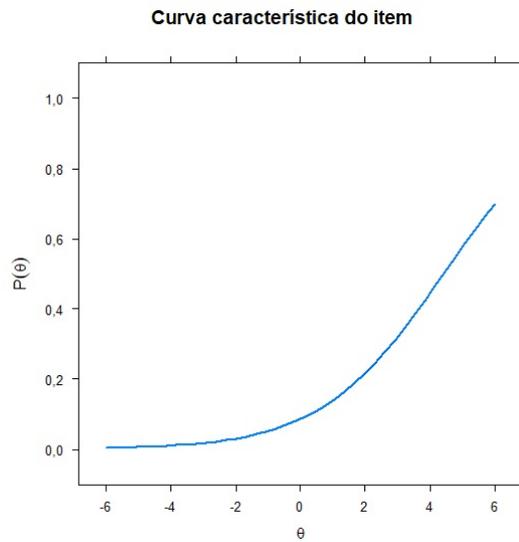
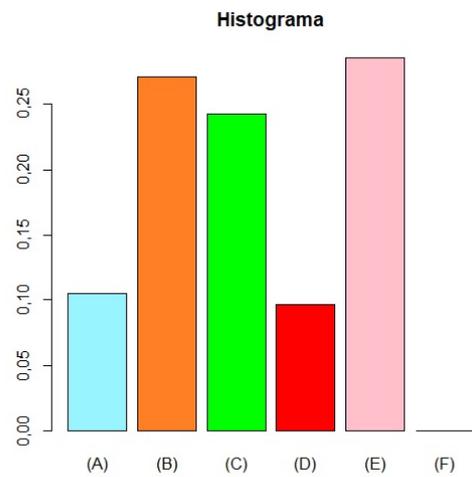
(b) $-\frac{3}{2x+2}$

(c) $\frac{3}{2x-2}$

(d) $\frac{3}{2-2x}$

(e) não sei

Tópico: Fatoração e simplificação de expressões algébricas.

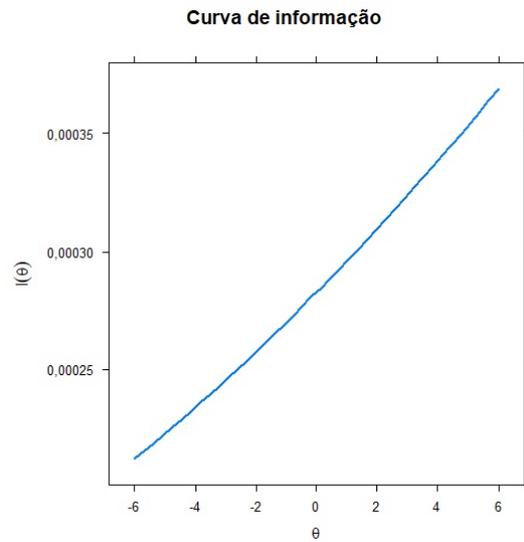
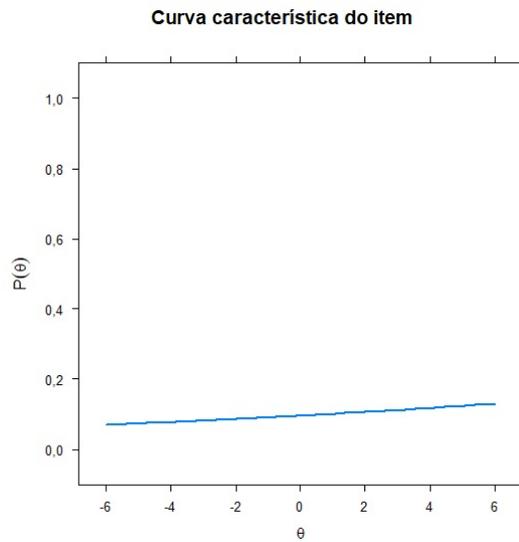
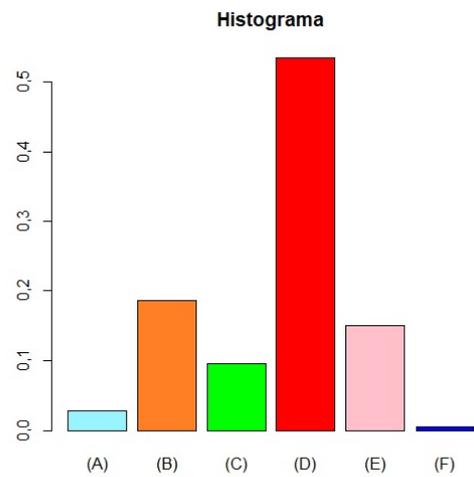


Análise do item 17

Item 17. Um dos fatores irredutíveis da expressão polinomial $81t^4 - 625$:

- (a) $9t + 5$
- (b) $9t - 5$
- (c) $9t^2 + 25$
- (d) $9t^2 - 25$
- (e) não sei

Tópico: Fatoração e simplificação de expressões algébricas.



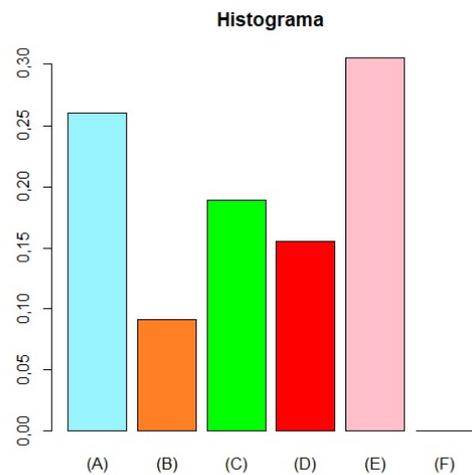
Análise do item 18

Item 18. Para $x < 0$, é equivalente à $\frac{\sqrt{x^2 + 1}}{x^2}$:

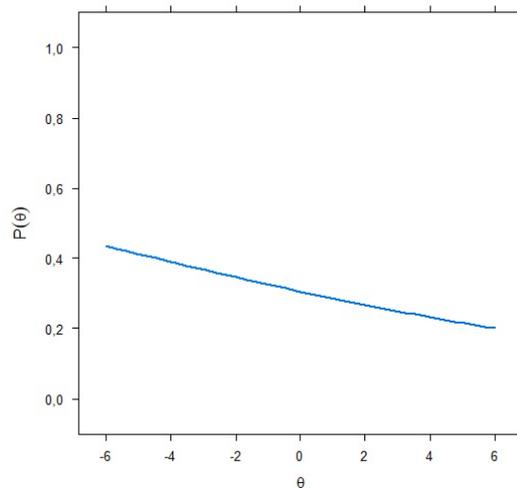
- (a) $\frac{1}{\sqrt{x^2 + 1}}$
- (b) $-\sqrt{1 + x^{-2}}$
- (c) $\sqrt{1 + x^{-2}}$
- (d) $-\frac{1}{\sqrt{x^2 + 1}}$
- (e) não sei

Tópico: Fatoração e simplificação de expressões algébricas.

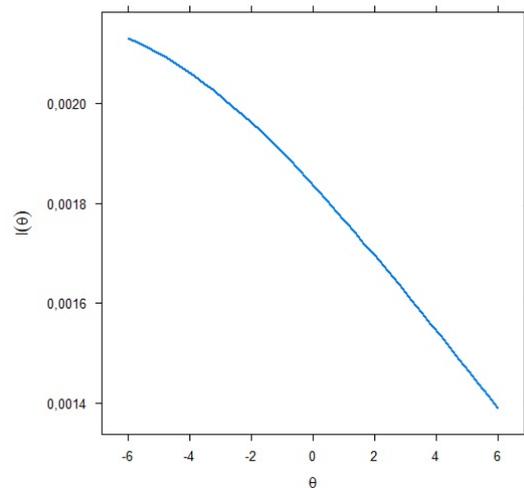
Observação: Embora o gabarito aponte para a opção (E), o item não apresenta opção correta.



Curva característica do item



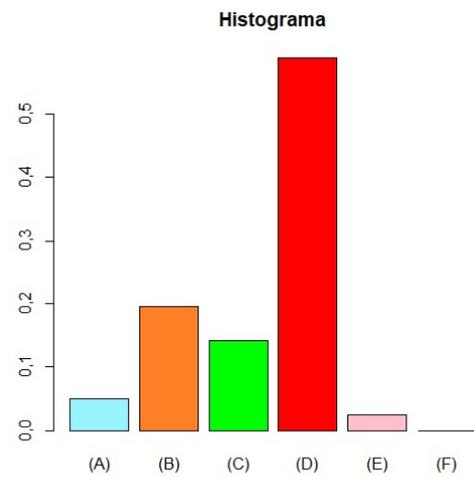
Curva de informação



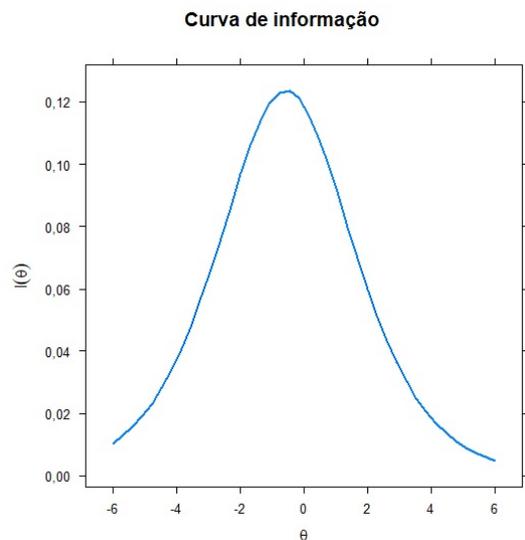
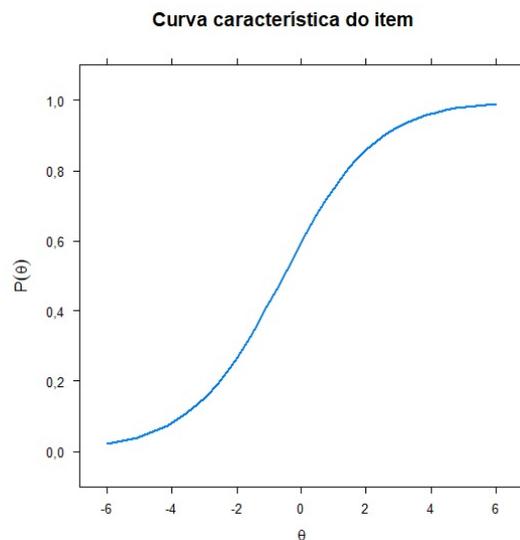
Análise do item 19

Item 19. É equivalente a afirmação: “A distância entre x e 3 é pelo menos igual a 2”

- (a) $|x - 2| \geq 3$
- (b) $|x + 3| \geq 2$
- (c) $|x + 2| \geq 3$
- (d) $|x - 3| \geq 2$
- (e) não sei



Tópico: Valor absoluto e propriedades.



Análise do item 20

Item 20. A soma das soluções da equação

$$x + |2x - 3| = 2 \text{ é:}$$

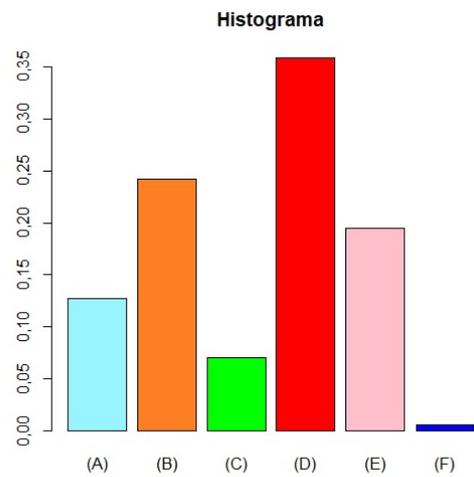
(a) $\frac{4}{3}$

(b) $\frac{8}{3}$

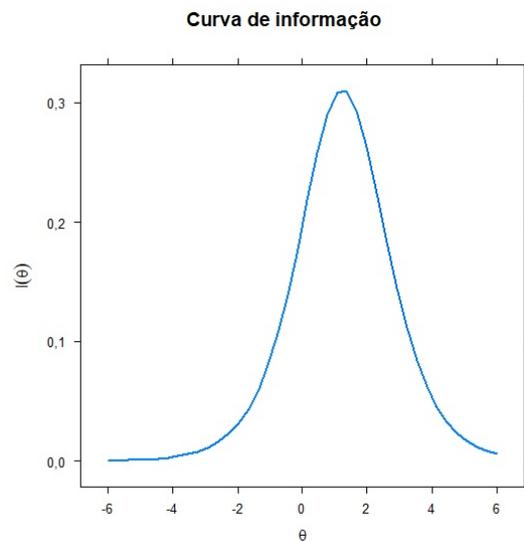
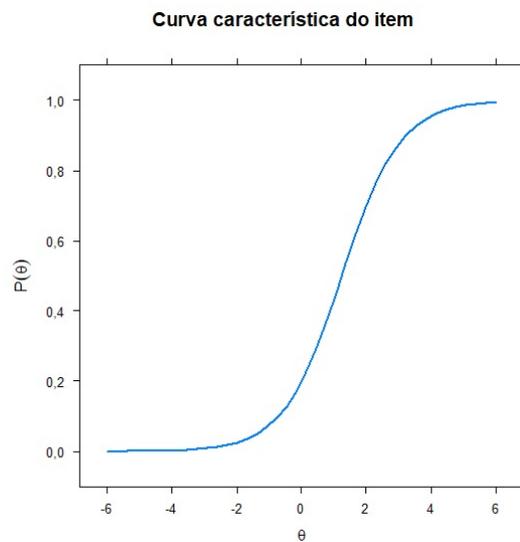
(c) $\frac{7}{3}$

(d) $\frac{5}{3}$

(e) não sei



Tópico: Equações e inequações modulares.

