



Universidade de Brasília  
IE - Departamento de Estatística  
Estágio Supervisionado 2

**Aplicação da Teoria da Resposta ao Item na  
avaliação das Disciplinas de Serviço do  
Departamento de Estatística da UnB**

**João Marcelo Brito Alves de Faria**

Relatório Final

Orientador: Prof.º Dr Afrânio Márcio Corrêa Vieira

Brasília

Novembro de 2011

## Resumo

A Teoria de Resposta ao Item tem passado por um forte desenvolvimento teórico e está sendo utilizada como base em alguns exames e provas importantes no Brasil. A base estatística da teoria permite aos Estatísticos certa desenvoltura no trato e na interpretação dos resultados obtidos. Porém, a Avaliação Educacional possui outras características que não são abordadas pelos estatísticos, mas estão inseridas em outras áreas do conhecimento e aplicações.

Para um maior conhecimento da aplicabilidade da Teoria de Resposta ao Item esse trabalho propõe-se a desenvolver um processo avaliativo da síntese do instrumento aplicado até a posterior análise dos parâmetros estimados.

O instrumento desenvolvido contém 9 itens de múltipla escolha com 4 opções de resposta. As respostas dos indivíduos foram dicotomizadas para a posterior utilização do modelo dicotômico. O modelo utilizado para a síntese de foi modelo de 3 parâmetro, denominado também como modelo de Birnbaum. A estimação dos parâmetros do modelo supracitado fora sintetizada com a utilização das ferramentas presentes no ambiente R, utilizando o pacote **ltm** e a função pertencente a esse, **tpm**. A estimação propriamente dita é utilizada via máxima verossimilhança marginal.

O estudo evidenciou que a Teoria de Resposta ao Item é um processo contínuo e deve estar presente em todo o processo avaliativo. A síntese de itens é momento crucial deste e a sua eficiente utilização deve ser acompanhada de calibração e análises constantes.

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>7</b>
<b>2</b>	<b>AVALIAÇÃO EDUCACIONAL E TÉCNICAS DE MENSURAÇÃO</b>	<b>11</b>
2.1	AVALIAÇÃO EDUCACIONAL . . . . .	11
2.1.1	SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BRASILEIRA (SAEB) . . . . .	16
2.1.2	EXAME NACIONAL DO ENSINO MÉDIO(ENEM) . . . . .	18
2.2	TEORIA DE RESPOSTA AO ITEM . . . . .	18
2.2.1	CONCEITOS BÁSICOS . . . . .	18
2.2.1.1	MODELOS E SEUS PARÂMETROS . . . . .	20
2.2.1.2	MODELO LOGÍSTICO DE 3 PARÂMETROS . . . . .	21
2.2.2	INTERPRETAÇÃO DOS PARÂMETROS DOS ITENS . . . . .	23
2.2.2.1	ESCALAS . . . . .	25
2.2.2.2	MÉTODOS DE ESTIMAÇÃO . . . . .	27
<b>3</b>	<b>RESULTADOS E ANÁLISES</b>	<b>31</b>
3.1	ITEM 1 . . . . .	33

3.2	ITEM 2 . . . . .	34
3.3	ITEM 3 . . . . .	36
3.4	ITEM 4 . . . . .	37
3.5	ITEM 5 . . . . .	38
3.6	ITEM 6 . . . . .	40
3.7	ITEM 7 . . . . .	41
3.8	ITEM 8 . . . . .	42
3.9	ITEM 9 . . . . .	43
3.10	CONSIDERAÇÕES FINAIS . . . . .	45
	<b>APÊNDICE I</b>	<b>48</b>
	<b>APÊNDICE II</b>	<b>50</b>
<b>4</b>	<b>REFERÊNCIAS</b>	<b>52</b>

# Lista de Figuras

2.1	Curva Característica do Item . . . . .	22
2.2	Curvas Características para diferentes $b$ . . . . .	23
2.3	Curvas Características para diferentes $a$ . . . . .	24
2.4	Método da Quadratura . . . . .	30
3.1	CCI 1 a 5 . . . . .	34
3.2	CCI 6 a 9 . . . . .	34
3.3	Curva Característica Item 1 . . . . .	35
3.4	Curva Característica Item 2 . . . . .	36
3.5	Curva Característica Item 3 . . . . .	37
3.6	Curva Característica Item 4 . . . . .	38
3.7	Curva Característica Item 5 . . . . .	39
3.8	Curva Característica Item 6 . . . . .	40
3.9	Curva Característica Item 7 . . . . .	41
3.10	Curva Característica Item 8 . . . . .	42
3.11	Curva Característica Item 9 . . . . .	43

# Lista de Tabelas

3.1	Acertos e erros . . . . .	32
3.2	Estimativas para os Parâmetros dos Itens . . . . .	32
3.3	Estimativas para os Parâmetros dos Itens ( $c_i$ ) . . . . .	33
3.4	Estimativas para os Parâmetros dos Itens ( $a_i$ ) . . . . .	33
3.5	Estimativas para os Parâmetros dos Itens ( $b_i$ ) . . . . .	33

# Capítulo 1

## INTRODUÇÃO

Avaliação Educacional hoje não é apenas um campo com teorias, processos e métodos específicos, mas também um campo abrangente que comporta subáreas, com características diferentes: avaliação de sistemas educacionais, de desempenho escolar em sala de aula, de rendimento escolar com objetivo de macroanálises, de programas, avaliação institucional e auto-avaliação. Admite-se ainda diferentes enfoques teóricos como avaliação sistêmica, avaliação iluminativa ou compreensiva, avaliação participativa etc (Gatti, 2002).

A dificuldade em avaliações se deve, em parte, à posição marginal que essa área esteve na Academia Brasileira, sem que essa vertente tivesse espaço para produção científica e conseqüente desenvolvimento.

Historicamente, o primeiro grupo no Brasil dedicado a esse tipo de problema foi o CETPP - Centro de Estudos de Testes e Pesquisas Psicológicas, da Fundação Getúlio Vargas do Rio de Janeiro, criado em 1966. Em 1995, o Saeb - Sistema Nacional de Avaliação da Educação Básica utilizou a Teoria de Resposta ao Item como a diretriz para a obtenção do desempenho acadêmico dos estudantes provenientes da Educação Básica brasileira. A dificuldade principal, no contexto de avaliação educacional,

consiste em não se conseguir distinguir a habilidade do respondente e a dificuldade presente no instrumento de avaliação.

As habilidades de alunos em determinadas áreas não podem ser mensuradas diretamente. De Ayala (2009) define um construto como *característica que não pode ser observada diretamente*, e, portanto, as habilidades supracitadas podem ser classificadas com um construto, ou *traço latente*. O processo de mensuração deste traço latente envolve a formalização do conceito do construto em questão, no nosso caso, a habilidade em determinada área de conhecimento. Então, após essa formalização podem ser definidas as manifestações que levarão a inferência do traço latente.

Uma técnica de mensuração de traços latentes é a Teoria Clássica dos Testes, que tem como problema grave o fato de o instrumento construído depender intrinsecamente do objeto medido (Pasquali, 2003). Outro entrave nessa teoria consiste em que os parâmetros dos itens de um teste dependem da amostra dos respondentes que foram utilizados para estimação daqueles. A teoria clássica também tem como incongruência a estimação dos parâmetros de discriminação, pois a discriminação de cada item é testada contra o escore total que é constituído por todos os itens do teste, inclusive o item que está sendo analisado.

A Teoria de Resposta ao Item (TRI) tem grande importância, pois a habilidade do respondente não depende do teste, e portanto, podem ser realizadas comparações entre os respondentes, independente do teste a que foram submetidos, evidente, na mensuração do mesmo construto.



Outro ponto que incorpora grande importância a TRI é que a estimação dos parâmetros dos itens pode ser realizada sem que essa esteja condicionada à amostra em particular. Paralelamente, a TRI proporciona a possibilidade de mensuração da dificuldade dos itens na escala das habilidades dos respondentes, ou seja, os respondentes podem ser posicionados aos itens mais prováveis de responder, sejam itens mais fáceis ou mais difíceis, logo, as habilidades estimadas dos respondentes podem ser comparadas entre si.

O objeto de estudo neste projeto é a avaliação das habilidades em Análise Exploratória de Dados. O objetivo consiste na produção de um instrumento, questionário, que utilize a TRI para avaliar as habilidades dos indivíduos no uso da Análise Exploratória de Dados (AED). A AED tem grande importância nas mais variadas profissões, pois engloba o uso de técnicas gráficas e quantitativas para obtenção de informações a cerca de determinada variável. Em áreas como a Biologia, Agronomia e Farmácia é de suma importância para uma perspectiva inicial sobre o problema em questão. As ciências sociais utilizam essa ferramenta como alicerce de diretrizes a serem seguidas em diversas análises. Na área das ciências exatas propõe uma familiarização do profissional com informações ocultas na estrutura da variável em questão, comportamento da variável em estudo e sua variação. Na Economia é de grande valia haja vista a Econometria que utiliza de ferramental estatístico para análise de variáveis econômicas.

A técnica AED é ensinada na UnB pelo Departamento de Estatística em disciplinas de serviço, Estatística Aplicada, Bioestatística e Probabilidade e Estatística.

Essas disciplinas, como ditas, são caracterizadas como de serviço, ou seja, são aplicadas a um grande número de estudantes da UnB, então, variados professores são responsáveis por lecioná-las. Logo, essa vasta quantidade de professores causa variabilidade na forma como o conteúdo supracitado é lecionado. Então, faz-se necessária avaliar habilidades dos estudantes, para uma percepção do Departamento de qual é a absorção da AED lecionada para os variados grupos de alunos que utilizam desse serviço.

Uma maneira de tentar mensurar as habilidades dos estudantes quanto a esse conteúdo é por meio da utilização da TRI, pois assim podem ser comparadas as habilidades dos estudantes dessas disciplinas.

Poderão ser feitas considerações a respeito dos itens avaliados e uma posterior comparação de como estão relacionados à síntese dos itens de avaliação com as formas como os estudantes se comportam a frente dos itens propostos.

## Capítulo 2

# AVALIAÇÃO EDUCACIONAL E TÉCNICAS DE MENSURAÇÃO

### 2.1 AVALIAÇÃO EDUCACIONAL

A Avaliação Educacional é uma área de atuação acadêmica que permaneceu esquecida na pesquisa brasileira durante a maior parte do século passado. Diz-se esquecida quando comparada a atenção dispensada à mesma essa área quando comparada a outros países(Gatti, 2002). Porém, após a crescente necessidade de uma melhoria na qualidade da Educação Brasileira fez-se necessária uma atualização na metodologia utilizada para avaliá-la.

