
Teoria e prática de um Teste Adaptativo
Informatizado

Gilberto Pereira Sassi

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Teoria e prática de um Teste Adaptativo Informatizado

Gilberto Pereira Sassi

Orientadora: Profa. Dra. Mariana Cúri

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

USP – São Carlos
Maio de 2012

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

S252t Sassi, Gilberto
 Teoria e a prática de um teste adaptativo
 informatizado / Gilberto Sassi; orientadora Mariana
 Cúri. -- São Carlos, 2012.
 76 p.

 Tese (Doutorado - Programa de Pós-Graduação em
 Ciências de Computação e Matemática Computacional) --
 Instituto de Ciências Matemáticas e de Computação,
 Universidade de São Paulo, 2012.

 1. Teste adaptativo informatizado. 2. Algoritmo
 de seleção. 3. modelo logístico unidimensional de três
 parâmetros. 4. Teoria de Resposta ao Item. I. Cúri,
 Mariana, orient. II. Título.

À minha família e amigos.

Agradecimentos

Primeiramente agradeço à minha família por sempre me apoiar e dar incentivo .

Aos meus amigos da república Duendes Verdes, pelos bons momentos vividos em São Carlos.

Aos meus amigos de mestrado, em especial Alessandra Sibim e Daiane de Souza que me ajudaram a redigir este texto lendo e dando dicas valiosas.

À minha orientadora, Profa.Dra. Mariana Cúri, pelo incentivo e condução durante a elaboração desta dissertação.

Ao ICMC - USP pela infraestrutura que permitiu que realizasse os estudos e atividades para elaborar este trabalho.

Finalmente agradeço à CAPES pelo suporte financeiro.

“A vida é feita pra cair mil vezes e outras mil vezes se levantar.”

Resumo

O objetivo deste trabalho é apresentar os conceitos relacionados a Teste Adaptativo Informatizado, ou abreviadamente TAI, para o modelo logístico unidimensional da Teoria de Resposta ao Item. Utilizamos a abordagem bayesiana para a estimação do parâmetro de interesse, chamado de traço latente ou habilidade. Apresentamos os principais algoritmos de seleção de itens em TAI e realizamos estudos de simulação para comparar o desempenho deles. Para comparação, usamos aproximações numéricas para o Erro Quadrático Médio e para o Vício e também calculamos o tempo médio para o TAI selecionar um item. Além disso, apresentamos como instalar e usar a implementação de TAI desenvolvida neste projeto chamada de TAI2U, que foi desenvolvido no VBA-Excel usando uma interface com o R.

Palavras chaves: Teste Adaptativo Informatizado, Algoritmo de seleção de item, modelo logístico unidimensional, Teoria de Resposta ao Item.

Abstract

The main of this work is to introduce the subjects related to Computerized Adaptive Testing, or briefly CAT, for the unidimensional three-parameter logistic model of Item Response Theory. We use bayesian approach to estimate the parameter of interest. We present several item selection algorithms and we perform simulations comparing them. The comparisons are made in terms of the mean square error, bias of the trait estimates, the average time for item selection and the average length of test. Furthermore, we show how to install e use the CAT implementation of this work called built in MICROsoft Excel - VBA using interface with the statistical package R.

Keywords: Computerized Adaptive Testing, Item selection algorithms, unidimensional three-parameter logistic model, Item Response Theory.

Lista de Figuras

2.1	Curva Característica do Item - CCI para um ML3	12
4.1	Erro padrão do estimador para um banco com 50 itens no final do TAI . .	47
4.2	Erro padrão do estimador para um banco com 300 itens no final do TAI . .	48
4.3	Erro padrão do estimador para um banco com 700 itens no final do TAI . .	49
5.1	Passo 3) - Instrução número 1.	55
5.2	Passo 3) - Instrução número 2.	55
5.3	Passo 3) - Instrução número 3.	56
5.4	Passo 3) - Instrução número 4.	57
5.5	Passo 3) - Instrução número 5.	57
5.6	Passo 3) - Instrução número 6.	58
5.7	Passo 3) - Instrução número 7.	59
5.8	Passo 3) - Instrução número 8.	59
5.9	Passo 3) - Instrução número 9.	60
5.10	Passo 6) - Instrução número 1.	61
5.11	Passo 6) - Instrução número 2.	61
5.12	Passo 6) - Instrução número 3.	62

5.13 Passo 6) - Instrução número 4.	62
5.14 Passo 6) - Instrução número 5.	63
5.15 Passo 6) - Instrução número 6.	63
5.16 Passo 6) - Instrução número 7.	64
5.17 RExcel instalado com sucesso.	65
5.18 Instalando o TAI - Instrução número 1.	66
5.19 Instalando o TAI - Instrução número 2.	67
5.20 Configurando o TAI - Instrução número 1.	68
5.21 Configurando o TAI - Instrução número 2.	69
5.22 Configurando o TAI - Instrução número 3.	70
5.23 Configurando o TAI - Instrução número 4.	71
5.24 Executando o TAI - Instrução número 1.	72
5.25 Executando o TAI - Instrução número 2.	73

Lista de Tabelas

4.1	EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 50 itens.	41
4.2	EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 300 itens.	42
4.3	EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 700 itens.	43
4.4	Resultados da simulação para o tempo de seleção em segundos para um banco de 50 itens	44
4.5	Resultados da simulação para o tempo de seleção em segundos para um banco de 300 itens	45
4.6	Resultados da simulação para o tempo de seleção em segundos para um banco de 700 itens	46