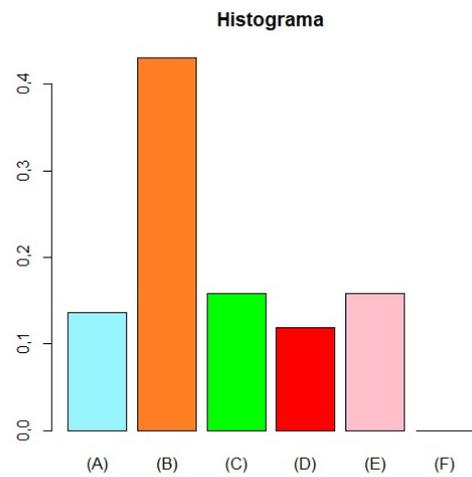


Análise do item 21

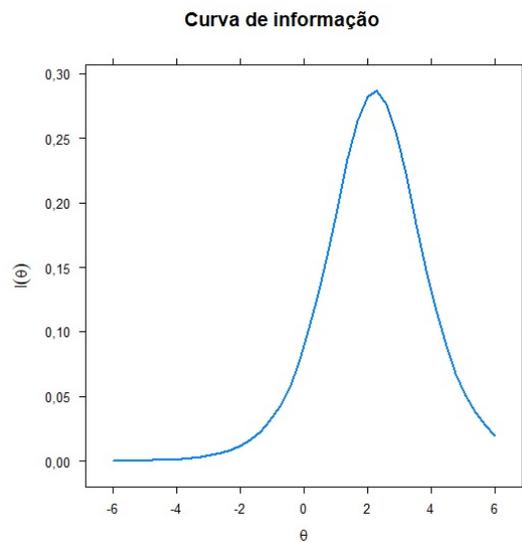
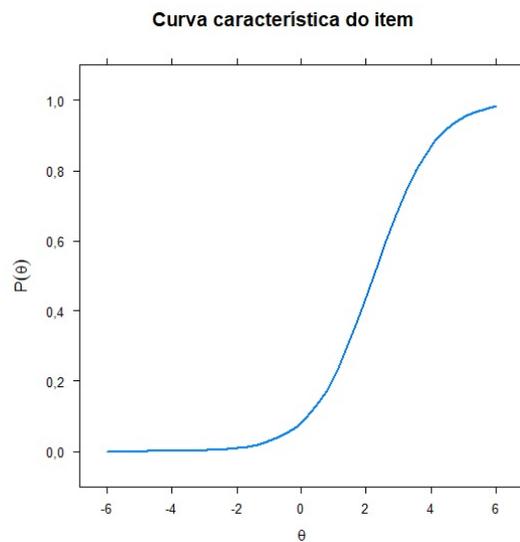
Item 21. O conjunto solução da inequação

$$\frac{2x + 1}{x + 1} > 2 \text{ é:}$$

- (a) $(-\infty, 1)$
- (b) $(1, \infty)$
- (c) $(-1, \infty)$
- (d) $(-\infty, -1)$
- (e) não sei



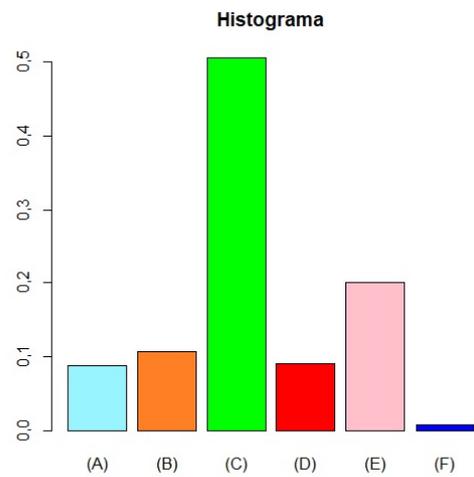
Tópico: Equações e inequações racionais.



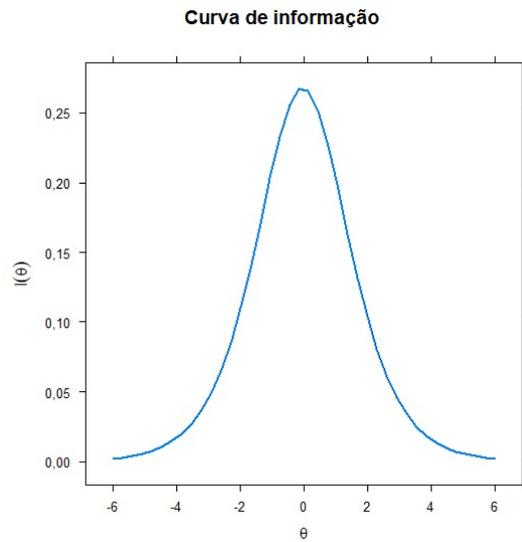
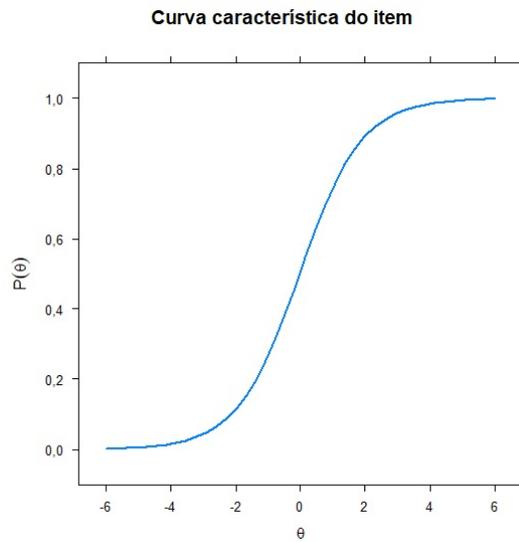
Análise do item 22

Item 22. Determine a equação da reta que passa pelos pontos $P(-1, 1)$ e $Q(1, 5)$.

- (a) $y = x + 2$
- (b) $y = 2x + 1$
- (c) $y = 2x + 3$
- (d) $y = 3x + 2$
- (e) não sei



Tópico: Retas no plano e coeficiente angular.



Análise do item 23

Item 23. Encontre o domínio de: $f(x) = \frac{1}{\sqrt{1+x}}$.

(a) $x \neq -1$

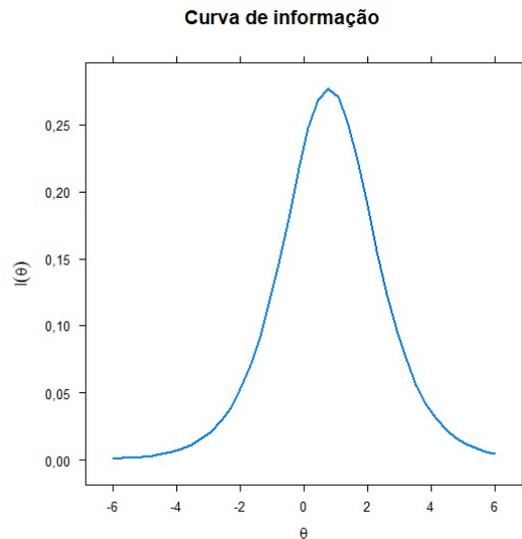
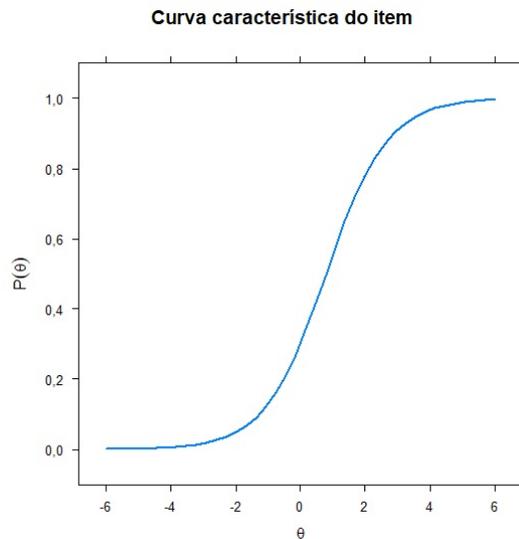
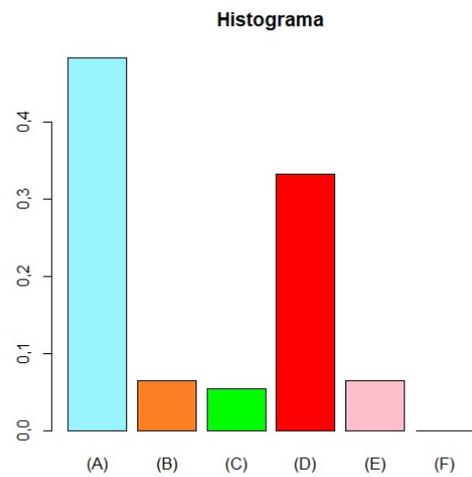
(b) $(-\infty, 1]$

(c) $x \neq 0$

(d) $(-1, \infty)$

(e) não sei

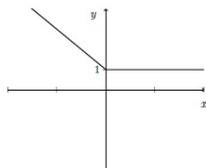
Tópico: Funções: definição. Domínio e imagem.



Análise do item 24

Item 24. Seja o gráfico de uma função real

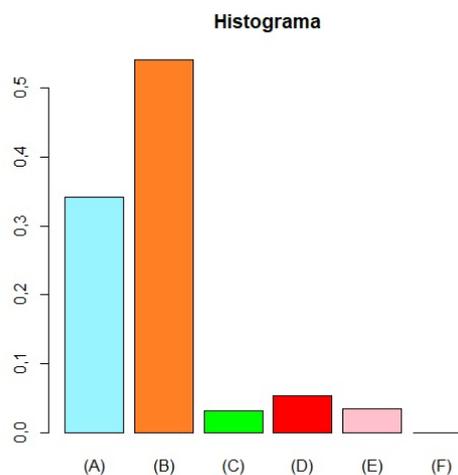
f .



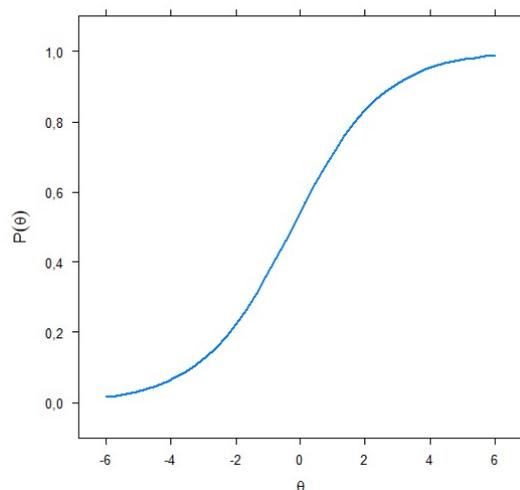
Com base neste gráfico, podemos dizer que o conjunto imagem de f é:

- (a) \mathbb{R}
- (b) $[1, +\infty)$
- (c) $[-1, +\infty)$
- (d) $[0, +\infty)$
- (e) não sei

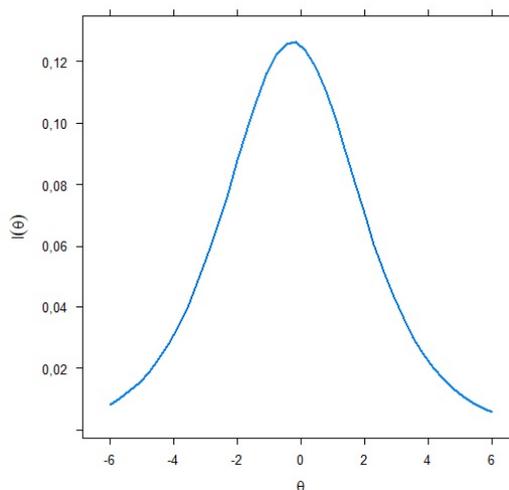
Tópico: Funções: definição. Domínio e imagem.