Ainda que tenha tido início somente no final do século passada, esse processo, atualmente, abrange diversos processos, métodos e teorias bastantes específicas. O referido desenvolvimento foi impulsionado em função dos programas avaliativos em variados níveis do sistema educacional brasileiro. Para a realização desses programas fez-se necessária a formação de pesquisadores no exterior e a preparação especial de funcionários dos variados níveis da administração do sistema escolar.

O problema inicial referente a Avaliação Educacional faz referência a essa ser confundida com a simples aplicação de provas, porém ela tem grande responsabilidade

no contexto educacional, pois, as consequências dessa são aprovação, reprovação e evasão escolar. Outro problema é consequência de uma focalização completa no aluno, pois o processo avaliativo deveria, também, inserir o professor nesse contexto. Surge, então, uma estrutura paradoxal, pois, quem avaliaria mais fidedignamente o professor seriam os próprios alunos.

O que pode ser considerada a primeira iniciativa em relação ao aprimoramento da avaliação educacional brasileira deu-se com o Centro de Estudos de Testes e Pesquisas Psicológicas (CETPP), em 1966. Nesse momento foi realizada uma pesquisa que tinha como objetivo compreender o nível de conhecimento dos alunos de Ensino Médio e buscou-se conhecer algumas características sócio-econômicas destes, e, para tanto, foram desenvolvidos cursos sobre elaboração de provas objetivas.

No mesmo período a Fundação Carlos Chagas buscou formar especialistas em medidas e testes, porém esses conhecimentos adquiridos não foram utilizados para avaliação nas redes de ensino. Processos seletivos para universidades, cargos públicos utilizaram do conhecimento supracitado em suas atividades.

Na década de 70, realizou-se um estudo com finalidade de síntese de um instrumento que permitisse a verificação do desempenho dos alunos da 1ª série do Ensino Fundamental.

Após essa crescente busca de aprimoramento da avaliação educacional observou-se um hiato nessa área, pois as Administrações Públicas subsequentes não mostraram interesse nesse elemento estrutural. A exceção encontrada foi o município de São Paulo que realizou um estudo avaliativo da rede de ensino, com o objetivo de utilização

desses dados para estruturação de atividades, tanto escolares como extra-escolares.

O sistema nacional de avaliação da educação básica é um grande marco nessa vertente de estudos. Anteriormente a síntese, de fato, desse conjunto de índices algumas atividades foram realizadas, como estudos pilotos, para que, com maior chance, fosse construído um sistema válido para entendimento da educação básica brasileira.

Avaliaram-se alunos da 1<sup>a</sup>, 3<sup>a</sup>, 5<sup>a</sup> e 7<sup>a</sup> série de escolas públicas em 10 capitais de Estados. Fora utilizada a teoria clássica e, portanto, o que despendeu esforço foi a construção de provas que pudessem ser válidas para as diferentes realidades das cidades supracitadas. Em decorrência desse projeto inicial, outros foram desenvolvidos como a avaliação dos alunos de 3<sup>a</sup> série do ensino médio, o Estado do Paraná realizou a aquilatação de alunos da 2<sup>a</sup> e 4<sup>a</sup> séries e ,por fim, foram feitas práticas semelhantes em escolas privadas de 11 Estados e do Distrito Federal.

Avaliações de rendimento escolar e a participação do Brasil no segundo Programa Internacional de Avaliação de Proficiência Educacional também contribuíram fortemente para a implantação do Sistema de Avaliação da Educação Brasileira (SAEB).

A avaliação educacional deixou de ser utilizada apenas para verificar o desempenho de alunos, possuindo caráter mais amplo, relativo a qualidade de programas, cursos, instituições dentre outros.

Na literatura encontra-se uma divisão em quatro gerações do processo avaliativo. A primeira geração se concentrava em mensurar rendimento escolar. A segunda geração deu-se graças a importância dada ao objeto da avaliação, tentando assim

entender melhor esse. Ou seja, a avaliação passou a ter, também, como objetivo determinar se os objetivos educacionais foram alcançados e em que proporção. A terceira geração incorporou o julgamento de mérito ou de valor ao processo avaliativo. Nesse momento deu-se a necessidade da presença de juízos de valor na avaliação e elucida-se a importância do processo de tomada de decisão após a realização de julgamento de valor. A quarta geração tem caráter político pois trás como característica central a negociação ou consenso entre pessoas de valores diferentes (Somera).

Após a análise das contribuições dadas pelas gerações, conclui-se que a avaliação educacional é instituição mutante e inacabada, pois, está sempre inserida em contextos históricos distintos e, por isso, tende a se adaptar a realidade na qual é observada. Fator relevante que, por vezes, se dilui em outros aspectos faz referência as características de um avaliador. Esse não pode ser alçado ao cargo de avaliador, apenas e exclusivamente, por estar inserido em uma Instituição de Ensino. Para que um avaliador seja de fato competente e confiável esse deve ter algumas características, são elas :

- Competência técnica;
- Competência no trato do objeto da avaliação;
- Habilidades em relações interpessoais;
- Objetividade.

Outras características devem ser observadas em um avaliador que possa, de fato, avaliar e ter resultados relevantes e satisfatórios no processo avaliativo. A reunião

dessas características em um único indivíduo não tem grande probabilidade de ocorrer, portanto, uma equipe avaliadora pode, de forma mais eficiente, sintetizar processos avaliativos. Para tanto, as avaliações devem ser embutidas de interações entre os avaliadores e aqueles que serão avaliados.

Alguns problemas são encontrados nas conseqüências observadas após a comunicação dos resultados provenientes de avaliações em grande escala. Uma delas consiste na resistência de alguns professores e alunos quanto a generalizações de conclusões, em virtude do caráter amostral de certas avaliações.

A questão motivacional também deve ter seu espaço, pois, na maioria dos casos a avaliação é impositiva. Em virtude dessa obrigatoriedade, os respondentes dessas avaliações de grande escala respondem às questões pertencentes a essas sem a essência requerida, para que os dados provenientes dessas provas possam gerar informações, de fato, relevantes. Além disso, múltiplas variáveis devem ser consideradas nesse processo, pois, não apenas a relação professor-aluno é responsável pelo rendimento desses em uma avaliação, equivalência entre idade e série, horas de estudo extra classe, participação da família no acompanhamento nas atribuições escolares, dentre outros aspectos também influem nos resultados obtidos por avaliados.

Na década de 90 ocorreram variadas tentativas pela busca de incremento das informações sobre a Educação Brasileira. Ficou evidenciado, nesse primeiro momento, a possibilidade de utilização de instrumentos referenciados a critério ou referenciados a normas. Algumas dessas avaliações tiveram caráter censitário, porém a maioria dessas foram amostrais. As censitárias traziam alguns pontos interessantes no pro-

cesso avaliativo, como, por exemplo, requeria a participação completa dos estados não apenas na problemática da logística da aplicação dos instrumentos mas, também, a participação desses no processo de síntese dos instrumentos avaliativos e também de uma correção preliminar. Fator relevante nesse processo foi a modificação da sofisticação do tratamento estatístico quem eram conferidos aos dados provenientes dessas avaliações. A TRI já estava, de certa forma, consolidada em alguns países desenvolvidos, porém a maioria dos professores, tanto do ensino fundamental como os do ensino médio não teriam conhecimento técnico suficiente para entender os resultados advindos da utilização dessas novas técnicas.

Após os primeiros movimentos nesse sentido as avaliações passaram a convergir para um mesmo modelo. Optou-se, quase totalitariamente, pelas avaliações amostrais e pela utilização do Banco de Dados do Instituto Nacional de Estudos e Pesquisas Educacionais com itens anteriormente já testadas, calibradas.

### **2.1.1 SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BRASILEIRA (SAEB)**

O SAEB foi implantado em 1993 pelo Ministério da Educação juntamente com as Secretarias Estaduais de Educação (Gatti, 2002). Inicialmente, tinha como estrutura um modelo de amostra nacional de alunos e provas objetivas no modelo clássico. A estrutura utilizada sofreu aperfeiçoamentos e passou a ser aplicada anualmente, entre séries do ensino fundamental e médio e com diretrizes advindas da TRI. Essa nova abordagem utilizada permitiria a obtenção de informações mais precisas a respeito da real conjuntura da educação brasileira, além de condições de comparabilidade em escala.



O SAEB é subdivido em duas grandes vertentes, a primeira está relacionada com o acesso ao ensino básico onde são verificados o atendimento à demanda e a eficiência, na segunda encontram-se pontos relacionados ao estudo da qualidade, que implica um detalhamento em:

- o desempenho do aluno e o desenvolvimento de competências;
- perfil sócio-econômico, hábitos, perfis dos alunos, condições de trabalho entre outros aspectos inerentes a todas as escolas;
- planejamento, projeto pedagógico, estratégias de ensino;
- infra-estrutura, instalações e equipamentos.

A última atualização no SAEB faz referência a agregação ao modelo estudos de casos de situações específicas consideradas relevantes para análises mais profundas. Recentemente, o Ministério da Educação expandiu a avaliação de desempenho e instituiu a *Prova Brasil*, integrado com o SAEB. Pôde-se assim construir um indicador mais interessante, o Índice de Desenvolvimento de Educação Básica (IDEB). Pôde-se, também, encontrar um valor desse indicador para o conjunto de escolas do Brasil, dos Estados, dos Municípios e de cada escola em particular.

Em suma, as atividades desenvolvidas, tal como o SAEB e outras iniciativas de outras frentes, como Secretarias de Educação Estaduais, ou o próprio Ministério da Educação, fomentaram mudanças na conotação dos processos avaliativos, retirando-lhes o caráter punitivo e transformando-os em instrumentos objetivos e possuidores de finalidades representativas da real estrutura de centros de ensino. Os resultados

dessa transformação na Avaliação gerou dados em âmbitos nacional, estadual e municipal, porém a forma como esses dados são disponibilizados e a maneira como são utilizados para interferirem na síntese de projetos para as instituições educacionais precisam ser aperfeiçoados.

### **2.1.2 EXAME NACIONAL DO ENSINO MÉDIO(ENEM)**

Algumas questões interessantes sobre os movimentos iniciais do Exame Nacional do Ensino Médio (ENEM) podem ser enfocadas, por exemplo, a não obrigatoriedade do exame, a questão da taxa cobrada para a participação do exame, o uso dos resultados para ingresso no ensino superior, dentre outras. Esse exame foi concebido para verificar competências e habilidades,segundo a formulação dos seus responsáveis, e pretende avaliar cinco competências e vinte e uma habilidades (Vianna, 2003).