Sumário

1	Introdução	1
1.1	Desenvolvimento do banco de itens e tópicos relacionados a segurança	4
1.2	Algoritmos de seleção	6
1.3	Organização dos capítulos	10
2	Teoria de Resposta ao Item	11
2.1	Curva Característica do Item	11
2.2	Modelo logístico de três parâmetros	13
2.3	Estimação	13
2.3.1	Notação	14
2.3.2	Estimador de máxima verossimilhança	15
2.3.3	Problemas com o estimador de máxima verossimilhança	15
2.3.4	Abordagem bayesiana	16
3	Algoritmos de seleção	18
3.1	Notação	19
3.2	Critério da Máxima Informação de Fisher	20
3.3	Critério de Owen	22

3.4	Algoritmos ponderados pela distribuição <i>posteriori</i>	23
3.4.1	Critério da Máxima Informação ponderada pela distribuição a <i>posteriori</i>	23
3.4.2	Distribuição preditiva em TAI	24
3.4.3	Critério da máxima informação esperada	26
3.4.4	Critério da Máxima Informação Esperada ponderada pela <i>posteriori</i>	28
3.4.5	Critério da Mínima Variância a <i>posteriori</i>	29
3.5	Algoritmo do Critério da Máxima Informação Global	30
3.5.1	Informação de Kullback-Leibler	31
3.5.2	Critério da Máxima Informação Global	34
4	Resultados	36
4.1	Medidas para comparar algoritmos de seleção	36
4.2	Estudos de simulação	38
4.3	Resultados da simulação	39
4.4	Apresentação de resultado	51
4.5	Implementação	51
5	Instalando o TAI2U	53
5.1	Como Instalar	54
5.1.1	Primeira fase	54
5.1.2	Segunda fase	64
5.2	Como configurar	67
5.3	Como executar	70

Referências 74

Capítulo 1

Introdução

Em diversas situações, temos como principal objetivo conhecer alguma característica não-observável, tal como conhecimento em uma determinada disciplina na área educacional e presença ou intensidade de uma determinada patologia ou comportamento na área da saúde (Fragoso, 2010). Essa característica não-observável é denominada traço latente ou habilidade. Para mensurar o traço latente, a abordagem tradicional é aplicar um teste e quantificar a habilidade pelo escore, ou seja, o número de respostas corretas.

Apesar da abordagem tradicional ser simples e direta, podemos citar algumas brechas dessa abordagem (Wainer, 2001):

1. O teste deve ser capaz de avaliar toda a população, sendo assim é necessário ter itens ¹ fáceis para indivíduos pouco hábeis, itens com dificuldade média para valores médios do traço latente e itens difíceis para os respondentes mais hábeis. Nesse contexto, um teste pode ser composto por muitos itens;
2. É difícil determinar o número adequado de itens em um teste na abordagem tradicional;
3. Indivíduos podem ser obrigados a responder itens que não correspondem ao seu nível de habilidade;

¹Segundo a notação da literatura, item é uma pergunta do teste.

4. Existe um tempo de espera pelo resultado do teste;
5. Testes muito longos podem causar fadiga e produzir respostas não confiáveis.

Para cobrir essas brechas, os Testes Adaptativos Informatizado, ou de forma abreviada TAI, surgiram na década de 1970 e 1980 (Meijer & Nering, 1999). O TAI é um método de realizar avaliações em que itens são apresentados via computador. Após cada resposta, a estimativa do traço latente é recalculada e um novo item é selecionado, tal item é o mais adequado para a estimativa interina da habilidade. Notemos que em TAI é necessário termos um *software* em que o teste é realizado e ainda é necessário um banco de itens com parâmetros da Teoria de Resposta ao Item conhecidos ². Segundo Van Der Linden (2000), podemos descrever um Teste Adaptativo Informatizado pelos seguintes passos:

Passo 1) O primeiro item é selecionado;

Passo 2) Estimamos o traço latente baseado na resposta ao primeiro item, denotamos este valor por $\hat{\theta}_1$;

Passo 3) Escolhemos o próximo item a ser respondido no TAI. Tal item deve ser “o mais adequado” para a estimativa pontual interina da habilidade, indicada por $\hat{\theta}_{k-1}$;

Passo 4) Recalculamos a estimativa do traço latente baseado nas respostas anteriores, denotamos esse valor por $\hat{\theta}_k$;

Passo 5) Repita os passos 3 e 4 até não ser mais necessária nenhuma resposta, segundo algum critério pré-estabelecido, chamado de critério de parada.

Segundo Wainer (2000), para o **Passo 1)** existem duas possibilidades principais: sortear o primeiro item ou usar a informação *a priori*. As duas abordagens têm vantagens e desvantagens: sortear o primeiro item não utiliza a informação que já temos sobre o traço latente, mas evita que o mesmo item seja selecionado como primeiro item a cada

²Para uma apresentação da Teoria de Resposta ao Item vide Capítulo 2 e para maiores detalhes vide Hambleton & Swaminathan (1989).

seção do TAI, o que não é ideal, por exemplo, em situações onde o TAI é utilizado como classificatório. Usar a distribuição *a priori* faz com que sempre utilizemos toda informação disponível em todas as etapas de um TAI, contudo pode ocorrer que um item, ou conjunto de itens, seja selecionado de forma recorrente para ser o primeiro. Nesse trabalho, decidimos usar a segunda abordagem, pois as aplicações desenvolvidas não serão utilizadas para estabelecer qualquer tipo de classificação e é ideal usar toda informação disponível para selecionar um item. Na seção 1.2 e no Capítulo 3, explicamos como utilizar a distribuição *a priori* para selecionar o primeiro item.