Curva característica do item



Curva de informação

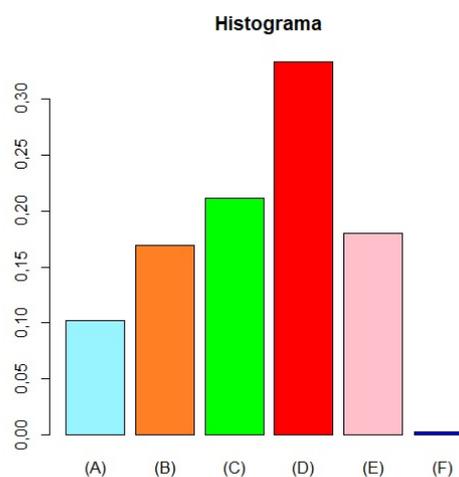


Análise do item 25

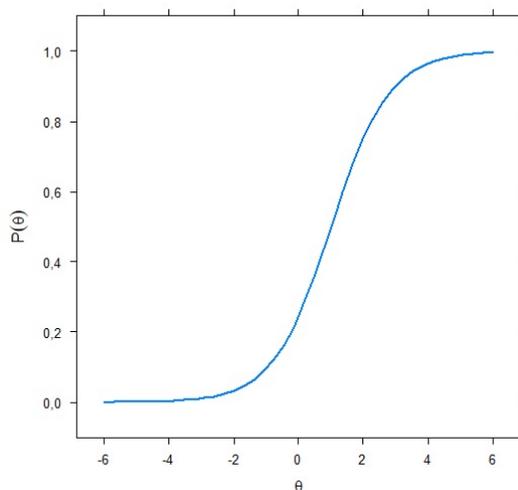
Item 25. Os pontos de interseção de $f(x) = \frac{x+4}{x-2}$ com os eixos x e y são, respectivamente:

- (a) $(-4, 0)$ e $(-2, 0)$
- (b) $(4, 0)$ e $(0, 2)$
- (c) $(0, -4)$ e $(2, 0)$
- (d) $(-4, 0)$ e $(0, -2)$
- (e) não sei

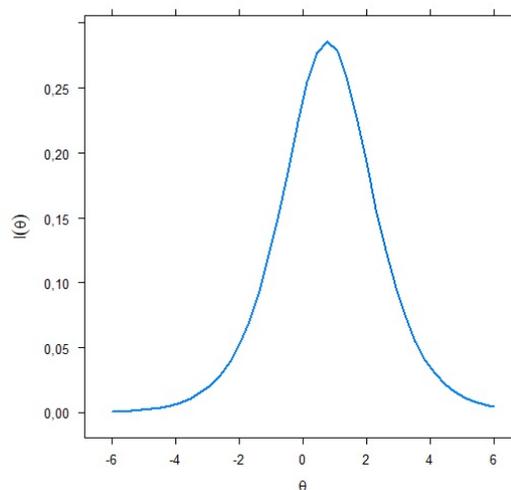
Tópico: Funções: gráficos. Interseção com os eixos coordenados.



Curva característica do item



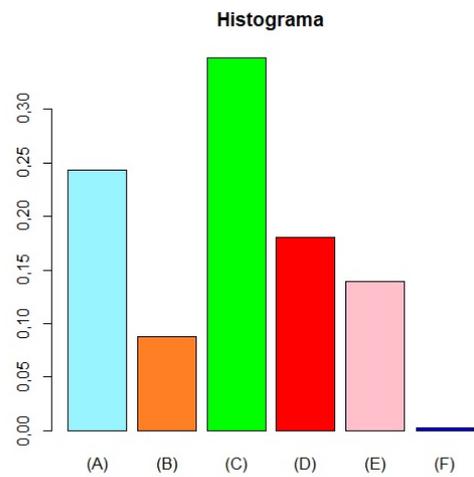
Curva de informação



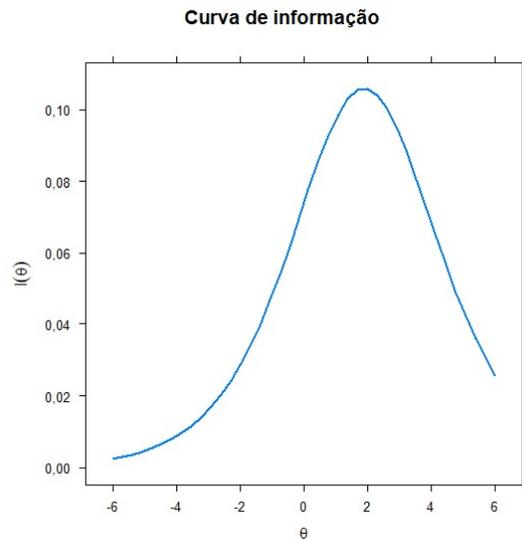
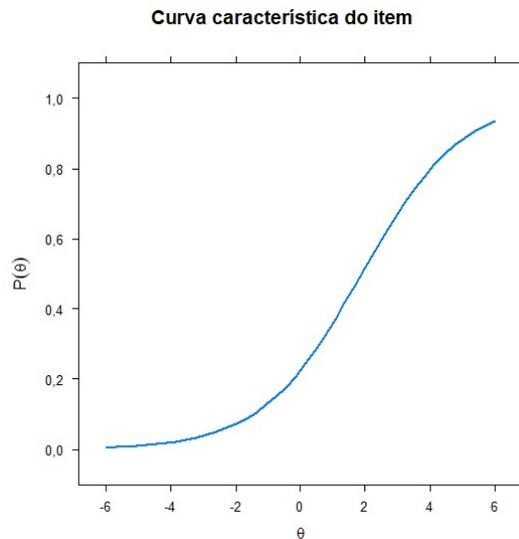
Análise do item 26

Item 26. A abscissa do ponto máximo de $f(x) = x^2$ no intervalo $[-3, 2]$ é:

- (a) $x = -3$
- (b) $x = 0$
- (c) $x = 9$
- (d) $x = 2$
- (e) não sei

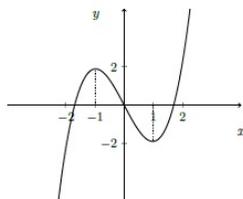


Tópico: Máximos e mínimos locais e globais.



Análise do item 27

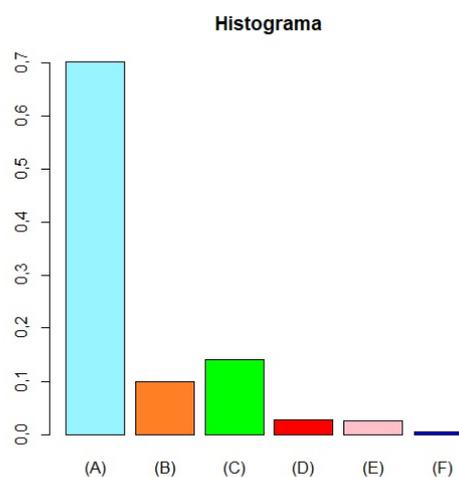
Item 27. Seja f uma função real cujo gráfico é exibido abaixo:



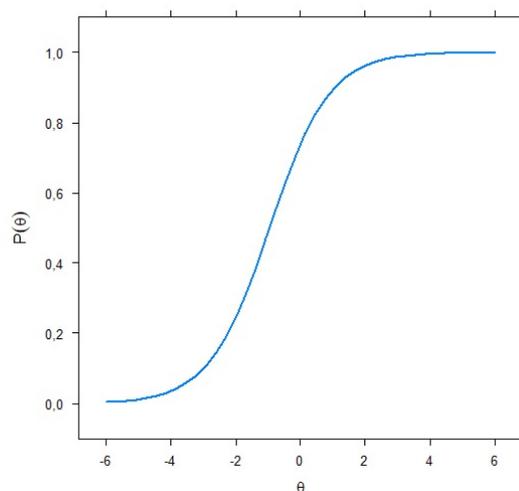
Com base neste gráfico, pode-se afirmar que:

- (a) f é decrescente em $(-1, 1)$
- (b) f é crescente em $(0, +\infty)$
- (c) f é decrescente em $(-\infty, -1]$
- (d) f é sempre crescente
- (e) não sei

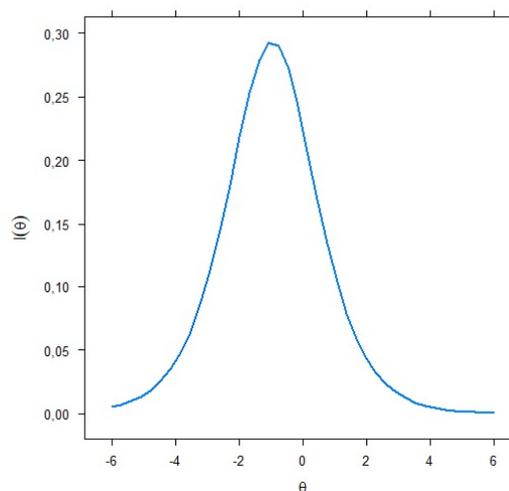
Tópico: Intervalos de crescimento e decrescimento.



Curva característica do item



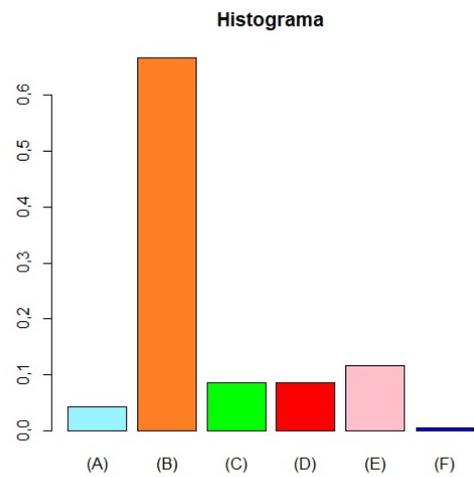
Curva de informação



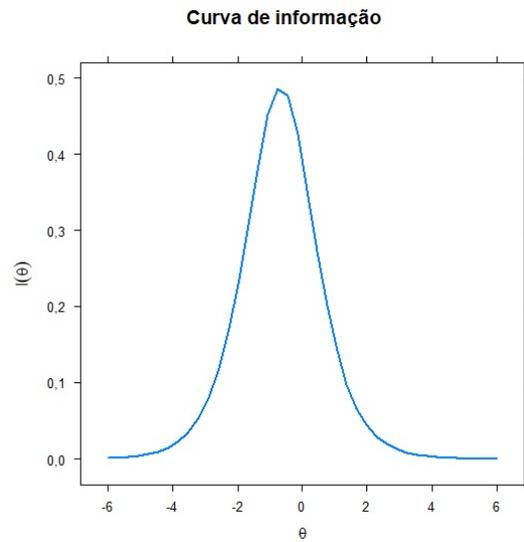
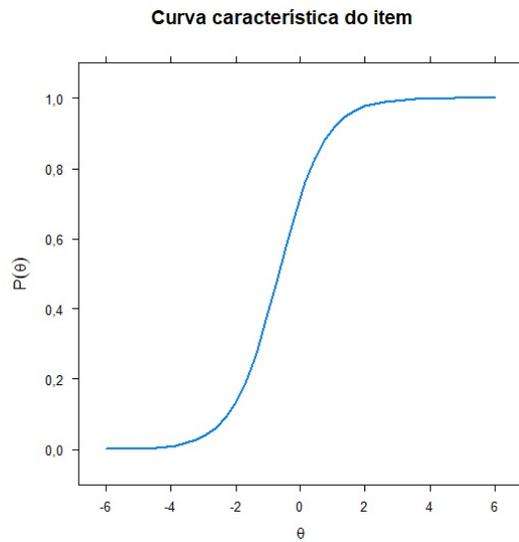
Análise do item 28

Item 28. Se $f(x) = x^2 + 1$ e $g(x) = 2x - 1$,
então $f(g(2))$ é:

- (a) 11
- (b)
- (c) 9
- (d) 5
- (e) não sei



Tópico: Composição de funções.