O órgão responsável pelo ENEM indica que o exame não mede conteúdos, mas apenas competências e habilidades. Bloom indica que quando são avaliados conteúdos, implicitamente, avaliam-se as capacidades. Vianna indica que o ENEM poderia ser mais eficiente e palatável caso ficasse restrito apenas a duas capacidades básicas, a verbal e a numérica.

## **2.2 TEORIA DE RESPOSTA AO ITEM**

### **2.2.1 CONCEITOS BÁSICOS**

Por vezes, faz-se necessário mensurar alguma característica que não pode ser mensurada diretamente. Medidas de altura, espessura, peso, entre outras, possuem a características de serem diretamente mensuráveis. Porém existem outras características que não acomodam esse padrão.

Habilidades intelectuais, preferências perante certos aspectos da sociedade, expectativas, entre outras, devem ser mensuradas de maneira indireta, por vezes, via questionário, teste ou outro instrumento que consiga indiretamente quantificar a característica de interesse. Esse tipo de característica é denominada traço latente, ou construto. Dá-se um comportamento probabilístico para o padrão de respostas para os respondentes e, assim, podem ser estimados tanto parâmetros dos itens, como, dificuldade, discriminação e chance de acerto ao acaso, e, também, parâmetros relacionados com os respondentes, a latência em questão.

Uma questão importante quanto ao uso das técnicas relacionadas a TRI faz referência a questão do construto. Deve-se ter em mente que ao se utilizar as técnicas provenientes da TRI a latência em questão não pode estar relacionada com outros construtos, pois, caso ocorra a estimação do traço latente em estudo não gera informações plausíveis sobre o assunto em questão.

Após a análise dos dados, deve-se criar uma escala para a posterior interpretação entre as diferenças dos traços latentes estimados e também da dificuldade presente no item, haja vista que essas grandezas devem ser pertencentes à mesma escala.

### 2.2.1.1 MODELOS E SEUS PARÂMETROS

No início do desenvolvimento desse ferramental estatístico, em meados da década 1950, fora proposto um modelo de probabilidade para a resposta do respondente a determinado teste, que possuía apenas itens dicotômicos, dois parâmetros relacionados ao item, a dificuldade e a discriminação desse, e o parâmetro relacionado com o respondente. A forma como esses padrões de resposta se acomodavam era proveniente de uma função ogiva normal. Esse modelo foi proposto por Lord em 1952. Após algum tempo de trabalho com esse modelo, Lord percebeu que havia a necessidade de se incorporar um terceiro parâmetro, relacionado com o acerto casual. Porém, a função ogiva normal traz dificuldades no processo de estimação, passou então, a se utilizar a função logística, pois essa é uma função explícita dos parâmetros em estudo. Essa alteração foi realizada por Birnbaum em 1968. Paralelamente ao trabalho de Lord, fora desenvolvido um modelo que possuía apenas um parâmetro relacionado com o item em questão. Inicialmente, Rasch, em 1960, utilizou a função ogiva normal, logo após, Wright, em 1968 descreveu o mesmo modelo utilizando a função logística.

Fica evidente que em alguns casos existia a necessidade de se estimar tanto os parâmetros relacionados com os itens como os relacionados com os respondentes. Inicialmente, utilizou-se o método da máxima verossimilhança conjunta, porém quando existente a necessidade da estimação simultânea supracitada, deparava-se com grandes problemas computacionais (Andrade et al, 2000). Fora sugerida a utilização do método da máxima verossimilhança marginal, que então, era realizada em

duas etapas. O primeiro passo consistia em estimar os parâmetros relacionados com os itens, assumindo determinada distribuição probabilística para as habilidades e na segunda etapa realizava-se a estimação dos parâmetros referentes as habilidades. Recentemente, fora sugerida a utilização de métodos bayesianos para o processo de estimação dos parâmetros supracitados.

Os modelos mencionados são dependentes de três fatores fundamentais:

- 1 a natureza do item, referente a forma da resposta, dicotômicos ou não;
- 2 o número de populações envolvidas no estudo;
- 3 o número de construtos a serem investigados.

#### **2.2.1.2 MODELO LOGÍSTICO DE 3 PARÂMETROS**

O modelo, para ser válido, deve atender a algumas suposições:

- Unidimensionalidade : apenas um construto deve estar presente no instrumento.
- Os itens utilizados no teste devem ser relevantes para a estimação do construto;
- Independência local : a resposta a determinado item não pode influenciar na resposta a outro item;
- Deve existir um fator dominante para a estimação do traço latente, para que, assim, apenas um traço latente seja avaliado.

A relação existente entre a Função de Resposta ao Item e os parâmetros de determinado modelo são observadas, graficamente, pela Curva Característica do

Item(CCI). Observa-se uma CCI na Figura 2.1.

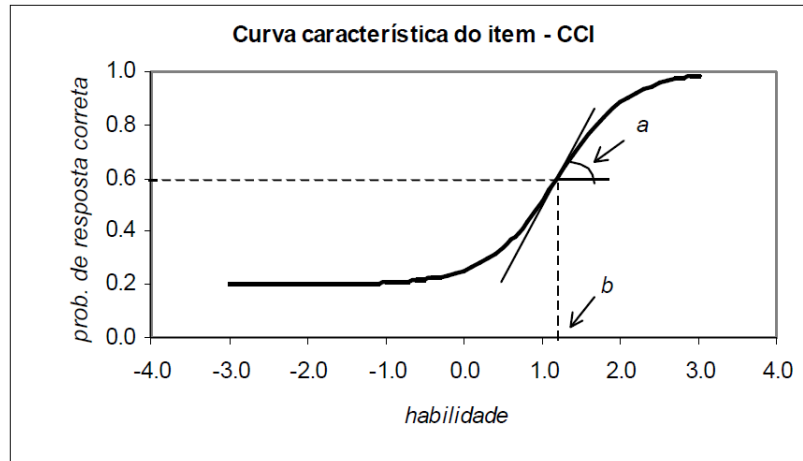


Figura 2.1: Curva Característica do Item

Deve-se ter em mente que os itens tratados nesse estudo devem ser aqueles que apresentam o seguinte comportamento: quão maior a habilidade do respondente maior será a chance de "acerto" para determinado item.

O modelo mais comumente utilizado é o modelo logístico de 3 parâmetros, para respostas dicotômicas, e é estruturado da seguinte maneira:

$$P(U_{ij}|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2.1)$$

onde

- $U_{ij}$  é uma variável dicotômica que recebe valor 1 quando o indivíduo  $j$  responde corretamente ao item e 0 caso contrário;
- $\theta_j$  corresponde ao valor do traço latente para o  $j$ -ésimo respondente;
- $P(U_{ij}|\theta_j)$  é a Função de Resposta ao Item;

- $a_i$  é o parâmetro referente a discriminação do  $i$ -ésimo item;
- $b_i$  é o parâmetro referente a dificuldade do  $i$ -ésimo item;
- $c_i$  é o parâmetro correspondente ao "chute", acerto casual sob baixa habilidade do  $i$ -ésimo item;
- $D$  é um fator de escala que relaciona os resultados encontrados com a utilização da função logística com os valores provenientes da função ogiva normal.

A Figura 2.2 mostra diferentes curvas características provenientes de diferentes itens que possuem variações no parâmetro  $b$ , dificuldade.

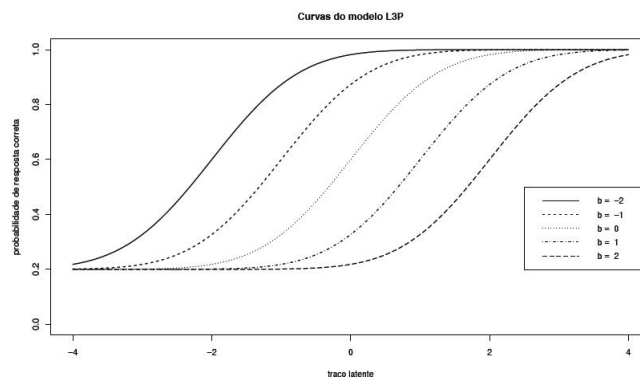


Figura 2.2: Curvas Características para diferentes  $b$

A Figura 2.3 mostra diferentes curvas características provenientes de diferentes itens que possuem variações no parâmetro  $a$ , discriminação.

## 2.2.2 INTERPRETAÇÃO DOS PARÂMETROS DOS ITENS

Os parâmetros do modelo de Birnbaum têm as seguintes interpretações :

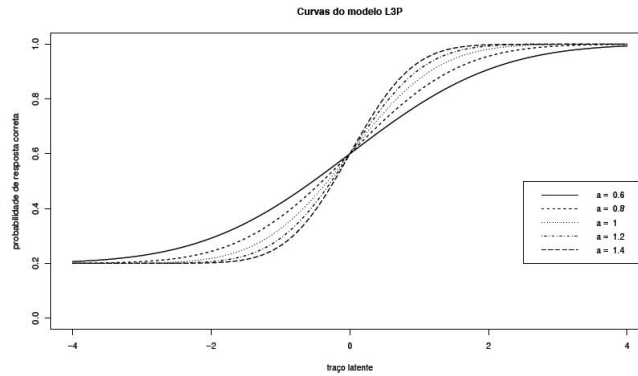


Figura 2.3: Curvas Características para diferentes  $a$

- O parâmetro  $a$  revela a inclinação da CCI no ponto de inflexão dessa. Quanto maior o valor de  $a$  menor é a região na escala na qual os respondentes têm, aproximadamente, a mesma probabilidade de acertar o item;
- O parâmetro  $b$  representa o nível de habilidade necessária para que a probabilidade de acerto seja igual a  $\frac{(1+c)}{2}$ . Esse parâmetro corresponde a dificuldade do item, portanto, quanto maior o  $b$  mais difícil é o item;
- O parâmetro  $c$  representa o "chute", ou acerto casual sob baixa habilidade, sendo a probabilidade do respondente acertar o item dado que a sua habilidade é baixa. Quando, na escala, a habilidade tende a  $-\infty$  o valor encontrado para a probabilidade de resposta correta determina o valor  $c$ .