Para o **Passo 2)** e o **Passo 4)**, utilizamos a abordagem bayesiana de estimação, pois, segundo Andrade *et al.* (2000) e Van Der Linden (2000), a abordagem clássica tem algumas vantagens no contexto de Teste Adaptativo Informatizado (vide Capítulo 2 para maiores detalhes).

O **Passo 3)**, denominado algoritmo de seleção, exerce um papel central em TAI. Ele estabelece como escolher o próximo item que será respondido baseado nas respostas anteriores. No Capítulo 3 apresentamos os algoritmos de seleção estudados nesse trabalho e na Seção 1.2 faremos uma breve revisão bibliográfica dos algoritmos de seleção.

O **Passo 5)**, chamado de critério de parada, determina o momento em que nenhum item será selecionado. De acordo com Van Der Linden (2000), existem três possíveis alternativas para critério de parada:

- i) O teste tem tamanho fixo, ou seja, todos os indivíduos responderão N itens, isto é, o TAI não apresentará mais questões quando N itens forem respondidos;
- ii) Nenhum item será selecionado quando o erro padrão do estimador da habilidade for pequeno, isto é, quando o erro padrão do estimador do traço latente for que δ ³ nenhum item será selecionado. Notemos que nessa abordagem indivíduos podem ter testes de tamanhos distintos;
- iii) Esse critério é uma junção de i) e ii). Ou o indivíduo responderá N itens ou o erro

³O valor δ é estabelecido pelo construtor do teste.

padrão do estimador para o traço latente é menor que δ , isto é, o TAI não apresentará mais itens quando o critério de parada **i)** ou o critério de parada **ii)** for alcançado.

Segundo Linden & Pashley (2010), pode ocorrer no critério de parada **ii)**, que o erro padrão não fique pequeno. Por esse motivo, nesse trabalho optamos pela terceira abordagem.

Segundo Meijer & Nering (1999), existem dois tópicos atuais de pesquisa em TAI:

- i)** Desenvolvimento do banco de itens e tópicos relacionados a segurança;
- ii)** Algoritmos de seleção.

Na seqüência, discutiremos sobre esses tópicos.

1.1 Desenvolvimento do banco de itens e tópicos relacionados a segurança

Segundo Van Der Linden & Veldkamp (2004), o desenvolvimento de TAI's operacionais mostrou problemas desconhecidos até então. Um deles é a freqüência que um item é utilizado pelo TAI. Verificou-se na prática que alguns itens eram selecionados com bastante freqüência e outros nem chegavam a ser selecionados. Isto é ruim, pois o desenvolvimento de um item é custoso e o uso excessivo ou o não-uso deste é um desperdício de recurso indesejável. Davey & Nering (1998) discutem esse problema extensivamente e descrevem dois modos principais de tratá-los:

- i)** Controlar a freqüência que um item é selecionado, denominada taxa de exposição do item;
- ii)** Gerenciar o banco de itens.

Segundo Meijer & Nering (1999), o objetivo dos algoritmos de seleção que têm controle da taxa de exposição dos itens é limitar o uso dos itens pela frequência com que os itens são respondidos. Sympson & Hetter (1985) propuseram um método em que a cada item é atribuído um parâmetro de controle da taxa de exposição e cada vez que um item é selecionado, gera-se um número aleatório ρ entre 0 e 1 e o item é respondido somente se ρ é menor que o parâmetro de controle da taxa de exposição. Van Der Linden & Veldkamp (2007) propuseram uma melhoria no método de Sympson & Hetter (1985) condicionando o parâmetro de controle de exposição ao valor do traço latente.

Segundo Van Der Linden (2000), apesar dos métodos de controle da taxa de exposição de itens limitarem o uso excessivo de itens, eles não implicam em um uso uniforme do banco de itens. Na verdade, implementações de TAI com controle de taxa de exposição mostram que grandes porções do banco são raramente usadas. O principal motivo para esse fato é que tais itens contribuem muito pouco para adaptação ao indivíduo no algoritmo de seleção de itens. Como a produção de um item é um longo e custoso processo, item com pouco uso é um desperdício indesejado de recurso. Assim, uma abordagem cabível é evitar o problema como um todo e construir um banco de itens planejado que seja usado de forma uniforme pelos respondentes.

Flaugher (1990) propõe usar simulações para avaliar o uso dos itens depois do banco já ter sido construído. Boekkooi-Timminga (1991) propôs utilizar a técnica de programação linear inteira para determinar um conjunto de especificações, chamado de *blueprint*, para construir o banco de itens para o modelo logístico de um parâmetro⁴. O banco de itens construído usando o *blueprint* apresenta um uso mais uniforme dos itens. Van Der Linden *et al.* (2000) também propôs utilizar programação linear inteira para produzir um *blueprint*, contudo a abordagem proposta permite que o banco tenha itens calibrados sob vários modelos da Teoria de Resposta ao Item⁵.

De acordo com Van Der Linden & Veldkamp (2004), embora o TAI tenha sido inicialmente motivado pela intenção de adaptar o teste ao indivíduo, os primeiros programas

⁴Para a definição do modelo logístico de um parâmetro vide Hambleton & Swaminathan (1989).

⁵Vide o Capítulo 2 para uma descrição da Teoria de Resposta ao Item e para maiores detalhes vide Weiss & Yoes (1991).

de TAI baseados somente nesse princípio conduziram a resultados irrealistas. Por exemplo, se os itens são selecionados somente para adaptar os itens ao respondente, o conteúdo pode facilmente tornar-se desequilibrado para alguns níveis de θ . Se os respondentes aprenderem sobre essa característica, eles podem mudar sua preparação para o teste e os resultados não serão mais válidos. Logo, um TAI somente será aceitável se o princípio de adaptação dos itens ao indivíduo for implementado em conjunção com outras especificações não-estatísticas que impõe uma restrição no algoritmo de seleção. Dessa forma, um TAI deve combinar adaptação dos itens aos respondentes e obediência a várias restrições não-estatísticas.