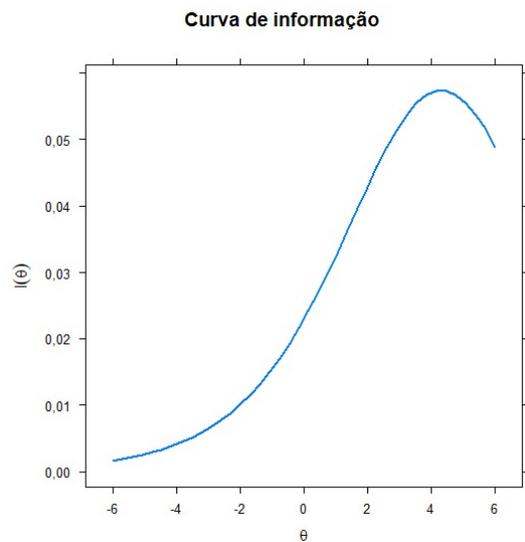
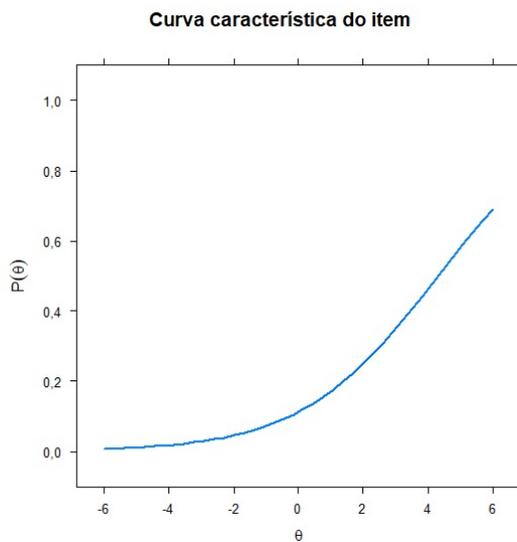
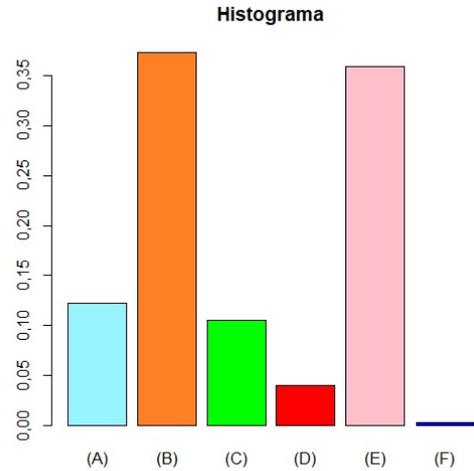


Análise do item 29

Item 29. Sejam as funções f e g funções tais que $g(x) = f(x + 2) + 3$. Então o gráfico de g pode ser obtido do gráfico de f por:

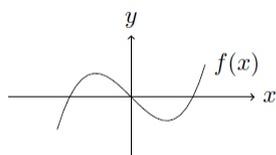
- (a) translação horizontal de 2 unidades para esquerda e translação vertical de 3 unidades para cima.
- (b) translação horizontal de 2 unidades para direita e translação vertical de 3 unidades para cima.
- (c) translação horizontal de 2 unidades para esquerda e translação vertical de 3 unidades para baixo.
- (d) translação horizontal de 2 unidades para direita e translação vertical de 3 unidades para baixo.
- (e) não sei

Tópico: Transformação de funções: translação, rotação, reflexão, contração e expansão.

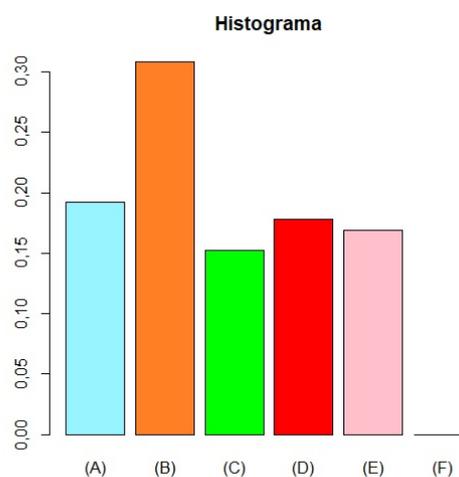


Análise do item 30

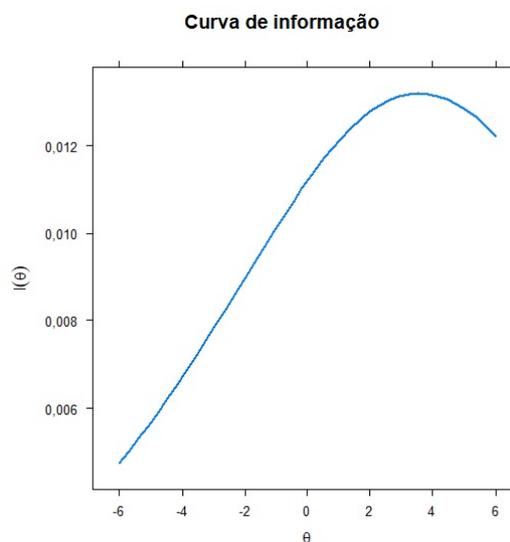
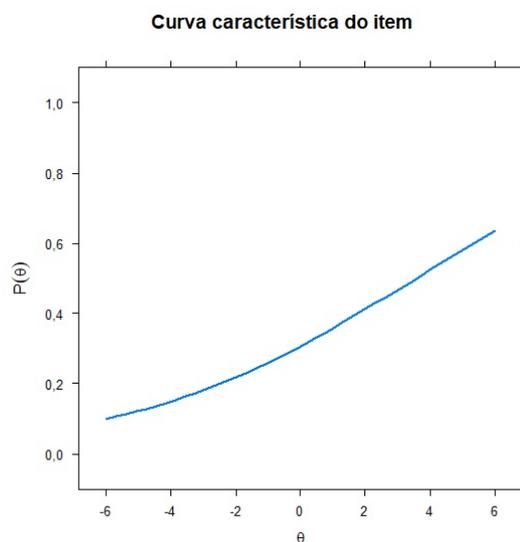
Item 30. O gráfico abaixo pode representar a função polinomial f definida por



- (a) $f(x) = x^2 - x$
- (b) $f(x) = x^3 - x$
- (c) $f(x) = x^4 - x^3$
- (d) $f(x) = x^4 - x$
- (e) não sei



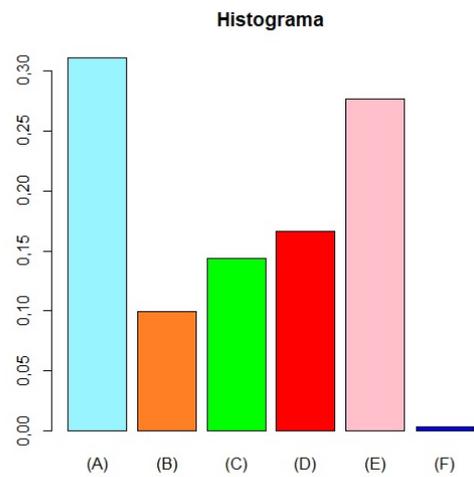
Tópico: Polinômios e funções polinomiais:
definição e propriedades.



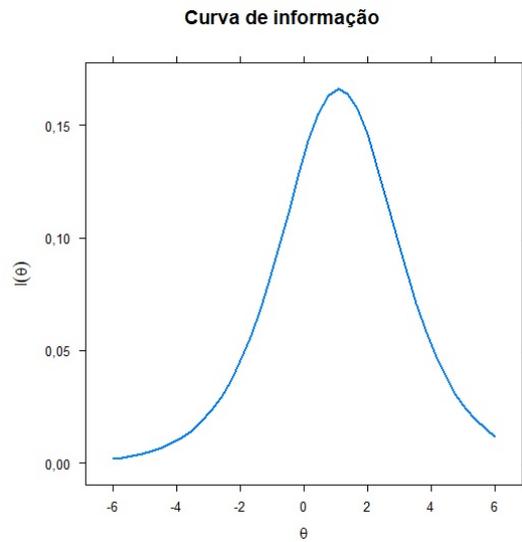
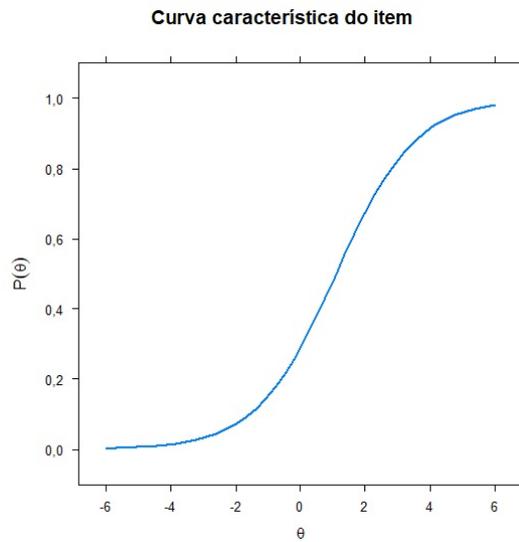
Análise do item 31

Item 31. O resto da divisão de $3x^3 + 3x^2 - 20x - 17$ por $4 - x^2$ é:

- (a) $-8x - 5$
- (b) $-9x - 4$
- (c) $-7x - 2$
- (d) $-6x - 8$
- (e) não sei



Tópico: Divisão de polinômios.



Análise do item 32

Item 32. Determine x que satisfaça

$$2(2x - 3) + 5(x + 1) = 6x - 7.$$

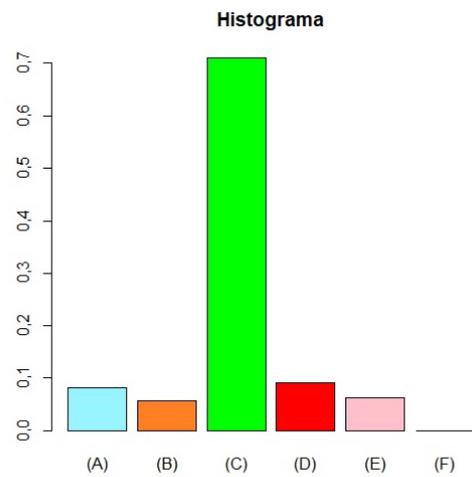
(a) $x = 2$

(b) $x = 4$

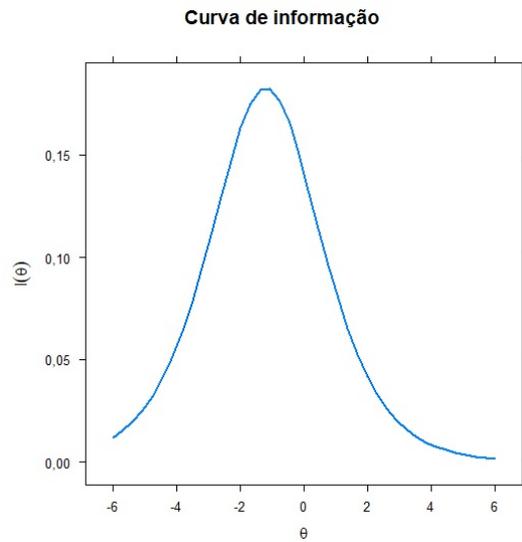
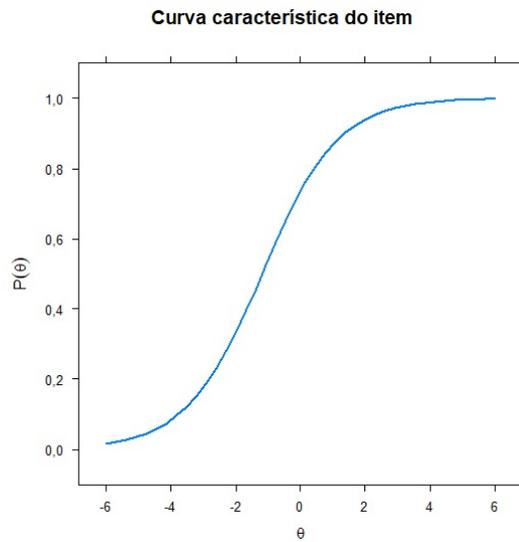
(c) $x = -2$

(d) $x = -4$

(e) não sei



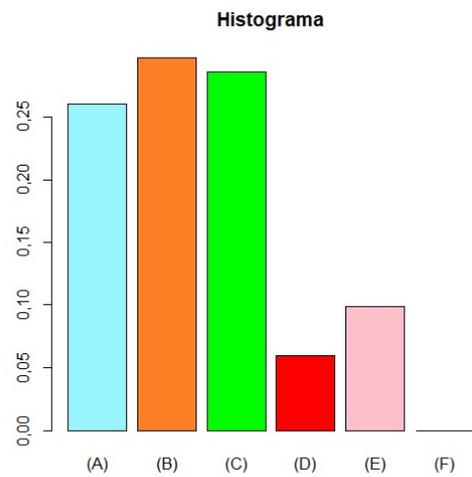
Tópico: Equações polinomiais.



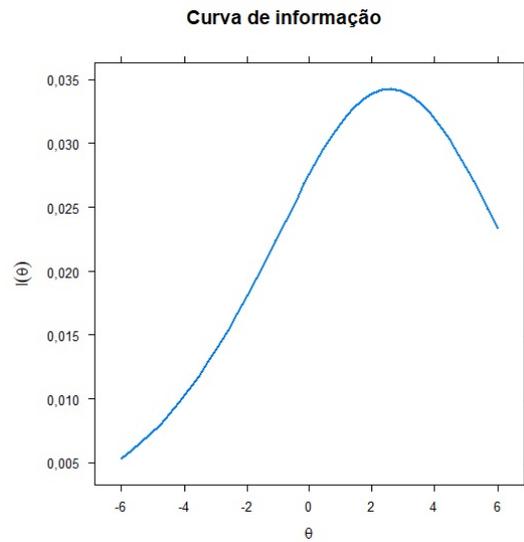
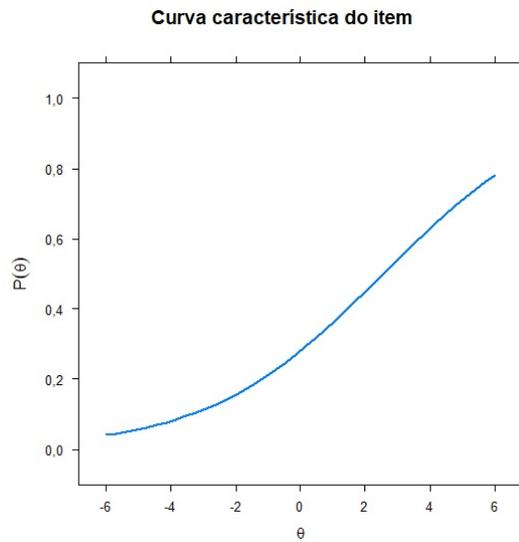
Análise do item 33

Item 33. A expressão $(x + 1)(x + 2) > 0$ é equivalente à:

- (a) $x > -1$
- (b) $-2 < x < -1$
- (c) $x < -2$ ou $x > -1$
- (d) $x < -2$
- (e) não sei



Tópico: Inequações polinomiais.