Outra forma de se interpretar um item é analisar a "função de determinação" deste. A função de informação de Fisher é obtida através da seguinte expressão

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)} \quad (2.2)$$



Onde  $P_i(\theta)$  corresponde a  $P(U_{ij} = 1|\theta_j)$  e  $Q_i(\theta)$  a  $1 - P_i(\theta)$ . Para o modelo logístico de 3 parâmetros tem-se a seguinte função de informação:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (2.3)$$

Portanto, algumas situações podem gerar um item com informação grande, quando a dificuldade se aproxima da habilidade do respondente, quando a discriminação do item é grande e, por fim, quanto menor for o valor de  $c_i$ . A expressão 2.2 revela a informação para um dado item, portanto, a informação do teste como um todo é obtida a partir de

$$I(\theta) = \sum_{i=1}^I I_i(\theta) \quad (2.4)$$

Deve-se ter em mente que os itens tratados nesse estudo devem ser aqueles que apresentam o seguinte comportamento: quão maior a habilidade do respondente maior será a chance de "acerto" para determinado item.

### 2.2.2.1 ESCALAS

Para a construção de escalas de conhecimento, conhecidas também como escalas de habilidade, interpretáveis faz-se necessário que os parâmetros em estudo estejam, após sua estimação, todos na mesma métrica. Haja vista o método utilizado para a análise das informações advindas de um banco de dados de respostas a itens educacionais sabe-se que essas podem ser comparáveis, porém elas não carregam nenhum significado pedagógico. Caso não sejam feitas ligações das estimativas encontradas com os conteúdos envolvidos na avaliação pode-se apenas comparar as estimativas em relação a média e desvio padrão encontrados no grupo em estudo. Em suma, não

é possível dizer o que determinado indivíduo com  $\theta$  superior a de outro respondente tem maior conhecimento.

A síntese dessas escalas tem como base *níveis âncora*, que são de fato estabelecidos por meio de *itens âncora*. Esses pontos são definidos na escala de habilidade, pelo especialista, para serem interpretados pedagogicamente. *Definição de item âncora* : Considere dois níveis âncora consecutivos  $Y$  e  $Z$  com  $Y \leq Z$ . Diz-se que determinado item é *âncora para o nível  $Z$*  se, e somente se, as 3 condições abaixo forem satisfeitas simultaneamente :

- $P(U = 1|\theta = Z) \geq 0,65$  e
- $P(U = 1|\theta = Y) \leq 0,50$  e
- $P(U = 1|\theta = Z) - P(U = 1|\theta = Y) \leq 0,30$

Como já citado, alguns problemas surgem da aplicação de testes em grande amostras, dentre eles, as escalas empregadas, o tipo de instrumento utilizado e os julgamentos comparativos que são comunicados na sociedade.

A utilização de diferentes escalas, "a priori", não traz problemas, a problemática pode surgir caso os referenciais utilizados não tenham pontos comuns, tornando-os assim, não equivalentes. Os referenciais que são comumente utilizados são a média, o desvio padrão e um determinado score  $z$ , que exprime a diferença entre os scores obtidos e a média grupo em termos do desvio padrão. Em termos gerais, geram-se escalas que vão de -3 a 3; porém escalas desse tipo são de trato relativamente fácil para técnicos, especialistas. Porém o entendimento das grandezas observadas na

escala não é trivial. Uma possibilidade de melhora para a compreensão dessa escala está em se transformar o escore, com valores pré-definidos.

O terceiro ciclo do SAEB trouxe alguns resultados importantes, dentre eles pode-se ressaltar a produção de escalas de proficiência. Uma interpretação educacional das habilidades e das competências associadas aos pontos da escala pode ser sintetizada. No presente projeto não serão caracterizados itens âncoras. Em trabalhos futuros com maior quantidade de itens esses devem ser construídos.

### 2.2.2.2 MÉTODOS DE ESTIMAÇÃO

Normalmente, quando são aplicados instrumentos de mensuração de habilidades, entre outros, os parâmetros dos itens não são conhecidos, como o de dificuldade, de discriminação e o de acerto ao acaso. Também não são sabidos os parâmetros de habilidade dos respondentes. Portanto, faz-se necessária a estimação paralela dos dois tipos de parâmetros. Uma técnica que pode ser utilizada para esse fim é a Estimação dos parâmetros por Máxima Verossimilhança Marginal.

Pode-se fazer uma analogia entre a diferença existente entre a Máxima Verossimilhança Marginal e a Máxima Verossimilhança Conjunta. Essa analogia é sintetizada utilizando as informações existentes entre a análise de efeitos mistos e efeitos fixos, respectivamente.

A probabilidade de resposta do vetor  $\mathbf{x}$  é

$$p(\underline{x}|\underline{\theta}, \underline{\eta}) = \prod_{j=1}^L p_j^{x_j} (1 - p_j)^{1-x_j} \quad (2.5)$$

onde  $p(\underline{x}|\underline{\theta}, \underline{\eta})$  é a probabilidade condicional em relação a habilidade dos respondentes,  $\theta$ , e os parâmetros dos itens ( $\eta$  é um matriz de parâmetros dos itens), e  $L$  é

o número de itens presentes no teste.

Aplicando a idéia de amostra ,advinda da amostra de respondentes, faz-se necessária integração sobre a distribuição da população :

$$p(\underline{x}) = \int_{-\infty}^{\infty} p(\underline{x}|\underline{\theta}, \underline{\eta})g(\underline{\theta}|\underline{\zeta})d\underline{\theta} \quad (2.6)$$

onde  $p(\underline{x}|\underline{\theta}, \underline{\eta})$  é dado pela equação 2.5 e  $g(\underline{\theta}|\underline{\zeta})$  representa uma distribuição contínua de indivíduos,  $\zeta$  são os parâmetros de uma distribuição normal padrão. A equação 2.6 é denominada probabilidade marginal do vetor de resposta  $\mathbf{x}$ . E é dessa distribuição marginal que os parâmetros dos itens são estimados. Portanto, o que se faz é multiplicar a verossimilhança dos dados observados pela distribuição da população para a eliminação dos parâmetros de habilidades dos respondentes.

Como o parâmetro de habilidade dos respondentes foi eliminado do processo de estimação, o aumento do tamanho da amostra não gera aumento no número de parâmetros a serem estimados na função de verossimilhança marginal. Deve-se assumir uma forma para a distribuição da população *a priori* ou que seja estimada empiricamente. Para a resolução da integral vista na equação 2.6, podem ser utilizadas diversas técnicas para a resolução, como o método da quadratura de Gauss - Hermite ou baseadas nas fórmulas de integração estruturadas pelos processos de Newton-Cotes.

Existem diversas fórmulas de integração desenvolvidas, estruturalmente são semelhantes à seguinte forma:

$$\int_a^b f(x)dx = \sum_{i=0}^n w_i f(x_i) \quad (2.7)$$

ou seja, são somas ponderadas de  $n + 1$  valores de  $x_i$ . Na quadratura gaussiana, os

pontos não são igualmente espaçados, porém são escolhidos de forma que os  $n + 1$  valores resultam como uma integral exata. Para a compreensão da estrutura de Gauss-Hermite, faz-se necessária uma explanação à respeito dos polinômios de Hermite,  $H_n(x)$ , são caracterizados como ortogonais no intervalo  $[-\infty, \infty]$ , em relação a função peso  $e^{-x^2}$ , ou seja,

$$\int_{-\infty}^{+\infty} e^{-x^2} H_n(x) H_m(x) dx = 0, \quad n \neq m \quad (2.8)$$

$$\int_{-\infty}^{+\infty} e^{-x^2} [H_n(x)]^2 dx = c(n) = 0 \quad (2.9)$$

os primeiros polinômios de Hermite são

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

e a relação recursiva geral é dada por:

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \quad (2.10)$$

Então a quadratura de Gauss-Hermite pode ser escrita da seguinte forma:

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x) d(x) = \sum_{i=0}^n w_i f(x_i) \quad (2.11)$$

onde  $x_i$  são as raízes do polinômio de Hermite de grau  $n + 1$  e o erro correspondente é dado por

$$E_n = \frac{(n+1)! \sqrt{\pi}}{2^{2n+1} (2n+2)!} F^{2n+2}(\zeta) \quad (2.12)$$

A quadratura gaussiana pode ser interpretada como pela soma das áreas de finitos retângulos. A figura a seguir, exemplifica esse método:

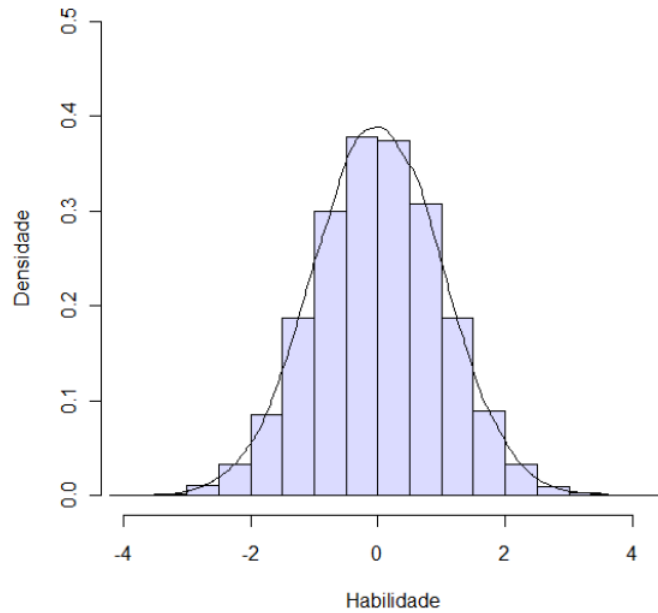


Figura 2.4: Método da Quadratura

Os pontos de quadratura são os pontos médios da base superior de cada retângulo, como visto na Figura 2.4 e são representados por  $\bar{\theta}_k$ , a soma ponderada supracitada terá como pesos  $A_k \equiv A(\bar{\theta}_k)$ . Utiliza-se então, alturas e larguras de acordo com os modelos que estão sendo aplicados.

## Capítulo 3

# RESULTADOS E ANÁLISES

O objetivo do presente projeto é a síntese de um conjunto de atividades tendo como suporte a TRI. As atividades foram:

- Construção do instrumento de mensuração com 9 questões de múltipla escolha tendo como conteúdo a Análise Exploratória de Dados, no Apêndice I encontra-se o questionário contendo os itens;
- Aplicação do instrumento em determinadas turmas de Estatística Aplicada e Bioestatística na UnB;
- Correção dos testes de maneira a "dicotomizar" as respostas dadas pelos respondentes;
- Utilização do banco de dados provenientes das respostas computadas supracitadas para estimação e análise dos parâmetros do Modelo de Birnbaum, denominado como Modelo Logístico de 3 parâmetros (modelo acima já explicitado).