Van Der Linden & Veldkamp (2004) propuseram solucionar esse problema através de uma série de testes intermediários, chamados de “*shadow-test*”, criada a partir de soluções de problemas de programação linear inteira (para maiores detalhes, vide Van Der Linden (2000) e Van Der Linden & Veldkamp (2004)). Kingsbury & Zara (1989) sugeriram particionar o banco de itens em grupos de acordo com especificações não-estatísticas. Quando o TAI está em execução, os números de itens em cada grupo são gravados. Para satisfazer as especificações não-estatísticas, o algoritmo seleciona o próximo item do grupo com a menor quantidade de itens selecionados. Wainer & Kiely (1987) propuseram modificar o banco de itens: em vez de itens discretos, o TAI selecionaria feixes de itens, denominados “*testlet*”. Os *testlets* são tratados como unidades a serem respondidas pelo indivíduo e são pré-montados para satisfazer especificações não-estatísticas.

1.2 Algoritmos de seleção

A principal idéia subjacente em TAI é a “adaptação” do TAI ao respondente, que é realizada pelo **algoritmo de seleção de item (ou critério de seleção)**. Por algoritmo de seleção entendemos o processo de escolher um item ainda não respondido baseado nas respostas anteriores. Tendo isso em mente, podemos, segundo Wainer (2001), estabelecer dois momentos no algoritmo de seleção: Primeiro, quando nenhum item foi respondido, ou seja, a escolha do primeiro item (na seção 1, chamamos essa etapa de

Passo 1)). Segundo, quando temos $k - 1$ (> 0) itens respondidos e queremos determinar qual o “melhor” item para o indivíduo responder. Notemos que essa escolha é baseada nas $k - 1$ primeiras respostas. Em cada um dos momentos, a escolha do item pode ser feita de uma maneira distinta (vide Capítulo 3 para maiores detalhes). A seguir, descrevemos, em ordem cronológica, os principais algoritmos de seleção encontrados na literatura para o modelo logístico unidimensional de três parâmetros ⁶ da Teoria de Resposta ao Item. Para uma revisão mais completa de algoritmos de seleção vide Van Der Linden (2010).

1968 Birnbaum (1968) propôs o Critério da Máxima Informação de Fisher. Essa abordagem é motivada pelo fato de que a variância do estimador de máxima verossimilhança é assintoticamente igual a recíproca da Informação de Fisher (Casella & Berger, 2001). Apesar desse critério ter sido motivado para uma abordagem clássica, ele também é muito usado em conjunção com métodos bayesianos (Linden & Pashley, 2010). Para maiores detalhes vide Seção 3.2.

1969 Owen (1969) foi o primeiro a utilizar a distribuição *a posteriori* no algoritmo de seleção. Este critério foi formulado numa época em que o poder computacional era reduzido e aproximações para a média e a variância *a posteriori* eram utilizadas. A sugestão de Owen (1969) foi usar a estimação pontual da habilidade para realizar a “adaptação”: o item escolhido é aquele que tem o parâmetro de dificuldade ⁷ mais próximo da estimativa pontual interina da habilidade. Uma crítica a essa abordagem é a utilização apenas do parâmetro de dificuldade para selecionar o item sem nenhuma tentativa de melhorar o processo com outras informações, como por exemplo o uso do parâmetro de discriminação ⁸. Para uma descrição completa desse método vide Seção 3.3.

1975 Owen (1975) propôs uma melhoria no algoritmo de seleção proposto em Owen (1969). Contudo, o poder computacional disponível na década de 1970 limitava o uso bem-sucedido da abordagem proposta: o item escolhido é aquele que faz a variância a

⁶Para a definição do modelo logístico unidimensional de três parâmetros vide o Capítulo 3.

⁷Para a definição de parâmetro de dificuldade vide Capítulo 2.

⁸Para a definição do parâmetro de discriminação vide 2.

posteriori ficar menor. Para maiores detalhes vide Seção 3.4.5.

1983 McBride & Martin (1983) propuseram um algoritmo de seleção que atendesse duas demandas: adaptar o item ao respondente e resolver o problema da taxa de exposição dos itens apresentado na Seção 1.1. A proposta deles foi selecionar um conjunto de k itens em vez de um único item, este conjunto é composto pelos k itens mais “adaptados”⁹ ao respondente. Deste conjunto, sorteamos um item e o escolhemos como o primeiro item no TAI, após este item ser respondido escolhemos um item não-respondido de forma aleatória de tal conjunto como segundo item e o processo continua até todos os itens deste conjunto serem respondidos e então escolhemos um novo conjunto.

1996 Chang & Ying (1996) também enunciaram um algoritmo de seleção que tenta atender duas demandas: “adaptar” os itens ao respondente e controlar a taxa de exposição conforme explicado na Seção 1.1. Eles basearam a proposta na seguinte afirmação: “Itens com valor alto no parâmetro de discriminação são valiosos, pois eles geralmente têm valor maior da Informação de Fisher, isto é, fazem o erro padrão do estimador ficar menor”. Chang & Ying (1996) alegam que, quando utilizamos o algoritmo de seleção proposto por Birnbaum (1968), escolher itens com valores grandes do parâmetro de discriminação no começo do teste, quando temos pouca informação sobre o traço latente, agrava o problema da taxa de exposição de item descrito na seção 1.1. Assim, sugeriram usar uma medida baseada na Informação de Kullback-Leibler, que nos fala quão bem um item distingue a estimativa interina da habilidade e o valor real do traço latente. O algoritmo de seleção escolhe o item com o maior valor na medida de Kullback-Leibler. Uma explicação detalhada deste algoritmo de seleção está na seção 3.5.