Análise do item 34

Item 34. Determine A e B tais que $\frac{A}{x-1}$

$$+ \frac{B}{x+2} = \frac{1}{(x-1)(x+2)}.$$

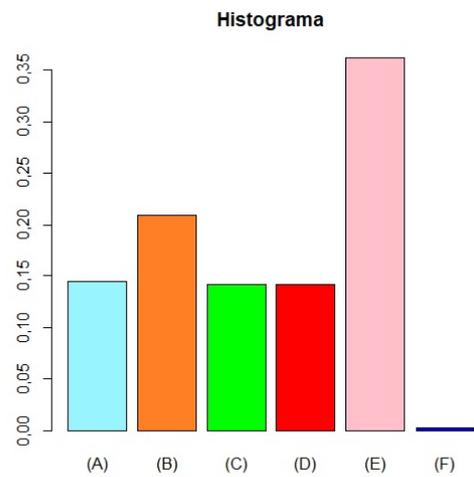
(a) $A = -\frac{1}{3}, B = \frac{1}{3}$

(b) $A = \frac{1}{3}, B = -\frac{1}{3}$

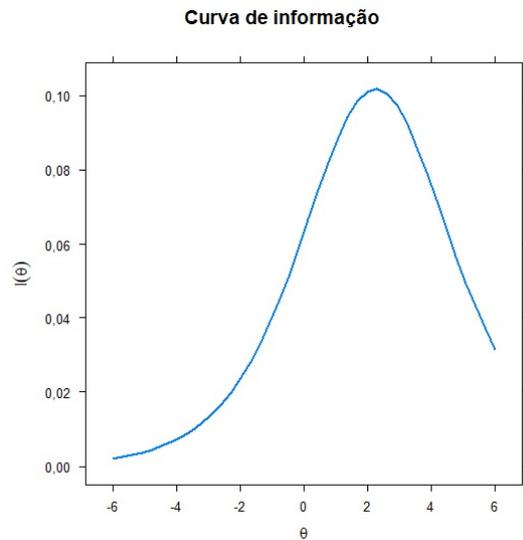
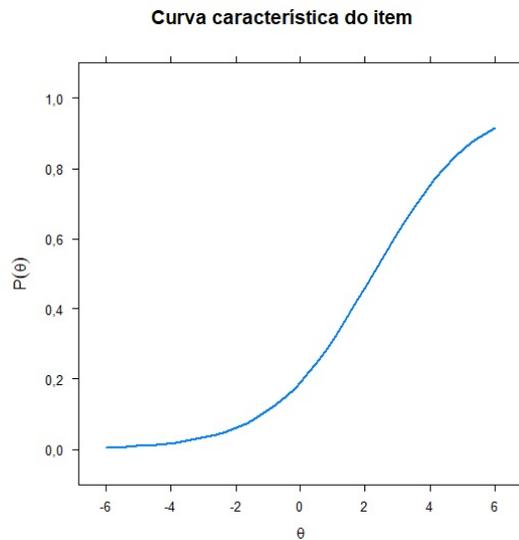
(c) $A = -\frac{1}{2}, B = -\frac{1}{3}$

(d) $A = \frac{1}{4}, B = -\frac{1}{3}$

(e) não sei



Tópico: MMC e MDC de polinômios. Decomposição em frações parciais.



Análise do item 35

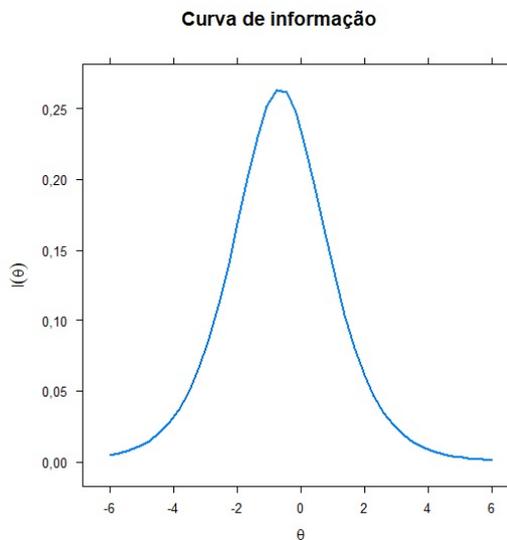
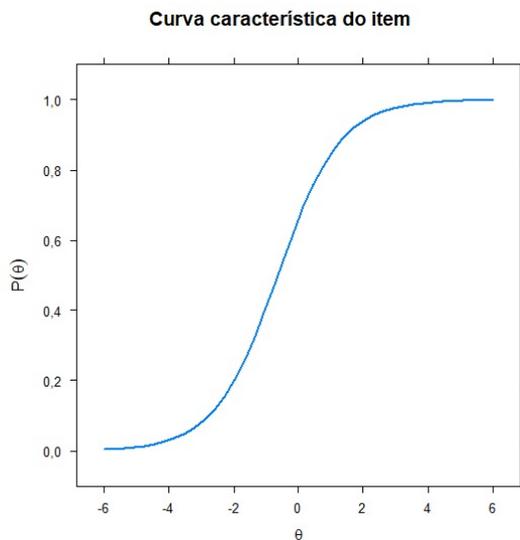
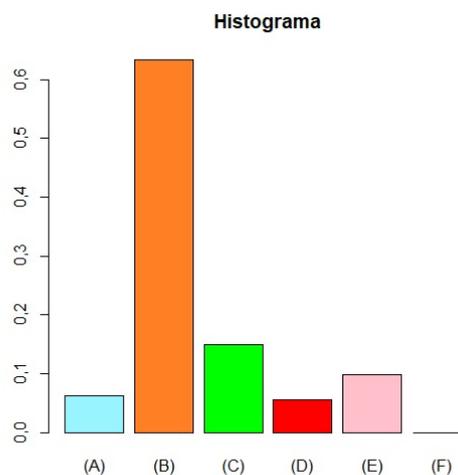
Item 35. Seja a função real f definida por:

$$f(x) = \begin{cases} 1 - x^2 & \text{se } x < 2 \\ -x & \text{se } x \geq 2 \end{cases}$$

O valor de $f(-1) + f(0) + f(3)$ é:

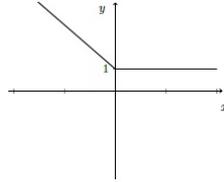
- (a) 1
- (b)
- (c) 0
- (d) -1
- (e) não sei

Tópico: Funções definidas por várias sentenças.



Análise do item 36

Item 36. Selecione a função que melhor se aproxima do gráfico:



(a)
$$f(x) = \begin{cases} -x + 1 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0 \end{cases}$$

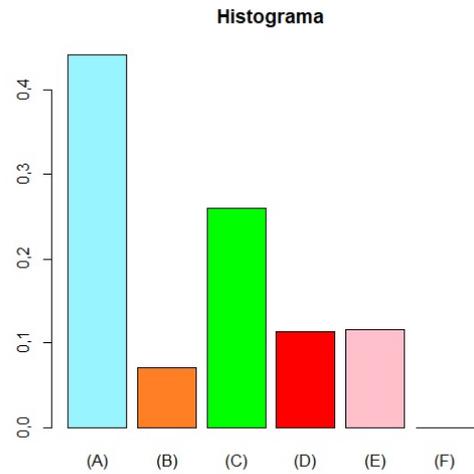
(b)
$$f(x) = \begin{cases} 1 & \text{se } x < 0 \\ -x + 1 & \text{se } x \geq 0 \end{cases}$$

(c)
$$f(x) = \begin{cases} x + 1 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0 \end{cases}$$

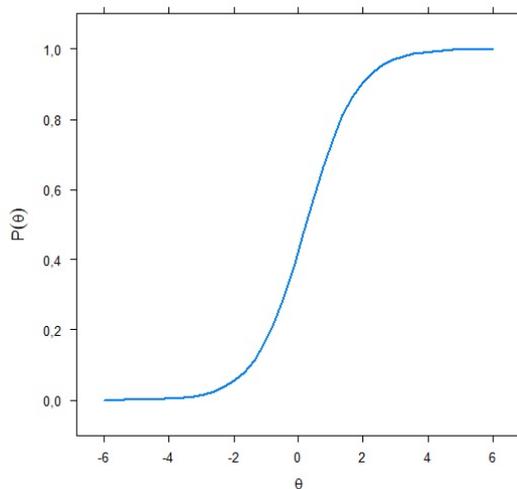
(d)
$$f(x) = \begin{cases} x & \text{se } x < 0 \\ x + 1 & \text{se } x \geq 0 \end{cases}$$

(e) não sei

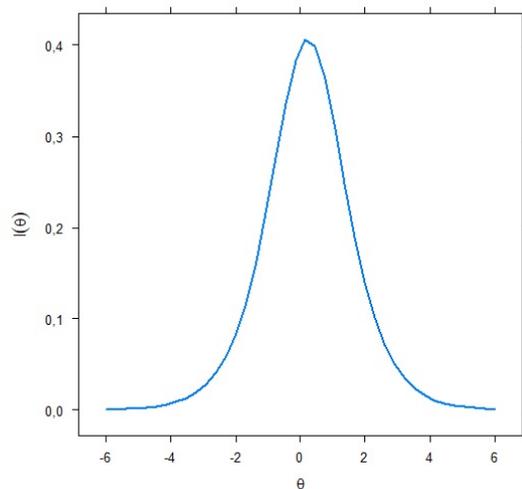
Tópico: Funções definidas por várias sentenças.



Curva característica do item

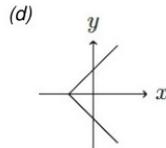
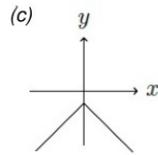
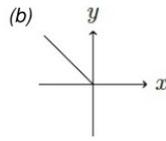
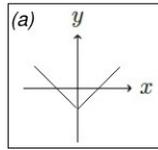


Curva de informação

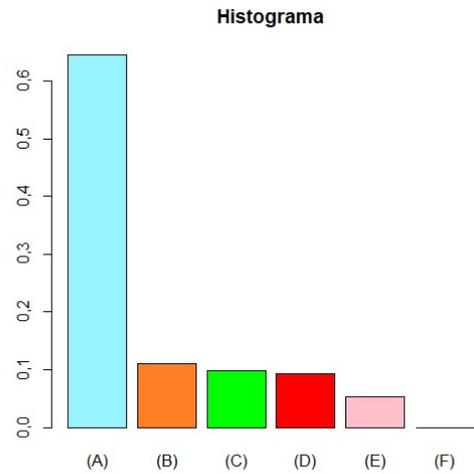


Análise do item 37

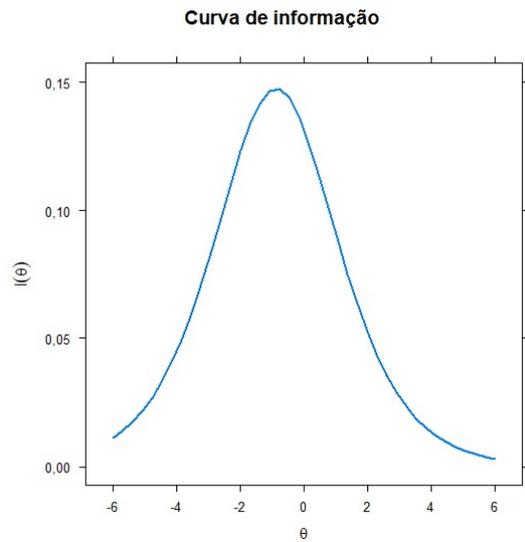
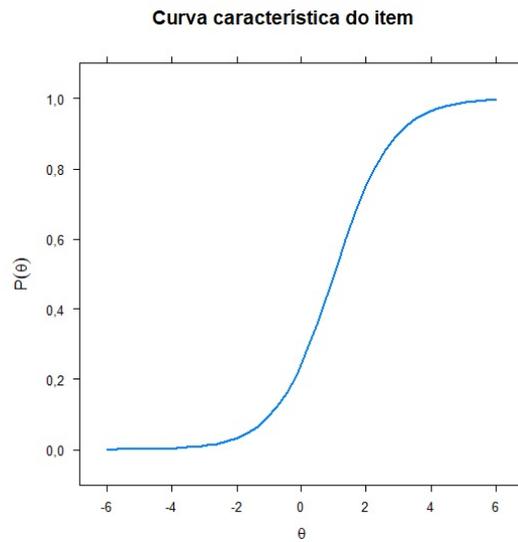
Item 37. O gráfico que melhor representa a função real f definida por $f(x) = |x| - 1$.