O processo de estimação dos parâmetros dos itens em questão deu-se com a utilização das ferramentas presente no ambiente R, com o pacote *ltm* e a função

*tpm*. A programação para a estimação dos parâmetros e para a síntese dos gráficos estão em anexo.

Após a aplicação de 230 questionários e posterior correção dos testes aplicados ao indivíduos alvo fora sintetizado um banco de dados e, a partir desse, buscou-se estimar os parâmetros referentes ao itens: dificuldade, discriminação e o referente ao acerto casual.

A Tabela 3.1 a seguir mostra as respostas para cada item

Tabela 1 3.1: Acertos e erros

	Acertos	Erros
Item 1	164	66
Item 2	39	191
Item 3	129	101
Item 4	90	140
Item 5	183	47
Item 6	118	112
Item 7	176	54
Item 8	115	115
Item 9	76	154

A Tabela 3.2 a seguir revela esses parâmetros para cada item em questão.

Tabela 3.2: Estimativas para os Parâmetros dos Itens

	Parâmetro "Chute"( $c_i$ )	Parâmetro Dificuldade( $b_i$ )	Parâmetro Discriminação( $a_i$ )
Item 1	0.633	0.798	6.330
Item 2	0.732	0.305	47.350
Item 3	0.506	1.535	2.331
Item 4	0.016	2.366	0.207
Item 5	0.000	-0.743	16.983
Item 6	0.392	7.450	0.189
Item 7	0.000	-1.972	0.652
Item 8	0.000	0.005	0.389
Item 9	0.165	1.036	2.510

As Figuras 3.1 e 3.2 revelam graficamente o comportamento desses parâmetros quando inseridos no modelo de 3 parâmetros.



Tabela 3.3: Estimativas para os Parâmetros dos Itens ( $c_i$ )

Parâmetro "Chute" ( $c_i$ )	Erro Padrão
Item 1	0.0585
Item 2	0.0390
Item 3	0.0708
Item 4	0.3861
Item 5	0.0136
Item 6	NaN
Item 7	0.0157
Item 8	0.0047
Item 9	0.0666

Tabela 3.4: Estimativas para os Parâmetros dos Itens ( $a_i$ )

Parâmetro "Discriminação" ( $a_i$ )	Erro Padrão
Item 1	28.5166
Item 2	15441.1200
Item 3	4.3231
Item 4	0.2241
Item 5	241.8745
Item 6	NaN
Item 7	0.2291
Item 8	0.1804
Item 9	2.4376

Tabela 3.5: Estimativas para os Parâmetros dos Itens ( $b_i$ )

Parâmetro "Dificuldade" ( $b_i$ )	Erro Padrão
Item 1	1.9301
Item 2	101.7010
Item 3	0.6252
Item 4	4.2603
Item 5	0.9214
Item 6	15.0039
Item 7	0.6443
Item 8	0.3521
Item 9	0.2472

### 3.1 ITEM 1

Como visto na Tabela 3.1 os parâmetros para o Item 1 são :

- Parâmetro referente a Dificuldade ( $b_1$ ) = 0.798

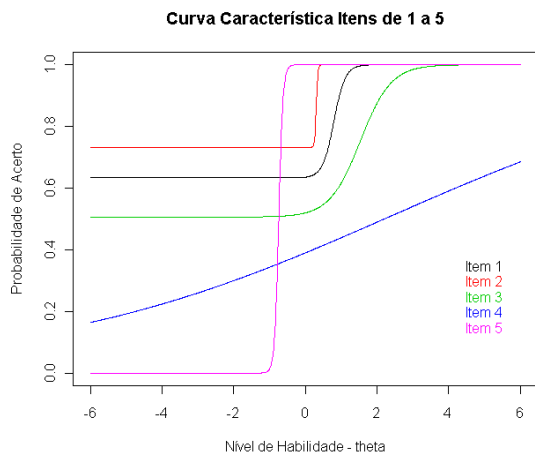


Figura 3.1: CCI 1 a 5

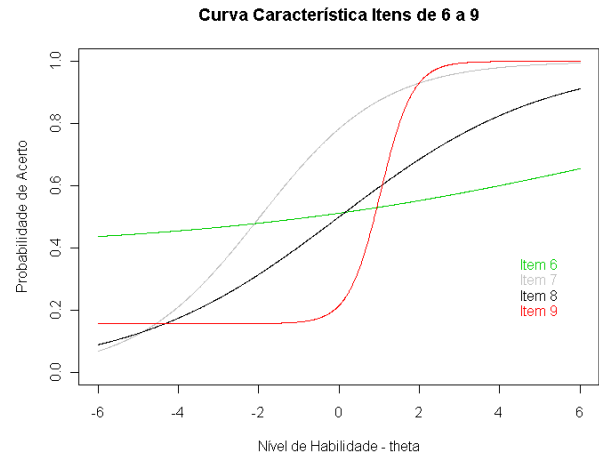


Figura 3.2: CCI 6 a 9

- Parâmetro referente a Discriminação ( $a_1$ ) = 6.33
- Parâmetro referente ao Chute ( $c_1$ ) = 0.633

O Item 1 apresenta dificuldade ( $b_1$ ) com valor mediano, 0.798; além de possuir valor de discriminação elevado, ou seja, para pequenas variações na habilidade do respondente a probabilidade de acerto sofre relevante alteração. Porém o mesmo item possui probabilidade de acerto casual elevada ( $c_1$ ) maior que 0.6, os especialistas indicam que itens com o parâmetro referente ao "chute" maior que 60% (Andrade et all, 2000) não devem ser utilizados para a estimação dos parâmetros referentes aos respondentes. Portanto, esse item deve sofrer alterações quanto aos distratores, pois essas estruturas trazem grande importância no que tange a probabilidade de acerto casual.

## 3.2 ITEM 2

Como visto na Tabela 3.1 os parâmetros para o Item 2 são :

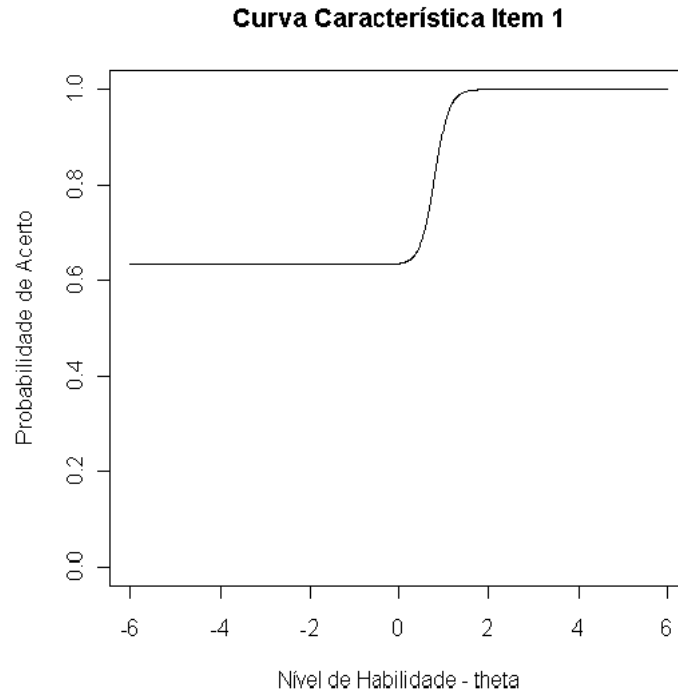


Figura 3.3: Curva Característica Item 1

- Parâmetro referente a Dificuldade ( $b_2$ ) = 0.305
- Parâmetro referente a Discriminação ( $a_2$ ) = 47.35
- Parâmetro referente ao Chute ( $c_2$ ) = 0.732

O Item 2 apresenta dificuldade ( $b_2$ ) com valor mediano, 0.305; porém o item possui probabilidade de acerto casual elevada ( $c_2$ ) maior que 0.7 além de possuir valor de discriminação elevadíssimo o que revela problemática na formulação do item. Especialistas indicam que itens com o parâmetro referente ao "chute" maior que 60% não devem ser utilizados para a estimação dos parâmetros referentes aos respondentes. Portanto, esse item deve sofrer alterações quanto aos distratores, pois essas estruturas trazem grande importância no que tange a probabilidade de acerto casual. Além de uma possível modificação para alteração do parâmetro  $a$ .

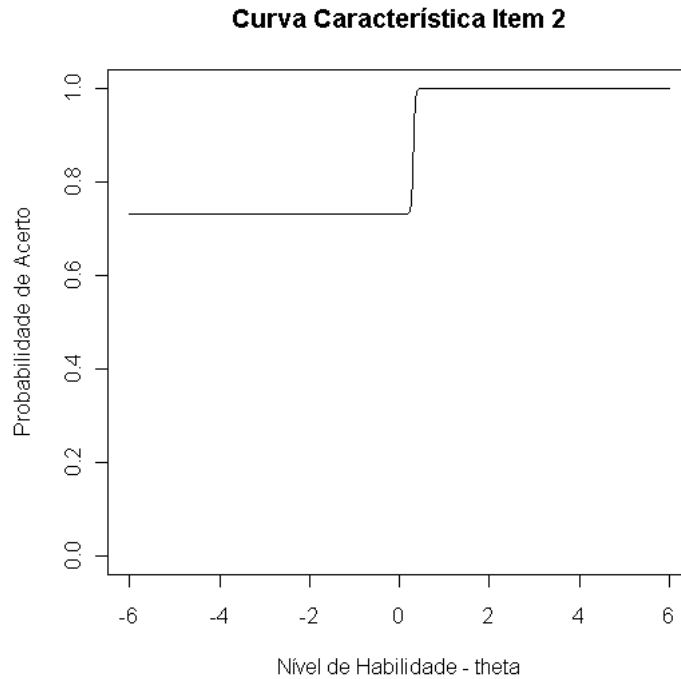


Figura 3.4: Curva Característica Item 2

### 3.3 ITEM 3

O item 3 tem a seguinte estrutura:

Como visto na Tabela 3.1 os parâmetros para o Item 3 são :

- Parâmetro referente a Dificuldade ( $b_3$ ) = 1.535
- Parâmetro referente a Discriminação ( $a_3$ ) = 2.331
- Parâmetro referente ao Chute ( $c_3$ ) = 0.506

O  $b_3 = 1.535$  mostra que o item possui dificuldade relativamente alta quando comparada com os demais itens presente no instrumento supracitado e  $a_3$  revela um poder de discriminação para um item que pode ser utilizado para a estimação das habilidades. Esse item tem potencial para ser um item altamente relevante para o