1997 Veerkamp & Berger (1997) propuseram uma melhoria na proposta de Birnbaum (1968). A motivação para este algoritmo de seleção foi de não ser ideal escolher o item com o maior valor da Informação de Fisher na estimativa pontual interina da

⁹McBride & Martin (1983) utilizaram a Informação de Fisher para determinar os mais “adaptados”.

habilidade no início do teste, pois há pouca informação sobre o traço latente com a estimativa imprecisa e vários outros valores têm substancial chance de ser o valor real da habilidade. Notemos que desejamos escolher o item com o maior valor na Informação de Fisher no valor real (ou próximo) da habilidade e não em um valor “distante”. Assim, Veerkamp & Berger (1997) propuseram usar a função de verossimilhança como medida de plausibilidade de um valor ser a habilidade verdadeira em vez de selecionar o item com máxima Informação de Fisher na estimativa pontual interina do traço latente. Dessa maneira, escolhemos o item cuja Informação de Fisher ponderada pela verossimilhança e integrada sobre um intervalo finito é máxima. Neste trabalho, substituímos a função de verossimilhança pela densidade da distribuição *a posteriori*. Para uma elucidação deste algoritmo de seleção vide Capítulo 3.

1998 No passado, o poder computacional limitava o uso da distribuição *a posteriori* sem aproximações no algoritmo de seleção. Isso já não é mais um empecilho e Van Der Linden (1998) faz uma revisão da abordagem proposta por Owen (1975) e sugere vários novos algoritmos de seleção. Vide Capítulo 3 para descrição completa destes.

1999 Seguindo a mesma motivação de Chang & Ying (1996), em que a escolha de valores grandes do parâmetro de discriminação no começo do teste é contraproducente quando usamos a Informação de Fisher no algoritmo de seleção, Chang & Ying (1999) propuseram criar grupos de itens, cada grupo com valores do parâmetro de discriminação parecidos. No começo do teste, escolhemos os itens do grupo com os menores valores do parâmetro de discriminação, e quanto mais itens forem respondidos escolhemos itens de grupos com valores maiores do parâmetro de discriminação.

Neste trabalho analisamos os algoritmos de seleção propostos na seção 1.2, com exceção das abordagens de McBride & Martin (1983) e Chang & Ying (1999). Para comparar o desempenho dos algoritmos de seleção, usamos uma aproximação numérica para o erro quadrático médio e para o vício, usamos também o tempo médio para selecionar um item e o tamanho médio de um teste, para maiores detalhes vide Capítulo 4. São apresentados estudos de simulação e uma implementação dos algoritmos de seleção

propostos por Birnbaum (1968), Owen (1969) e Chang & Ying (1996) é efetuada. A implementação é realizada através da interface entre o MS-Excel e o *software* R proposta por Baier & Neuwirth (2007).

1.3 Organização dos capítulos

No capítulo 2 apresentamos a Teoria de Resposta ao Item, o modelo logístico unidimensional de três parâmetros usado nesse trabalho e estimadores para habilidade. No capítulo 3 mostramos de forma detalhada o conceito de algoritmo de seleção e expomos todos os algoritmos de seleção analisados neste trabalho. No Capítulo 4 explicamos como foi realizado o estudo de simulação, apresentamos as aproximações utilizadas para o erro quadrático médio e o vício, explicamos como obtivemos o tempo médio de seleção de um item e o tamanho médio de um teste e expomos as conclusões feitas com base nos resultados obtidos. Por fim, no Capítulo 5 apresentamos a implementação de TAI desenvolvida com instruções de como instalar e utilizar.

Capítulo 2

Teoria de Resposta ao Item

Neste Capítulo, apresentaremos uma introdução à Teoria de Resposta ao Item (abreviadamente TRI) e, de maneira especial, o modelo logístico unidimensional de três parâmetros, que também indicaremos de forma abreviada como ML3, e estimadores para a habilidade subjacente ao ML3.

Segundo Ostini & Nering (2006), o princípio matemático da Teoria de Resposta ao Item é uma função que relaciona a probabilidade de um indivíduo acertar um item com o seu nível de habilidade. Em outras palavras, a função descreve através de probabilidade como um indivíduo com um valor alto do traço latente tem mais chances de acertar um item do que um indivíduo com um valor baixo da habilidade. Em TRI, existem vários modelos para esta função (vide Weiss & Yoes (1991) e Baker (2004) para uma revisão dos modelos da TRI), porém neste trabalho usamos o modelo logístico de três parâmetros descrito neste Capítulo.

2.1 Curva Característica do Item

Segundo Baker (2001), a Curva Característica do Item, ou abreviadamente CCI, é a pedra angular do ML3 e todos os construtos do modelo dependem da CCI. Existem três propriedades que descrevem a CCI. A primeira é a dificuldade ou parâmetro de

dificuldade que descreve como os itens são posicionados entre os valores da habilidade. Por exemplo, um item com um valor pequeno do parâmetro de dificuldade é apropriado para valores baixos da habilidade e um item com valor grande da dificuldade é apropriado para valores altos do traço latente. A segunda propriedade é a discriminação ou parâmetro de discriminação, que descreve quão bem um item diferencia entre indivíduos com habilidade acima e abaixo do parâmetro de dificuldade. Essa propriedade é essencialmente a curvatura da CCI no ponto de inflexão: quanto maior a curvatura, maior a discriminação entre indivíduos. Já a terceira propriedade é o parâmetro de acerto casual ou parâmetro de “chute”, que é a probabilidade de um indivíduo acertar um item sem nenhum conhecimento. A Figura 2.1 é um exemplo de CCI para o ML3 (vide DeMars (2010) para outros modelos da TRI).