(e) *não sei*



Tópico: Função modular.



Análise do item 38

Item 38. Se $f(x) = \frac{1}{x}$, calcule:

$$\frac{f(x) - f(2)}{x - 2}.$$

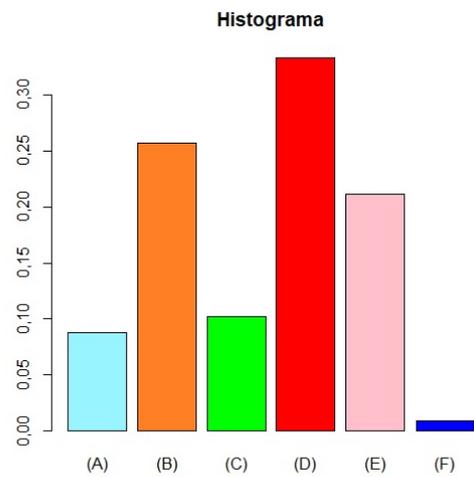
(a) $-\frac{1}{x}$

(b) $\frac{1}{2x}$

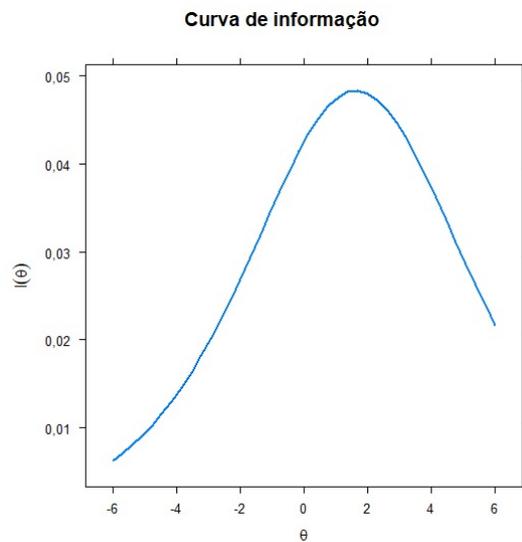
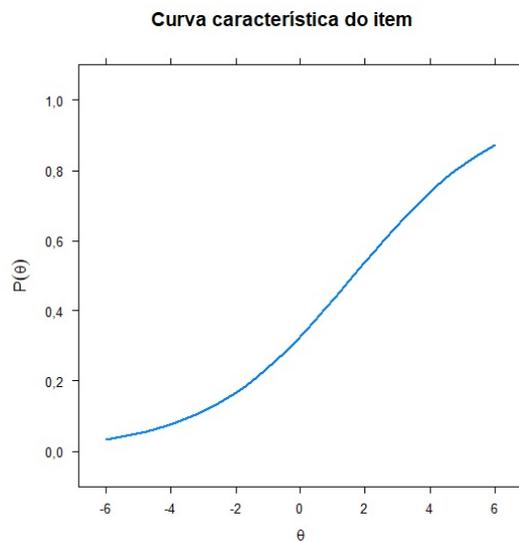
(c) $\frac{1}{x}$

(d) $-\frac{1}{2x}$

(e) não sei



Tópico: Taxa de variação média.

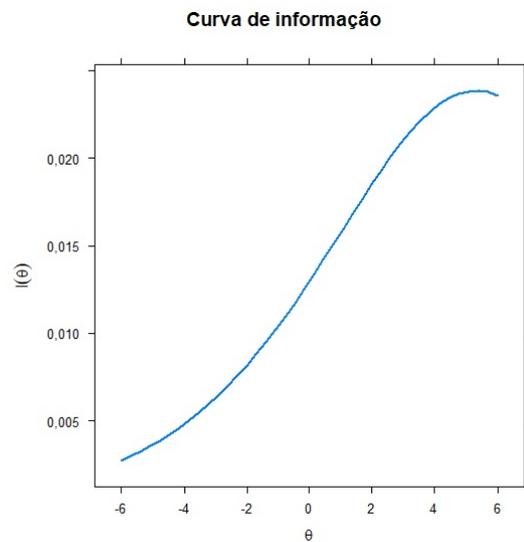
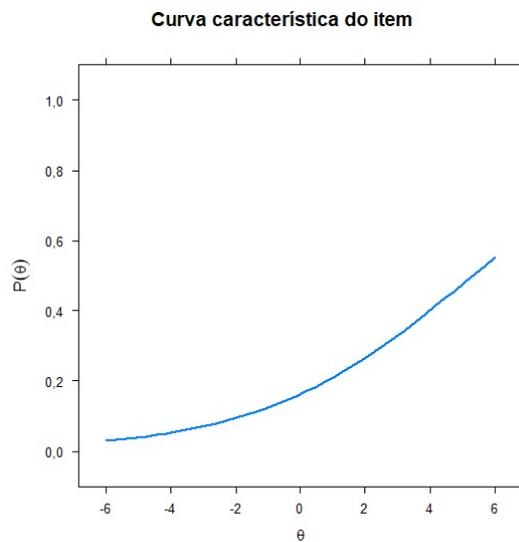
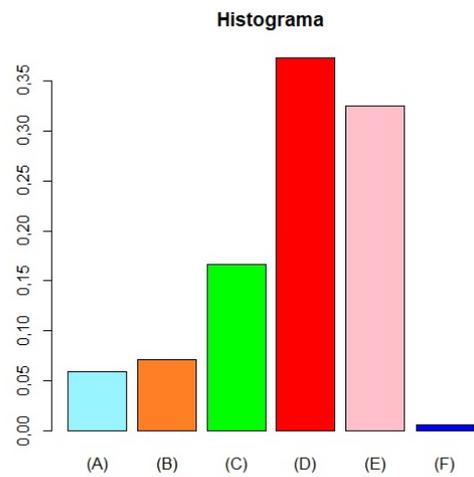


Análise do item 39

Item 39. Selecione o intervalo em que a função real f tal que $f(x) = (x + 1)^2 + 1$ é injetora.

- (a) $(-\infty, 1)$
- (b) \mathbb{R}_-
- (c) $[-1, \infty)$
- (d) \mathbb{R}
- (e) não sei

Tópico: Função sobrejetora, injetora e bijetora.



Análise do item 40

Item 40. Se $f(x) = \frac{x+3}{x}$, então:

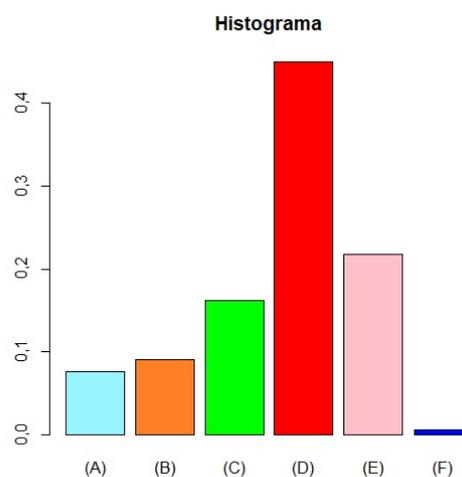
(a) $f^{-1}(x) = \frac{3}{3x-1}$

(b) $f^{-1}(x) = \frac{x-1}{3}$

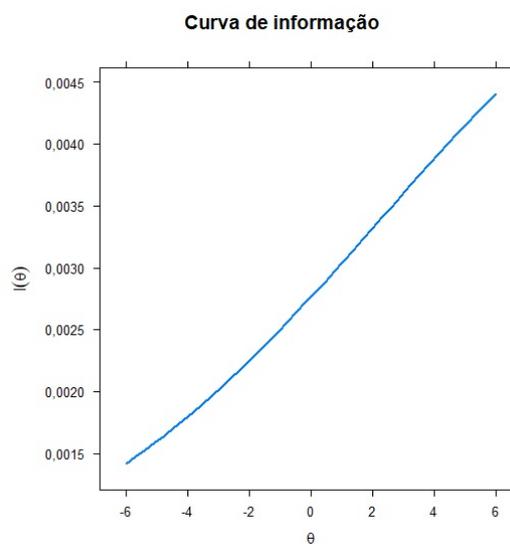
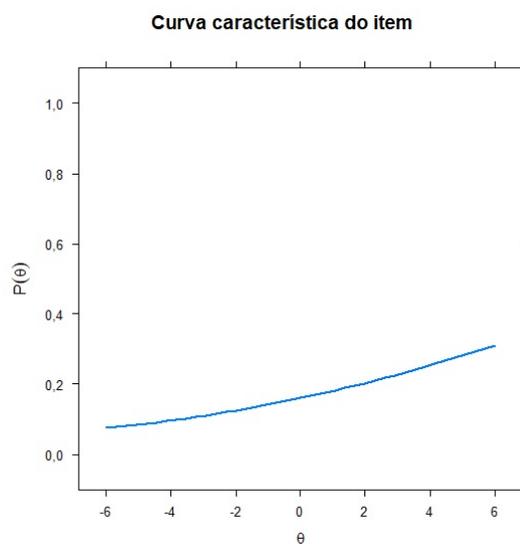
(c) $f^{-1}(x) = \frac{3}{x-1}$

(d) $f^{-1}(x) = \frac{x}{x+3}$

(e) não sei

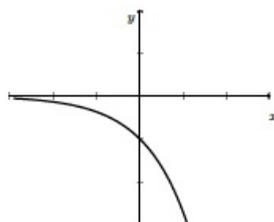


Tópico: Funções Inversa: Propriedades.

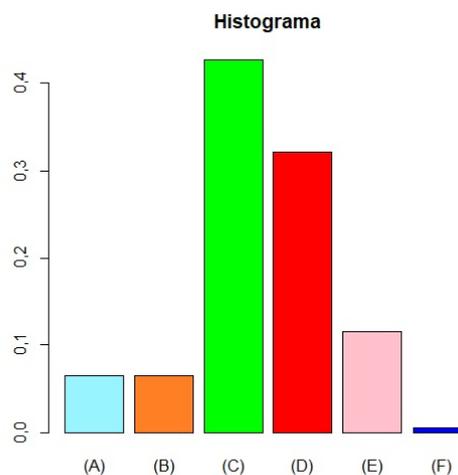


Análise do item 41

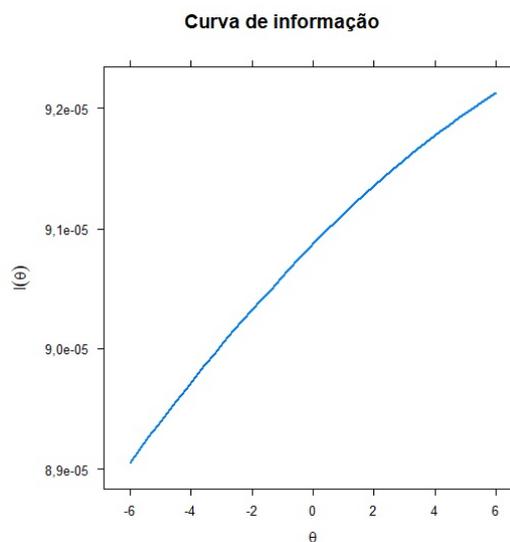
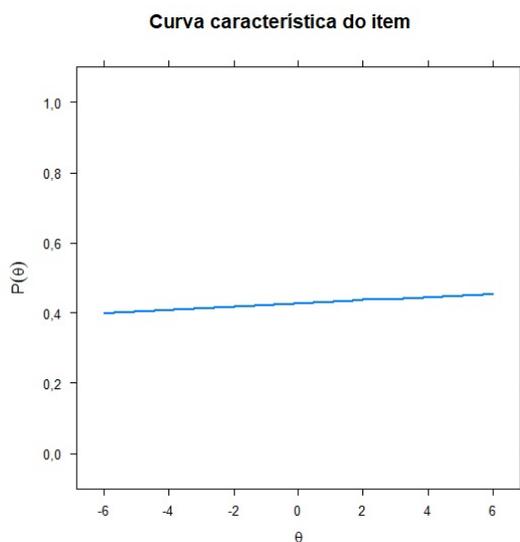
Item 41. Dentre as opções abaixo, determine a função exponencial f que melhor se aproxima do gráfico:



- (a) $f(x) = \left(\frac{1}{2}\right)^x$
- (b) $f(x) = 2^x$
- (c) $f(x) = -2^x$
- (d) $f(x) = -\left(\frac{1}{2}\right)^x$
- (e) não sei



Tópico: Função exponencial: definição e propriedades.