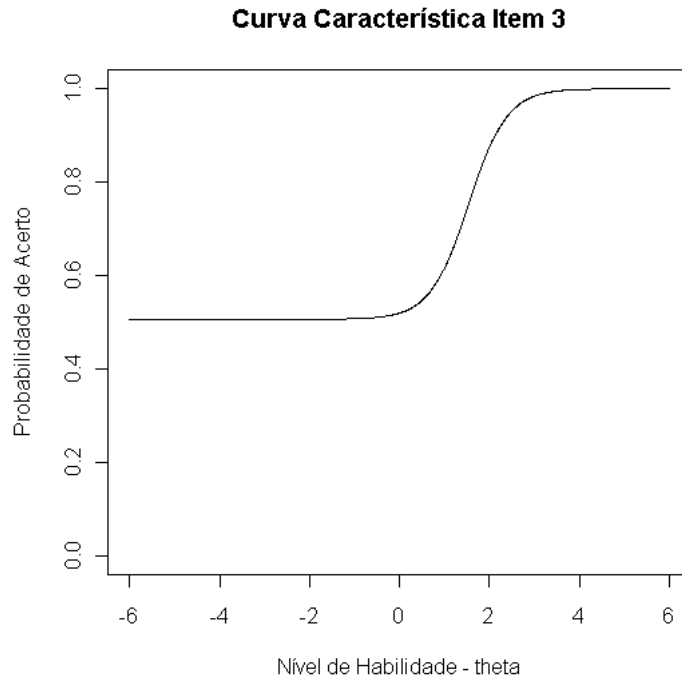


Figura 3.5: Curva Característica Item 3

estudo de aprendizado de AED porém a probabilidade de acerto casual ( $c_3$ )= 0.506 revela que esse aspecto do item deve ser modificado.

### 3.4 ITEM 4

Como visto na Tabela 3.1 os parâmetros para o Item 4 são :

- Parâmetro referente a Dificuldade ( $b_4$ ) = 2.366
- Parâmetro referente a Discriminação ( $a_4$ ) = 0.207
- Parâmetro referente ao Chute ( $c_4$ ) = 0.016

O item 4 apresentou comportamento distinto daquele esperado para uma avaliação educacional, pois a amplitude geralmente utilizada na análise das possíveis habilidades dos indivíduos não foi evidenciada probabilidade de acerto igual a 1, por

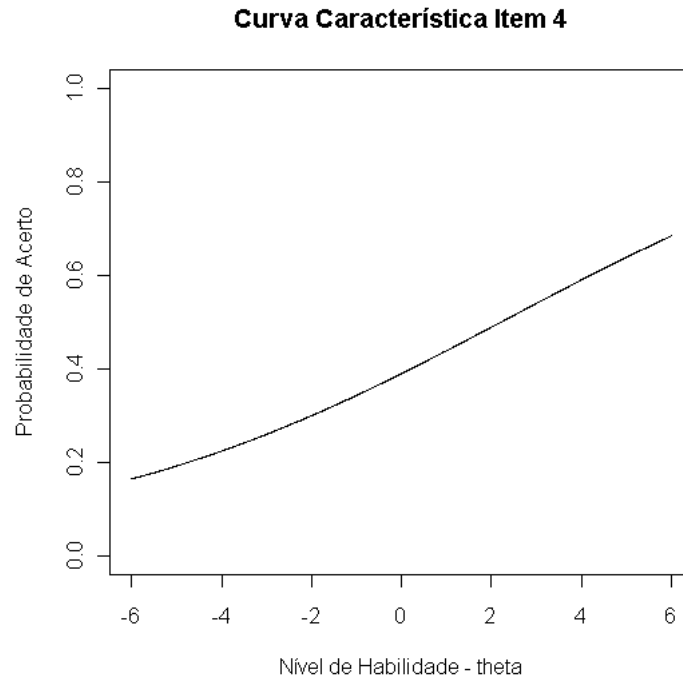


Figura 3.6: Curva Característica Item 4

outro lado também não fora observada probabilidade de acerto igual a 0. Portanto, o tópico abordado nessa questão deve ser utilizado para síntese de outro item.

### 3.5 ITEM 5

Como visto na Tabela 3.1 os parâmetros para o Item 5 são :

- Parâmetro referente a Dificuldade ( $b_5$ ) = -0.743
- Parâmetro referente a Discriminação ( $a_5$ ) = 16.983
- Parâmetro referente ao Chute ( $c_5$ ) = 0

O Item 5 apresentou dificuldade com intensidade relativamente baixa, quando comparado com os demais itens. Portanto o tópico abordado no item não é de difícil

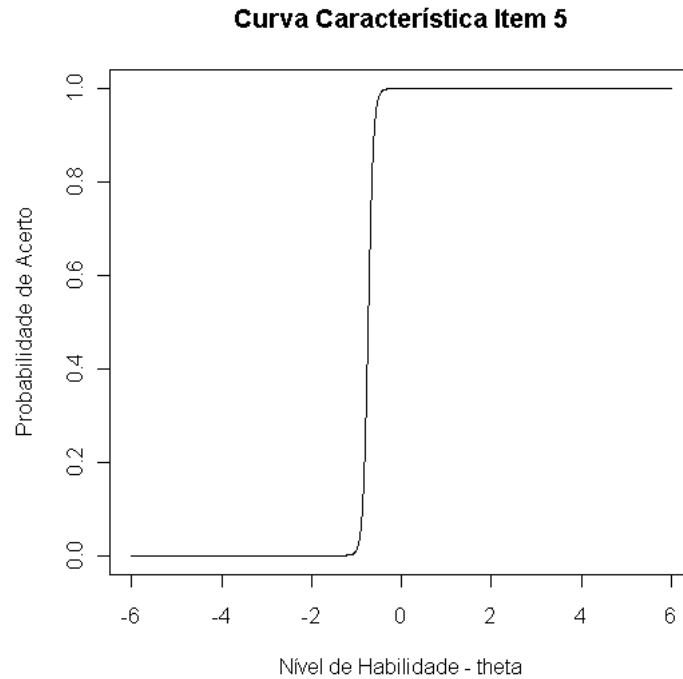


Figura 3.7: Curva Característica Item 5

compreensão por parte dos estudantes. O item tem  $c_5 = 0$  o que mostra um comportamento interessante em relação a possibilidade de acerto mediante chute. Portanto, nesse aspecto o item não precisa de alterações. Porém, o item apresenta  $a_5$ , muito elevado, ou seja, variações muito pequenas entre habilidade de respondentes leva a uma alteração elevada na probabilidade de acerto. Podem ser sintetizadas alterações nos distratores, alternativas incorretas presentes no item, pois não é relevante determinar tamanha precisão para se distinguir variações na probabilidade de acerto em relação a variação de habilidades dos respondentes.

### 3.6 ITEM 6

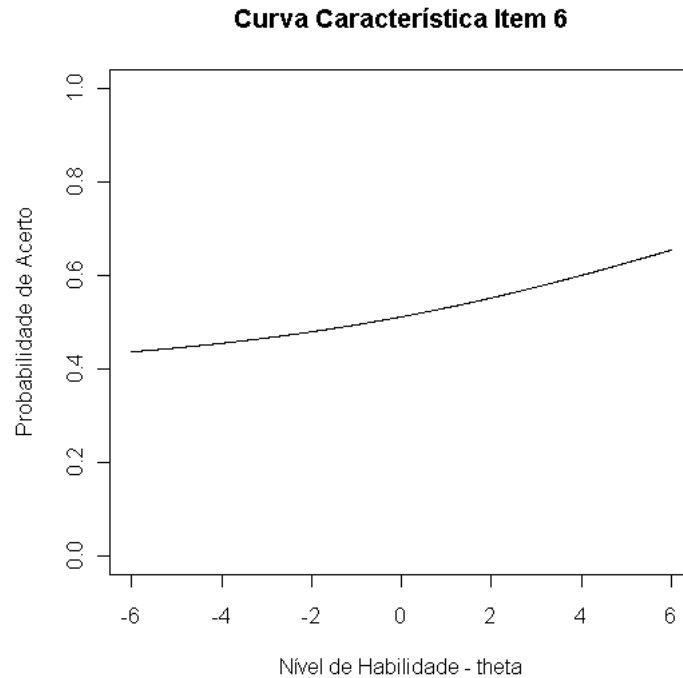


Figura 3.8: Curva Característica Item 6

Como visto na Tabela 3.1 os parâmetros para o Item 6 são :

- Parâmetro referente a Dificuldade ( $b_6$ ) = 7.45
- Parâmetro referente a Discriminação ( $a_6$ ) = 0.189
- Parâmetro referente ao Chute ( $c_6$ ) = 0.392

Assim como o que foi observado no Item 4 o presente item não apresentou o comportamento esperado para uma avaliação educacional. Mais uma vez não foram observadas probabilidades 1 e 0 para a amplitude de habilidades ( $\theta$ ) comumente utilizadas. O tópico abordado para a síntese do referido item deve ser utilizada para



a síntese de outro item que possa apresentar comportamento mais adequado ao que tange a Teoria de Resposta ao Item.

### 3.7 ITEM 7

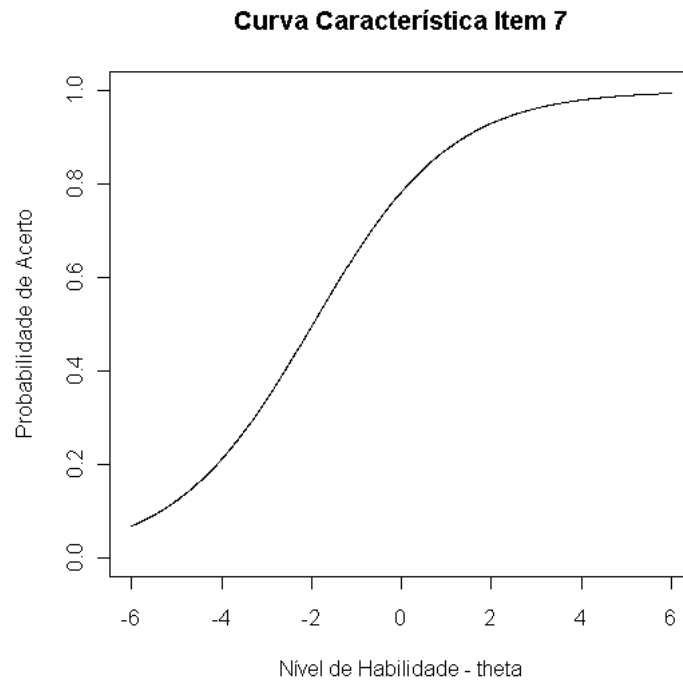


Figura 3.9: Curva Característica Item 7

Como visto na Tabela 3.1 os parâmetros para o Item 7 são :

- Parâmetro referente a Dificuldade( $b_7$ ) = -1.972
- Parâmetro referente a Discriminação( $a_7$ ) = 0.652
- Parâmetro referente ao Chute ( $c_7$ ) = 0

O Item 7 apresenta nível de dificuldade baixo quando comparado com os demais itens presentes no instrumento. Portanto, a carga teórica inserida no item não é

de difícil assimilação pelos alunos, respondentes submetidos ao teste. O parâmetro referente a discriminação é mediano. Já o parâmetro quanto ao acerto ao acaso é igual a zero na estimação com o banco de dados supracitado. O assunto referente a "outliers" deve ser abordado com alguma roupagem distinta para que a dificuldade do item possa ter um incremento e principalmente uma alteração na discriminação do presente item.