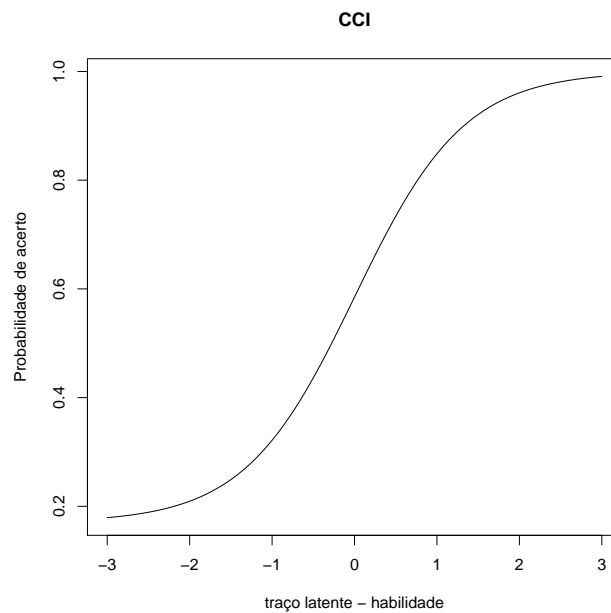


FIGURA 2.1: Curva Característica do Item - CCI para um ML3

Pela Figura 2.1, percebemos que quanto maior a habilidade do indivíduo, maior a probabilidade de acerto e que essa relação não é linear. De fato, a CCI tem forma de “S” com inclinação proporcional ao parâmetro de discriminação (Andrade *et al.*, 2000).

2.2 Modelo logístico de três parâmetros

Até o momento, apresentamos a Curva Característica do Item e as propriedades que a descrevem. Isto é útil para termos um entendimento intuitivo do modelo, mas falta a precisão e rigor para estimarmos a habilidade. Assim, expomos a equação matemática do modelo logístico de três parâmetros nesta seção.

O ML3 foi proposto por Lord *et al.* (1968) e descreve a probabilidade de um indivíduo acertar um item por

$$P(\theta) = c + (1 - c) \frac{1}{1 + \exp(-a(\theta - b))},$$

em que

a é o parâmetro de discriminação;

b é o parâmetro de dificuldade;

c é o parâmetro de acerto casual;

θ é o valor do traço latente do indivíduo.

É importante notar que o parâmetro de acerto casual não é função da habilidade e, segundo Baker (2001), valores maiores que 0.35 não são aceitáveis ¹. Note também que o parâmetro de discriminação é positivo, pois a CCI sempre é crescente (Andrade *et al.*, 2000).

2.3 Estimação

Uma das etapas fundamentais em um TAI é a estimação da habilidade. Como visto na Seção 2.2, a probabilidade de um indivíduo acertar um item depende da habilidade, do

¹Supondo que os itens tenham quatro alternativas.

parâmetro de dificuldade, do parâmetro de discriminação e do parâmetro de acerto casual. Contudo, no contexto de TAI, conhecemos a dificuldade, a discriminação e o parâmetro de “chute”, então precisamos estimar apenas a habilidade. Nesta seção, apresentamos o estimador de máxima verossimilhança e o estimador bayesiano para o traço latente.

2.3.1 Notação

Aqui, estabelecemos algumas notações usadas em todo o texto:

- os itens são indicados por i , com $i = 1, \dots, I$, sendo o I o número de itens no banco de itens;
- a posição dos itens no TAI é denotado por k , com $k = 1, \dots, N$, sendo N o número máximo de itens que o indivíduo responderá;
- i_k representa o k -ésimo item respondido no TAI;
- U_{i_k} é a variável dicotômica, chamada de variável resposta, que representa a resposta ao k -ésimo item do TAI com $U_{i_k} = 1$ se o indivíduo acerta o item e $U_{i_k} = 0$ se ele erra;
- u_{i_k} é a observação da variável U_{i_k} ;
- $\zeta_{i_k} = (a_{i_k}, b_{i_k}, c_{i_k})$ é um vetor contendo parâmetros do k -ésimo item, chamado de vetor de parâmetros do item ou simplesmente parâmetros do item, em que a_{i_k} é o parâmetro de discriminação do k -ésimo item, b_{i_k} é o parâmetro de dificuldade do k -ésimo item e c_{i_k} é parâmetro de acerto casual do item i_k ;
- $P(U_{i_k} = 1 \mid \theta, \zeta_{i_k})$ representa a probabilidade de um indivíduo com habilidade θ acertar um item i_k com parâmetros ζ_{i_k} segundo o ML3, ou seja,

$$P(U_{i_k} = 1 \mid \theta, \zeta_{i_k}) = c_{i_k} + (1 - c_{i_k}) \frac{1}{1 + \exp(-a_{i_k}(\theta - b_{i_k}))} ;$$

- Quando não houver ambiguidade, representaremos $P(U_{i_k} = 1 \mid \theta, \zeta_{i_k})$ simplesmente por $P_{i_k}(\theta)$.