Análise do item 42

Item 42. Determine o domínio da relação

$$y = \log_x(3 - x).$$

(a) $x > 0, x \neq 1$

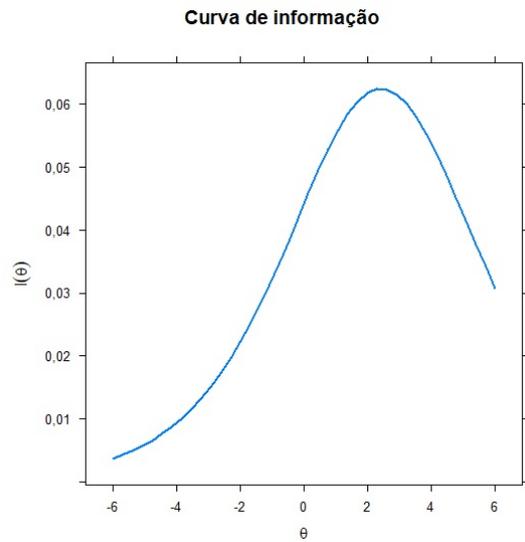
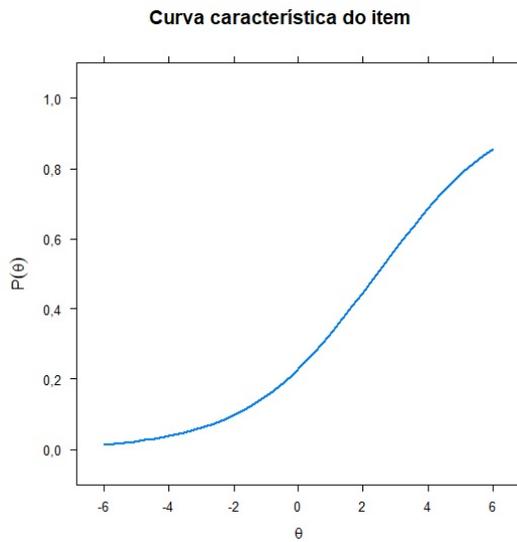
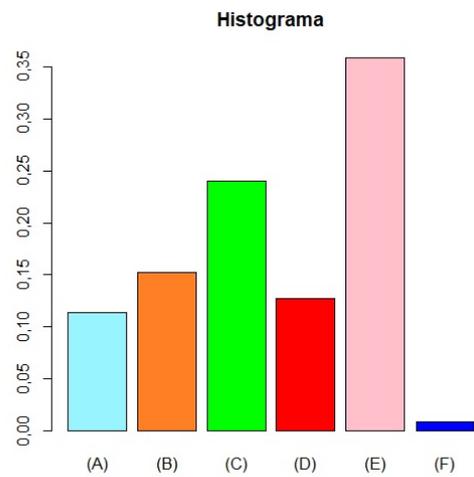
(b) $x > 3$

(c) $0 < x < 3, x \neq 1$

(d) $0 < x < 3$

(e) não sei

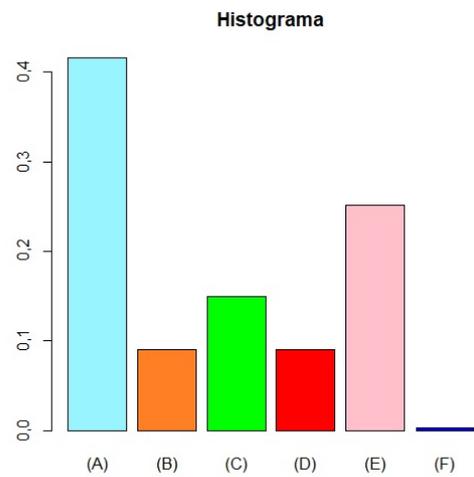
Tópico: Logarítmos: definição e propriedades. Mudança de base.



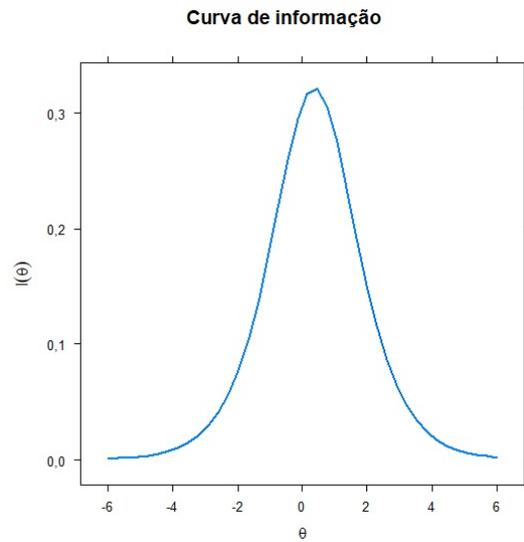
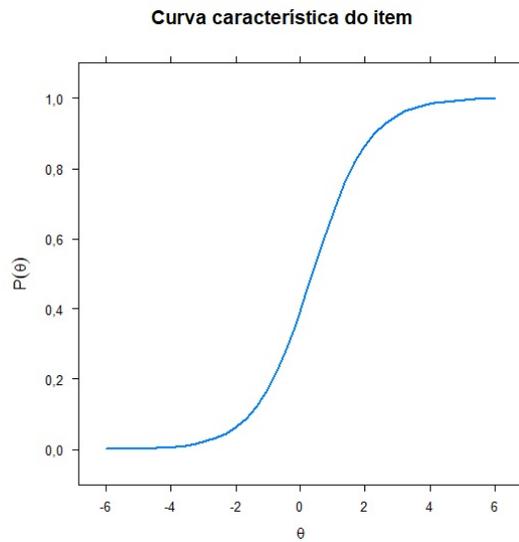
Análise do item 43

Item 43. Calcule: $\log_3(3^5)$

- (a) 5
- (b) 3
- (c) $\frac{1}{5}$
- (d) $\frac{1}{3}$
- (e) não sei



Tópico: Logaritmos: definição e propriedades. Mudança de base.



Análise do item 44

Item 44. Determine o conjunto solução da equação $5^{x^2} = 625$.

(a) $\{-4, 4\}$

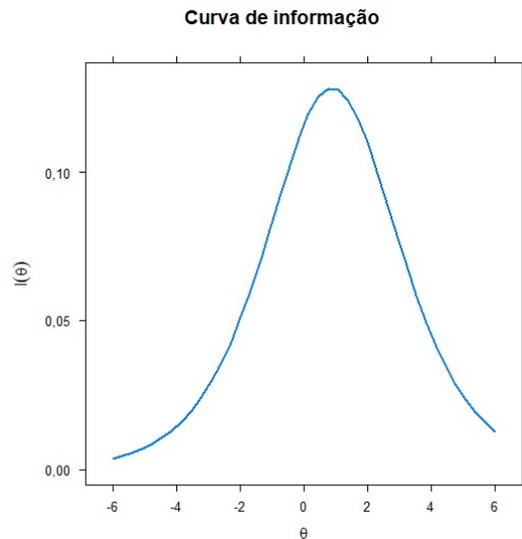
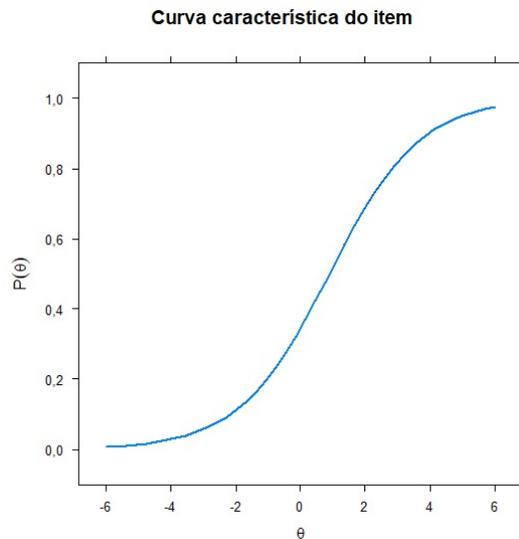
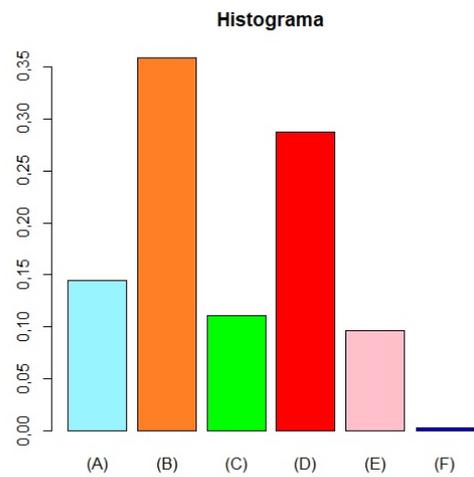
(b) $\{-2, 2\}$

(c) $\{4\}$

(d) $\{2\}$

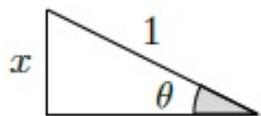
(e) não sei

Tópico: Inequações exponenciais e logarítmicas.



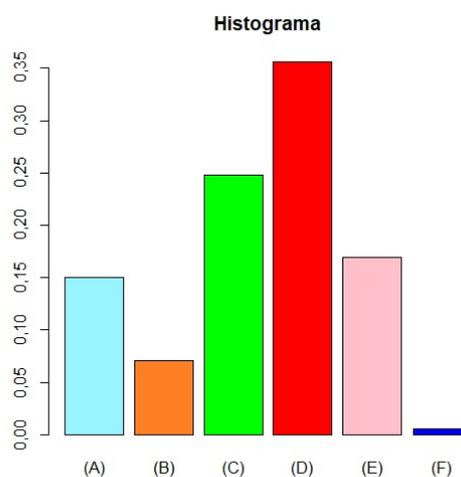
Análise do item 45

Item 45. Considere o triângulo retângulo:

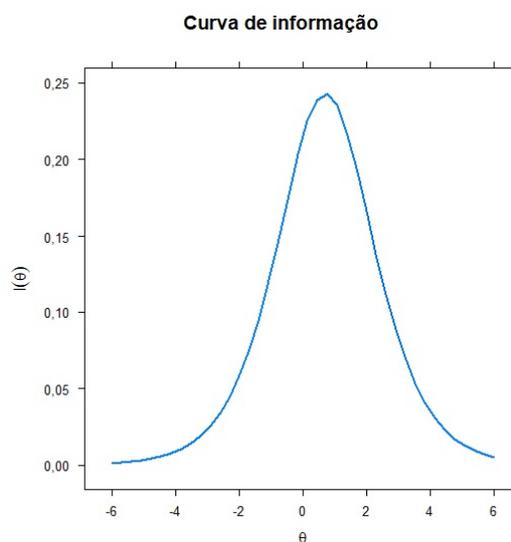
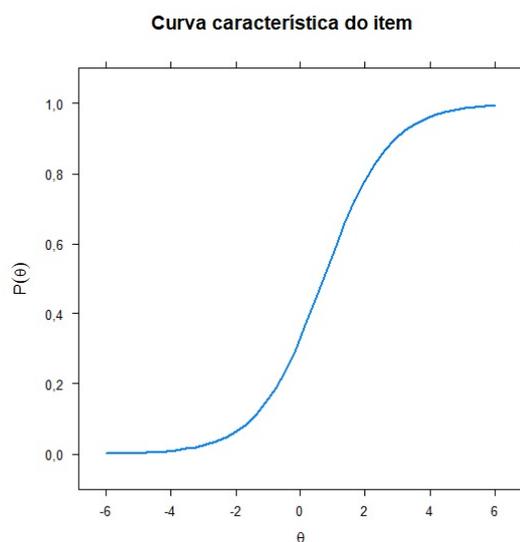


Então o valor de $tg \theta$ é:

- (a) x
- (b) $\frac{\sqrt{1-x^2}}{x}$
- (c) $\frac{1}{x}$
- (d) $\frac{x}{\sqrt{1-x^2}}$
- (e) não sei



Tópico: Relações trigonométricas no triângulo retângulo



Análise do item 46

Item 46. Se $\pi \leq \theta \leq \frac{3\pi}{2}$ e $\text{sen } \theta = -\frac{2}{3}$,
então $\text{cos } \theta$ é igual a:

(a) 1

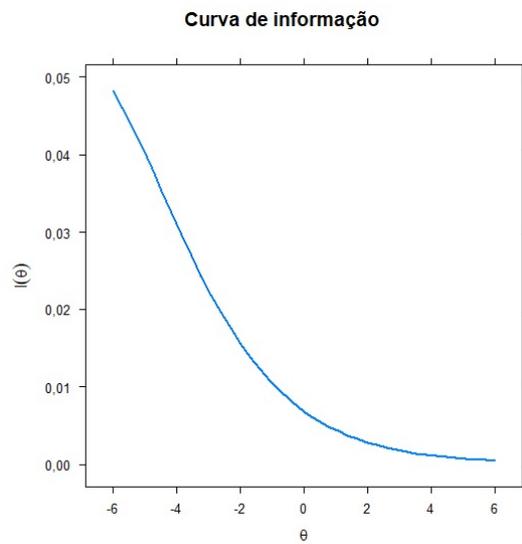
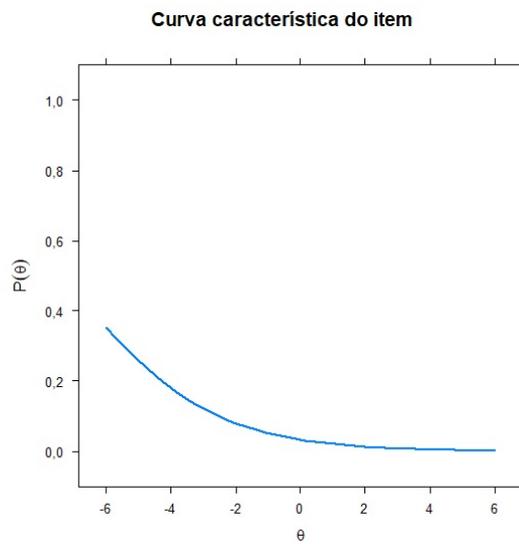
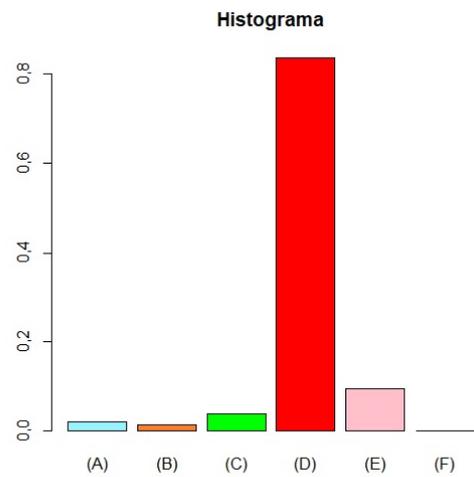
(b) $-\frac{1}{3}$

(c) $-\frac{\sqrt{5}}{3}$

(d) $-\frac{3}{5}$

(e) não sei

Tópico: O círculo trigonométrico.