### 3.8 ITEM 8

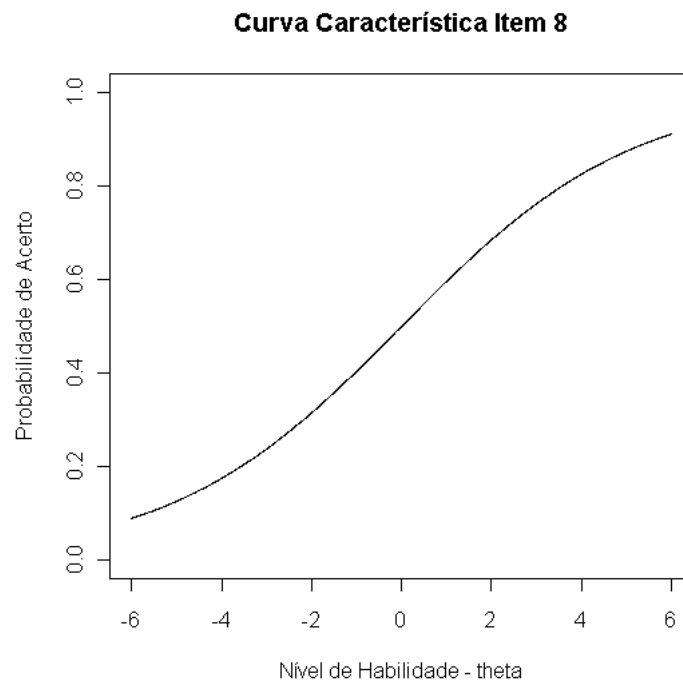


Figura 3.10: Curva Característica Item 8

Como visto na Tabela 3.1 os parâmetros para o Item 8 são :

- Parâmetro referente a Dificuldade ( $b_8$ ) = 0.005
- Parâmetro referente a Discriminação ( $a_8$ ) = 0.389

- Parâmetro referente ao Chute ( $c_8$ ) = 0

O Item 8 apresentou parâmetro relacionado ao acerto casual igual a zero pela estimativa sintetizada. A dificuldade do item é bastante próxima de zero o que indica que é um item relevante para a separação de alunos, respondentes, com habilidades baixas e altas. A curva mostrada na Figura 3.10 é bastante suave, ou seja, não consegue discriminar fortemente variações na habilidade. Portanto, devem ser feitas alterações para que o item possa discriminar com maior poderio.

### 3.9 ITEM 9

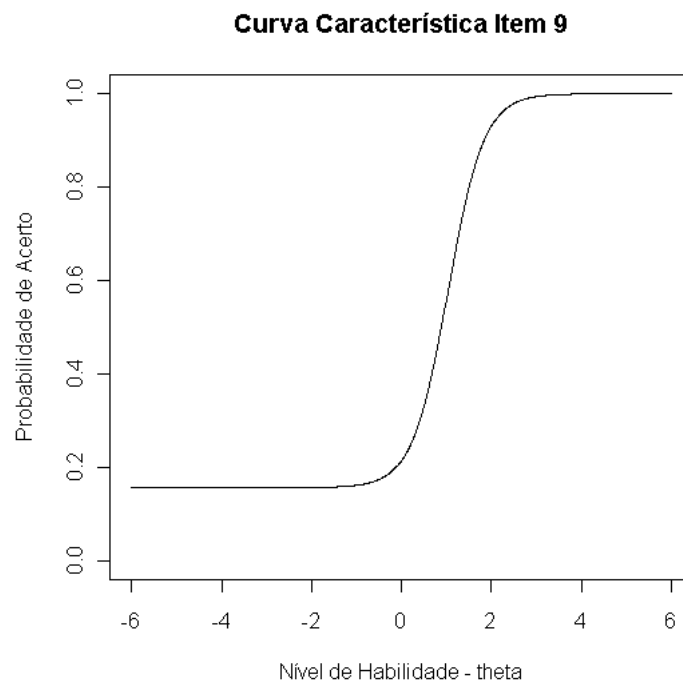


Figura 3.11: Curva Característica Item 9

Como visto na Tabela 3.1 os parâmetros para o Item 9 são :

- Parâmetro referente a Dificuldade ( $b_9$ ) = 1.036

- Parâmetro referente a Discriminação ( $a_9$ ) = 2.510
- Parâmetro referente ao Chute ( $c_9$ ) = 0.165

O Item 9 apresentou parâmetros interessantes para a compreensão da Teoria de Resposta ao Item além de ser o item que melhor resume a finalidade do projeto em questão. Com nível de dificuldade em um patamar ideal, para fins didáticos, não é um item extremamente fácil nem, tão pouco, extremamente difícil. Possui discriminação interessante que pode ser observada na Figura 3.11. O parâmetro referente a probabilidade de acerto mediante o "chute" é baixa e aceitável para uma aplicação onde o objetivo é a estimação das habilidades dos respondentes. O Item 9 conseguiu deter as características necessárias para que possa ser aplicado com o interesse de estimar as habilidades dos alunos. Dos itens propostos nesse projeto foi o que obteve melhor resultado.

### 3.10 CONSIDERAÇÕES FINAIS

A TRI é uma ferramenta que pode ajudar e melhorar a Avaliação Educacional em âmbito geral. Porém esse ferramental não é de uso para momentos específicos. No processo avaliativo deve-se ter a TRI como base em todo o processo, portanto para uma avaliação significativa essa teoria deve nortear todas as atividades do processo. Inicialmente, na síntese dos itens o estatístico já deve estar presente pois é nesse momento que são definidas as maneiras como se relacionará o construto que deseja ser mensurado com a questão para tomada de decisão dos respondentes. Esse processo não é simples nem ao menos pontual. Após o primeiro contato com as questões essas devem ser analisadas por terceiros para que possíveis erros ou problemas sejam detectados, pois a própria pessoa que cria as questões, por vezes, não consegue perceber determinados erros.

Após essa primeira fase, deve-se aplicar um questionário piloto. Essa fase é, por vezes, essencial pois é o primeiro contato que os respondentes, ou seja, quem de fato deve entender, analisar e responder, com as questões que estão sendo utilizadas para a estimação dos parâmetros dessas e das habilidades dos respondentes. Nessa fase do processo também podem ser encontrados erros que não foram percebidos pelo avaliador que produziu o instrumento nem pelos terceiros que analisaram os itens. E, principalmente, nesse momento podem ser feitas alterações cruciais para o desenvolvimento do banco de questões, pois ao sintetizar os itens o avaliador está direcionado para fazer as questões tendo como cerne determinado assunto como, por exemplo, classificação de variáveis. Mas os alunos, respondentes, não necessari-

amente estão direcionados a esse enfoque, então no momento da aplicação de um teste piloto consegue-se inferir se as informações que inicialmente foram desenvolvidas pelo avaliador chegam de maneira eficaz aos respondentes.

A calibração dos itens é a fase inicial do processo avaliativo que utiliza TRI. Porém é um dos dois blocos mais importantes do processo assim como a estimação das habilidades dos respondentes. Naquele momento podem ser tiradas informações de como, realmente, os itens foram interpretados pelos alunos. Aplica-se esse teste, já alterado pelas necessidades mostradas no teste piloto, e assim, consegue-se um banco de dados para a calibração dos itens. Fica evidente que nem todos os itens inicialmente propostos poderão ser utilizados para uma avaliação educacional.

Outros itens comportam-se bem e apresentam resultados relevantes como o Item 9. Então reitera-se a informação de que a TRI, no âmbito da Avaliação Educacional, deve ser um processo contínuo e não pontual. Devem ser desenvolvidas questões que contemplem o construto em questão, aplicar essas questões para um teste, ou seja, aplicar o piloto desse instrumento. Por vezes, nesse momento algumas modificações fazem-se necessárias. Após essas modificações deve-se aplicar os itens para a calibração desses, ou seja, conseguir estimar os parâmetros (dificuldade ( $b_i$ ), discriminação ( $a_i$ ) e probabilidade de acerto mediante escolha casual ( $c_i$ )). Nesse instante percebe-se que alguns problemas ou comportamentos distintos do esperado e devem ser realizadas alterações substanciais ou até mesmo excluir determinados itens que não mostraram bons resultados.

Este ciclo deve continuar até que se formem bancos de questões que possam ser utilizados para posterior estimação das habilidades dos respondentes.

Em suma, sintetizar itens para utilização em TRI é um processo contínuo que demanda análises consecutivas para se chegar a um ponto no qual os itens estejam se comportando de maneira ideal. Portanto o presente projeto é o passo inicial para a síntese de um banco de questões que possa ajudar e incrementar a Avaliação das matérias de serviço que o Departamento de Estatística dispõe para a Universidade de Brasília. A continuação dessas atividades são necessárias para que no futuro a Teoria de Resposta ao Item seja utilizada e o Departamento consiga, de forma sucinta, ter uma compreensão de como os alunos que cursam as matérias de serviço assimilam a parte da matéria denominada como Análise Exploratória de Dados.