2.3.2 Estimador de máxima verossimilhança

Antes de apresentar o estimador de máxima verossimilhança, ou abreviadamente EMV, definimos a função de verossimilhança (Andrade *et al.*, 2000) após o indivíduo responder k itens:

$$L(\theta \mid u_{i_1}, \dots, u_{i_k}) = \prod_{j=1}^k (P_{i_j}(\theta))^{u_{i_j}} (1 - P_{i_j}(\theta))^{1-u_{i_j}} \quad (2.1)$$

O estimador de verossimilhança é o ponto de máximo da função descrita pela equação 2.1. Não há solução analítica para encontrar o ponto de máximo e métodos numéricos são utilizados para encontrar o EMV (Andrade *et al.*, 2000).

2.3.3 Problemas com o estimador de máxima verossimilhança

De acordo com Andrade *et al.* (2000), alguns problemas ocorrem na estimação por máxima verossimilhança. Se o indivíduo obtém escore nulo, isto é, se $u_{i_1} = 0, \dots, u_{i_k} = 0$, então a verossimilhança resume-se a

$$L(\theta \mid u_{i_1}, \dots, u_{i_k}) = \prod_{j=1}^k (1 - P_{i_j}(\theta)).$$

Pela CCI (vide Figura 2.1), percebemos que

$$(1 - P_{i_j}(\theta)), \quad j \in \{1, \dots, k\}$$

são sempre decrescentes em θ , logo $L(\theta \mid u_{i_1}, \dots, u_{i_k})$ é decrescente em θ e o estimador de máxima verossimilhança é $\theta = -\infty$. Por outro lado, se o indivíduo acerta todas as questões que lhe são apresentadas (diz-se nesse caso que o indivíduo tem escore perfeito), a verossimilhança se resume a

$$L(\theta \mid u_{i_1}, \dots, u_{i_k}) = \prod_{j=1}^k P_{i_j}(\theta)$$

Pela CCI (vide Figura 2.1), percebemos que

$$P_{i_j}(\theta), \quad j \in \{1, \dots, k\}$$

são sempre crescente em θ , logo $L(\theta \mid u_{i_1}, \dots, u_{i_k})$ é crescente em θ e o estimador de verossimilhança é $\theta = \infty$.

Notemos que em TAI a chance de um examinado ter o escore nulo ou perfeito é grande, principalmente no início do TAI, quando poucos itens foram respondidos. Para contornar esse problema, nesse trabalho usamos a abordagem bayesiana para estimar a habilidade, que será apresentada na próxima seção.

2.3.4 Abordagem bayesiana

Na abordagem bayesiana, o traço latente é considerado uma variável aleatória baseando-se no princípio de que toda incerteza pode ser descrita através de probabilidade. A distribuição do parâmetro nesse contexto é uma junção da informação anterior a aplicação do TAI descrita por uma distribuição de probabilidade com a informação vinda das respostas aos itens apresentados pelo TAI descrita pela função de verossimilhança. Então, o estimador do parâmetro é esta distribuição chamada de distribuição a *posteriori* e usamos como estimativa pontual alguma característica desta distribuição, como por exemplo a

média.

Estamos supondo neste trabalho que a distribuição a *priori* da habilidade é normal padrão, ou seja, $\theta \sim \pi(\theta)$ em que

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\theta^2\right).$$

Assim, a nossa distribuição a *posteriori* após k itens respondidos é dada por

$$\pi(\theta | u_{i_1}, \dots, u_{i_k}) \propto \pi(\theta) \cdot L(\theta | u_{i_1}, \dots, u_{i_k}).$$

Notemos que no caso bayesiano não encontramos o problema do estimador não ser finito, isto é, não existir. Ele sempre existe e é uma distribuição de probabilidade. Como estimador pontual usamos alguma característica dessa distribuição. Neste trabalho, optamos por utilizar a média como estimador pontual da habilidade.

No próximo capítulo apresentaremos como construir um Teste Adaptativo Informatizado. Poderemos notar que a estimação da habilidade exerce um papel central nos algoritmos de seleção.

Capítulo 3

Algoritmos de seleção

Conforme notado no Capítulo 1, um Teste Adaptativo Informatizado (TAI) é aplicado segundo um algoritmo de seleção. Esse algoritmo especifica as questões que o indivíduo irá responder e sua ordem de apresentação. Podemos descrever esse algoritmo em três partes (Wainer, 2001):

Como começar? Qual o primeiro item a ser respondido?

Como continuar? Depois de cada resposta, qual o próximo item?

Quando parar? Quando o teste termina?

Ou mais precisamente:

Como começar? Neste trabalho, estamos supondo que a população tem distribuição normal padrão, ou seja, a distribuição *a priori* para o traço latente é normal padrão. Utilizamos esta distribuição *a priori* para selecionar o primeiro item.

Como continuar? Após cada nova resposta, atualizamos a estimativa da habilidade, ou seja, a distribuição *a posteriori* e a usamos para selecionar o próximo item.

Quando parar? Em TAI, é necessário estabelecer o momento em que o indivíduo não responderá mais itens, chamado de critério de parada. Assim, optamos por usar o

seguinte critério de parada: ou o erro padrão do estimador da habilidade fica menor que δ ou N itens são respondidos. O valor de δ e N é determinado pelo construtor do TAI. Esse critério de parada é utilizado em todos os algoritmos de seleção analisados neste trabalho.

3.1 Notação

Antes de apresentarmos os algoritmos de seleção analisados, precisamos complementar as notações apresentadas na Seção 2.3.1. A notação desse trabalho foi sugerida por Van Der Linden (2010).