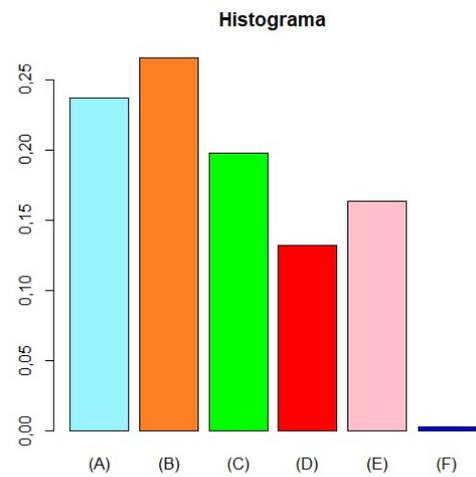


Análise do item 47

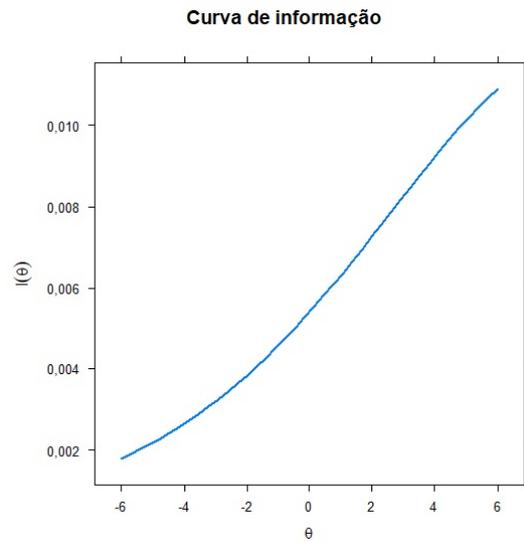
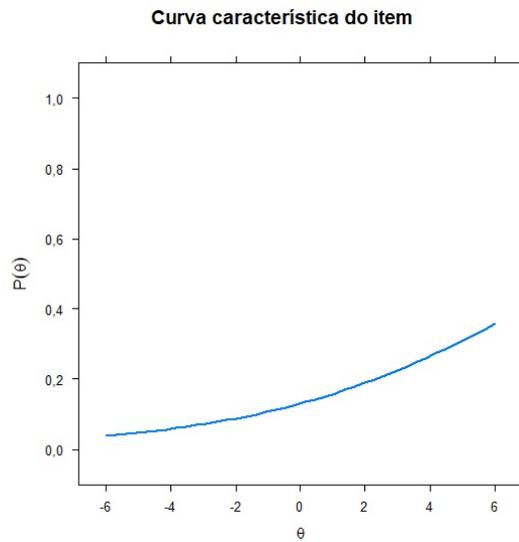
Item 47. Simplifique a expressão:

$$\operatorname{sen} x(\operatorname{cosec} x - \operatorname{sen} x).$$

- (a) $1 - \cos x$
- (b) $1 - \operatorname{sen} x$
- (c) $\operatorname{sen}^2 x$
- (d) $\boxed{\cos^2 x}$
- (e) não sei

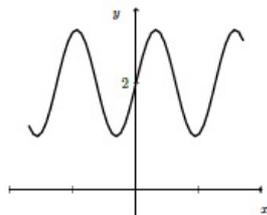


Tópico: Identidade trigonométrica

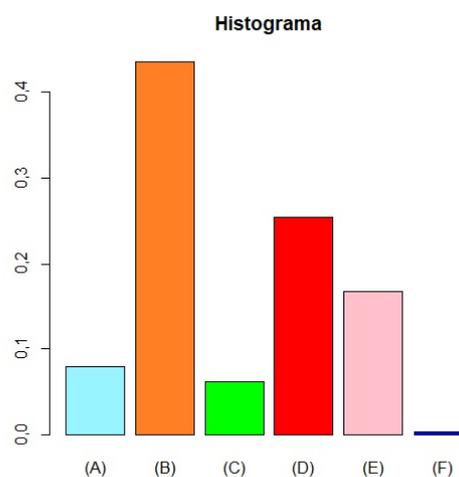


Análise do item 48

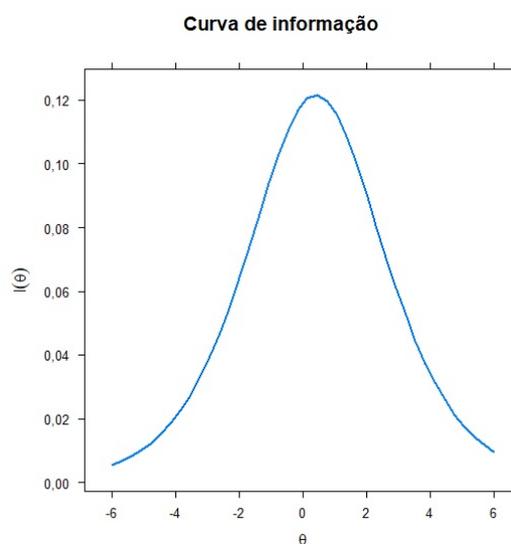
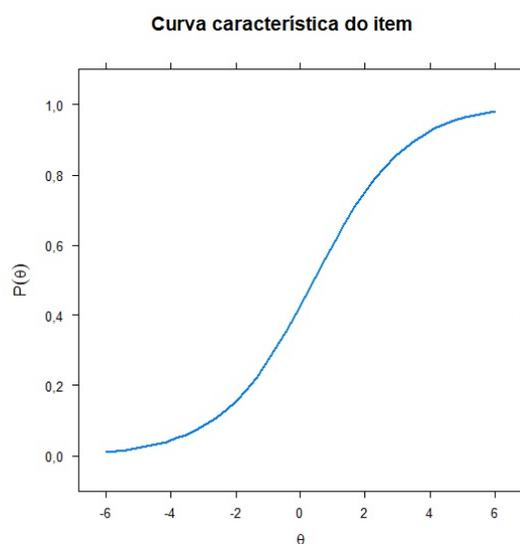
Item 48. Determine a função trigonométrica que melhor se aproxima do gráfico.



- (a) $\text{sen } x$
- (b) $2 + \text{sen } x$
- (c) $\text{cos } x$
- (d) $2 + \text{cos } x$
- (e) não sei

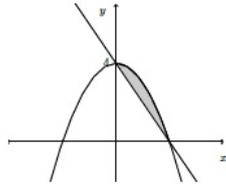


Tópico: Função trigonométrica.



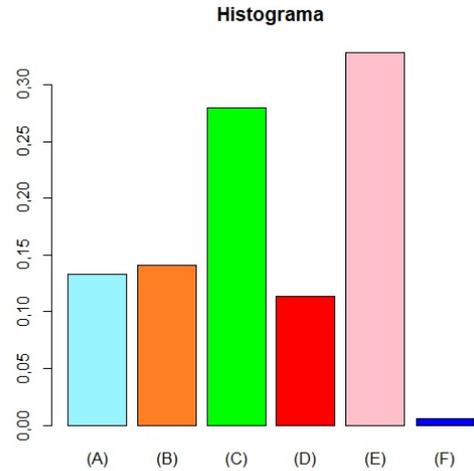
Análise do item 49

Item 49. Dada as funções f e g tais que $f(x) = -x^2 + 4$ e $g(x) = -2x + 4$, determine o conjunto que melhor descreve a região hachurada.

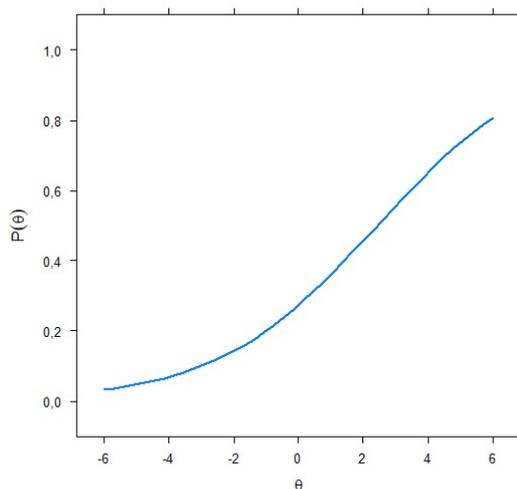


- (a) $\{(x, y) \in \mathbb{R}^2 | y \geq -x^2 + 4 \text{ e } y \leq -2x + 4\}$
- (b) $\{(x, y) \in \mathbb{R}^2 | y \geq -x^2 + 4 \text{ e } y \geq -2x + 4\}$
- (c) $\{(x, y) \in \mathbb{R}^2 | y \leq -x^2 + 4 \text{ e } y \geq -2x + 4\}$
- (d) $\{(x, y) \in \mathbb{R}^2 | y \leq -x^2 + 4 \text{ e } y \leq -2x + 4\}$
- (e) não sei

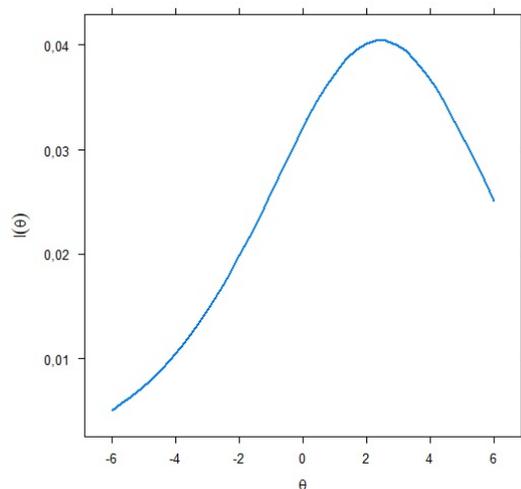
Tópico: Área entre curvas.



Curva característica do item

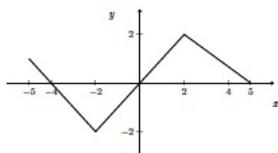


Curva de informação



Análise do item 50

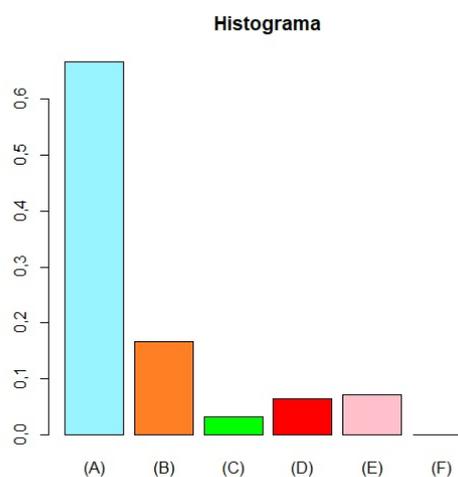
Item 50. Seja o gráfico da função real f :



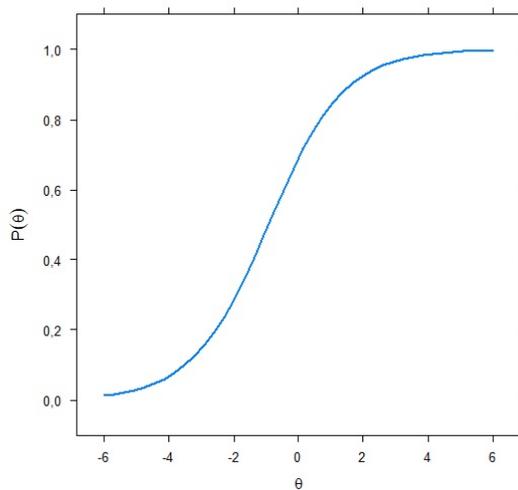
O intervalo em que $f(x) < 0$ é:

- (a) $(-4, 0)$
- (b) $(-4, -2)$
- (c) $(-5, -2) \cup (4, 5)$
- (d) $(-5, -2) \cup (2, 5)$
- (e) não sei

Tópico: Intervalos em que a função é positiva e negativa.



Curva característica do item



Curva de informação