# APÊNDICE I

```
bd<- read.table(file.choose(), header = F)

a<- tpm(bd)
summary(a)
theta<-seq(-6,6, by=0.01)

#Item 1
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 1")
q1<- 0.633 + (0.367*(1/(1 + exp(6.33*(0.798-theta))))))
lines(theta, q1)

#Item 2
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 2")
q2<- 0.732 + (0.268*(1/(1 + exp(47.35*(0.305-theta))))))
lines(theta, q2)

#Item 3
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 3")
q3<- 0.506 + (0.494*(1/(1 + exp(2.331*(1.535-theta))))))
lines(theta, q3)

#Item 4
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 4")
q4<- 0.016 + (0.984*(1/(1 + exp(0.207*(2.366-theta))))))
lines(theta, q4)

#Item 5
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 5")
q5<- 0 + (1/(1 + exp(16.983*(-0.743-theta))))
lines(theta, q5)

#Item 6
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 6")
q6<- 0.392 + (0.608*(1/(1 + exp(0.189*(7.45-theta))))))
lines(theta, q6)

#Item 7
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 7")
q7<- 0 + (1/(1 + exp(0.652*(-1.972-theta))))
lines(theta, q7)

#Item 8
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 8")
q8<- (1/(1 + exp(0.389*(0.005-theta))))
```



```

lines(theta, q8)

#Item 9
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Item 9")
q9<- 0.156 + (0.844*(1/(1 + exp(2.51*(1.036-theta))))))
lines(theta, q9)

#CCI's 5 primeiros itens
plot(c(-6,6), c(0,1), col="0", xlab="Nível de Habilidade - theta"
, ylab="Probabilidade de Acerto", main="Curva Característica Itens de 1 a 5")
lines(theta, q1)
lines(theta, q2, col=2)
lines(theta, q3, col=3)
lines(theta, q4, col=4)
lines(theta, q5, col=6)
text(5, 0.35, "Item 1", col=1)
text(5, 0.30, "Item 2", col=2)
text(5, 0.25, "Item 3", col=3)

```

# APÊNDICE II

## Questionário - Análise Exploratória de Dados

Tendo como base os conhecimentos relacionados à Análise Exploratória de Dados responda os itens subsequentes de acordo com o enunciado de cada uma das questões. O tempo para resolução é de 20 minutos.

### Questão 1.

Utilizando o conhecimento relacionado à classificação de variáveis, escolha a alternativa que classifica corretamente a variável a seguir:

Em uma fazenda de criação bovina, fez-se necessária a mensuração do peso dos animais de corte. Essa característica mensurada nos animais pode ser classificada como:

- a) Variável Qualitativa Ordinal
- b) Variável Quantitativa Discreta
- c) Variável Qualitativa Nominal
- d) Variável Quantitativa Contínua

### Questão 2.

Utilizando o conhecimento relacionado à distribuição de frequências, escolha a alternativa que completa corretamente a tabela abaixo:

Em uma pesquisa de Determinantes de Qualidade de Vida, deseja-se saber dos idosos se esses são, ou não, portadores de necessidades especiais. A tabela a seguir mostra as frequências observadas na pesquisa.

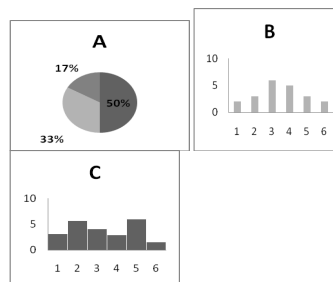
Necessidades Especiais	Frequência Absoluta ( $n_i$ )	Frequência relativa ( $f_i$ )
Portador	64	
Não Portador		
Total	160	1

Complete a tabela acima com a frequência de não portadores, a proporção de portadores e a proporção de não portadores, respectivamente:

- a) 96; 0,4 e 0,6
- b) 69; 0,6 e 0,4
- c) 96; 0,6 e 0,4
- d) 69; 0,4 e 0,6

### Questão 3.

Utilizando os conhecimentos relacionados a gráficos estatísticos, escolha a alternativa que relaciona corretamente os tipos de variáveis com os gráficos adequados para cada variável.



1. Variável Quantitativa Contínua
2. Variável Quantitativa Discreta
3. Variável Qualitativa

Escolha corretamente a alternativa que relaciona corretamente o tipo de gráfico, especificado pelas letras A, B e C, com os tipos de variáveis especificados pelos números 1, 2 e 3.

- a) (A-3), (B-1) e (C-2)
- b) (A-3), (B-2) e (C-1)
- c) (A-2), (B-1) e (C-3)
- d) Nenhuma das alternativas acima.

Para a resolução do exercício 4 utilize os seguintes dados.

Uma pesquisa escolar, a respeito de reprovação dos alunos, mostrou os seguintes dados, para a escola X:

Turma	Número de Reprovados
A	2
B	5
C	6
D	3
Total	16

Uma pesquisa escolar, a respeito de aprovação dos alunos, mostrou os seguintes dados, para a escola Y:

Turma	Número de Reprovados
E	4
F	3
G	2
H	3
Total	12

### Questão 4

Utilizando os conhecimentos relacionados a medidas de dispersão e correlatos. Escolha a alternativa correta.

Medidas de dispersão podem expressar informações mais precisas das características de determinado grupo de dados.

A variância do número de reprovados para a escola X e, para a escola Y, são respectivamente:

- a) 2,5 e 0,5.
- b) 5 e 2,5.
- c) 25 e 6,25.
- d) 6,25 e 5.

### Questão 5

Utilizando os conhecimentos relacionados a medidas de dispersão e correlatos. Escolha a alternativa correta.

A variância é importante medida na análise descritiva de dados, porém um entrave na interpretação desse resultado consiste na unidade de medida, pois a unidade é elevada ao quadrado. Então outra medida foi desenvolvida para resolver esse problema.

A medida desenvolvida e o valor dessa para duas populações com variâncias iguais a 16 e 25 são, respectivamente.

- a) **Coefficiente de Variação; 5 e 6.**
- b) **Desvio Padrão; 4 e 5.**
- c) **Coefficiente de Variação ; 5 e 4.**
- d) **Desvio Padrão; 8 e 1,5..**

### Questão 6

A respeito dos assuntos estudados sobre a medidas-resumo, medidas de dispersão e análise entre populações, escolha a alternativa correta.

Dois populações distintas apresentam diferentes observações de características mensuradas. Algumas medidas podem ser utilizadas para uma possível análise conjunta de duas populações.

Em um Hospital foram colhidos os dados referentes a pressão arterial(mmHg), a temperatura corporal( °C) e a capacidade pulmonar total(Litros) dos pacientes. A medida que pode trazer informações comparativas referentes a variabilidade dessas variáveis é:

- a) **Média.**
- b) **Desvio Padrão**
- c) **Coefficiente de Variação**
- d) **Nenhuma das alternativas**

### Questão 7

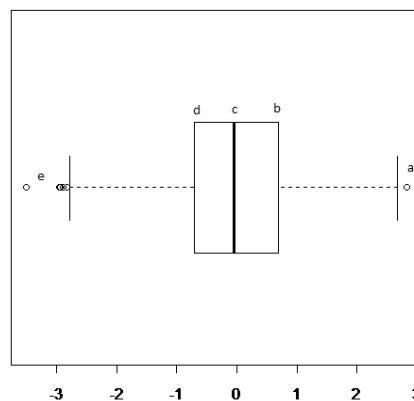
Levando em consideração assuntos correlatos a medidas-resumo, medidas de dispersão e quantis empíricos, escolha a alternativa correta.

A média aritmética e o desvio padrão podem não ser medidas adequadas para representar um conjunto de dados, quando:

- a) **Existem muitos dados no grupo observado**
- b) **A média é o triplo do desvio-padrão.**
- c) **O desvio-padrão é o triplo da média.**
- d) **São afetados, de forma exageradas, por valores extremos(outliers) .**

### Questão 8

Utilizando o conhecimento referente a análises gráficas e representações de distribuições, julgue o item a seguir.



A figura acima representa um diagrama largamente utilizado na análise exploratória de dados. Utilizando essa figura como base, escolha a alternativa que traz uma denominação do diagrama e relaciona corretamente as letras a, b, c, d e e, respectivamente.

- a) **Dotplot, 1º quartil, Moda, Mediana, Limite Inferior e Outliers**
- b) **Box – plot, Outliers, 3º quartil , Mediana, 1º quartil e Outliers.**
- c) **Dotplot, 3º quartil, Mediana, Média, 1º quartil e Outliers.**
- d) **Box – plot, Outliers, Mediana, 1º quartil, 3º quartil e Outliers.**

### Questão 9

Utilizando os conhecimentos referentes à análise bidimensional, julgue a alternativa correta.

As variáveis aleatórias podem ser classificadas como qualitativas e quantitativas. Podem ser observadas associações entre essas, e para uma análise dessa possível associação algumas medidas de associação podem ser utilizadas.

A associação entre variáveis **qualitativas** pode ser mensurada pela medida \_\_\_\_\_ e a associação entre variáveis **quantitativas** pode ser mensurada pela medida \_\_\_\_\_ .

- a) **Coefficiente de Contigência e Coeficiente de Variação.**
- b) **Coefficiente de Correlação de Pearson e Coeficiente de Variação.**
- c) **Coefficiente de Contigência e Coeficiente Correlação de Pearson.**
- d) **Coefficiente de Correlação de Pearson e Coeficiente de Contigência.**

# Capítulo 4

## REFERÊNCIAS

- ALVES, Suely; ANDRADE, Edgard; BICALHO, Ana; CAVEDINI, Sônia; DA CONCEIÇÃO, Gislaine; CORDEIRO, Maria; DE COUTO, Elze; JARDIM, Geralda, LAGE, Amazílis, PEREIRA, Carmelita. **Guia de Elaboração e Revisão de Questões e Itens de Múltipla Escolha**. Minas Gerais,
- PASQUALI, Luiz; PRIMI, Ricardo. **Fundamentos da Teoria da Resposta ao Item - TRI**. Revista Avaliação Psicológica, 2003.
- DE AYALA, Jaime. **The theory and practice of item response theory**. Guilford, New York, 2009.
- BAKER, Frank; Kim, Seock-Ho. **Item Response Theory : Parameter Estimation Techniques, Second Edition**(Statistics: A series of Text-books and Monographs). Dekker, New York, 2004.
- ANDRADE, Dalton; TAVARES, Heliton; VALLE, Raquel.**Teoria de Resposta ao Item : Conceitos e Aplicações**. SINAPE, 2000.
- QUADROS, Fernanda. **Comparação da Teoria Clássica dos Testes com**

- a **Teoria de Resposta ao Item aplicado ao subteste Semelhanças da WASI**. 2010. Monografia (Bacharelado em Estatística) - Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.
- GATTI, Bernardete. **Avaliação Educacional no Brasil**: pontuando uma história de ações. Ecoos Revista Científica, São Paulo, 2002.
  - SOMERA, Elizabeth. **Reflexões sobre vertentes da Avaliação Educacional**. Avesso do Avesso, Araçatuba, 2008.
  - VIANNA, Heraldo. **Avaliações Nacionais em Larga Escala**: análises e propostas. Estudos em Avaliação Educacional, 2003.