- Suponhamos que queremos selecionar o k -ésimo item no TAI após $k - 1$ itens respondidos. S_k representa o conjunto de todos os itens já respondidos, ou seja, $S_k = \{i_1, \dots, i_{k-1}\}$;
- Ainda supondo que desejamos escolher o item i_k , R_k representa o conjunto de todos os itens não respondidos, isto é, $R_k = \{1, \dots, I\} - S_k$;
- Denotamos a média a *posteriori* após $k - 1$ itens respondidos por $E(\theta \mid u_{i_1}, \dots, u_{i_{k-1}})$;
- Denotamos a variância a *posteriori* após $k - 1$ itens respondidos por $\text{Var}(\theta \mid u_{i_1}, \dots, u_{i_{k-1}})$;
- Quando não houver ambiguidade, representaremos as variáveis respostas dos itens já respondidos por \mathbf{U} , ou seja, $\mathbf{U} = (U_{i_1}, \dots, U_{i_{k-1}})$;
- Quando não houver ambiguidade, representaremos o valor observado de \mathbf{U} por \mathbf{u} ;
- Quando não houver ambiguidade, representaremos a distribuição a *posteriori* por $\pi(\theta \mid \mathbf{u})$, a média a *posteriori* por $E(\theta \mid \mathbf{u})$ e a variância a *posteriori* por $\text{Var}(\theta \mid \mathbf{u})$;
- A estimativa pontual da habilidade, após $k - 1$ itens respondidos, é denotada por $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}}$. Quando não houver ambiguidade, denotaremos $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}}$ simplesmente por $\hat{\theta}_{k-1}$;

- A estimativa pontual da habilidade quando não há itens respondidos é denotada por $\hat{\theta}_0$;
- Denotamos por R_0 o conjunto de todos os itens disponíveis em um TAI, ou seja, $R_0 = \{1, \dots, I\}$;
- Usamos as expressões “critério”, “critério de seleção” e “algoritmo de seleção” para designar algoritmo de seleção.

Nas seções que seguem apresentaremos os algoritmos de seleção estudados.

3.2 Critério da Máxima Informação de Fisher

O critério da máxima Informação de Fisher, proposto por Birnbaum (1968), foi tencionado pra diminuir a variância assintótica do EMV da habilidade. A variância assintótica é a recíproca da Informação de Fisher (Casella & Berger, 2001), em que Informação de Fisher é dada por (Cox, 2006)

$$I(\theta) = E \left(\frac{\partial}{\partial \theta} \log(L(\theta | \mathcal{U})) \right)^2.$$

Para o ML3, a Informação de Fisher para U_{i_1}, \dots, U_{i_k} é dada por (Andrade *et al.*, 2000)

$$I(\theta) = \sum_{j=1}^k \frac{\left(\frac{\partial}{\partial \theta} P_{i_j} \right)^2}{P_{i_j}(1 - P_{i_j})}.$$

Neste algoritmo de seleção, usamos a estimativa pontual da habilidade ¹ para selecionar itens. Quando não há itens respondidos, a distribuição a *posteriori* é normal

¹Neste trabalho, a estimativa pontual da habilidade é a média a *posteriori*.

padrão e, assim, $\hat{\theta}_0 = 0$. Logo, o primeiro item que o indivíduo responderá segundo o critério da máxima Informação de Fisher, é descrito por

$$\begin{aligned} i_1 &= \arg \max_m \{I(0) \mid m \in R_0\} \\ &= \left\{ \frac{\left(\frac{\partial}{\partial \theta} P_m(0)\right)^2}{P_0(m)(1 - P_m(0))} \mid m \in R_0 \right\}. \end{aligned}$$

Consideremos agora que já foram respondidos $k-1$ ($k-1 > 0$) itens e desejamos selecionar o k -ésimo item, então

$$\begin{aligned} i_k &= \arg \max_m \{I(\hat{\theta}_{k-1}) \mid m \in R_k\} \\ &= \arg \max_m \left\{ \sum_{j=1}^{k-1} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\hat{\theta}_{k-1})\right)^2}{P_{i_j}(\hat{\theta}_{k-1})(1 - P_{i_j}(\hat{\theta}_{k-1}))} + \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{k-1})\right)^2}{P_m(\hat{\theta}_{k-1})(1 - P_m(\hat{\theta}_{k-1}))} \mid m \in R_k \right\}. \quad (3.1) \end{aligned}$$

Notemos que (3.1) é equivalente a

$$i_k = \arg \max_m \left\{ \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{k-1})\right)^2}{P_m(\hat{\theta}_{k-1})(1 - P_m(\hat{\theta}_{k-1}))} \mid m \in R_k \right\}.$$

Segundo Van Der Linden (2010), uma das razões para a popularidade de TAI é a menor quantidade necessária de respostas, ou seja, testes menores. Contudo, isso contradiz o conceito de variância assintótica usado como motivação para o critério da máxima Informação de Fisher. Apesar disto, o critério da máxima Informação de Fisher é um dos algoritmos de seleção mais utilizados em estudos teóricos e em implementações de

TAI.

3.3 Critério de Owen

Owen (1969) propôs realizar a “adaptação” ao indivíduo usando a distribuição a *posteriori* e foi pioneiro nesse sentido (Van Der Linden, 2010). A proposta de Owen (1969) é encontrar itens através do parâmetro de dificuldade, isto é, escolher um item com valor da dificuldade que seja compatível com a estimativa pontual interina da habilidade. Assim, a escolha do primeiro é determinada por

$$i_1 = \arg \min_m \{ |b_m - \hat{\theta}_0| \mid m \in R_0 \}$$

como $\hat{\theta}_0 = 0$ (pois a distribuição a *priori* é normal padrão e usamos a média a *posteriori* como estimador pontual), temos que

$$i_1 = \arg \min_m \{ |b_m| \mid m \in R_0 \}.$$

Suponhamos que temos $k - 1$ itens respondidos e desejamos selecionar o k -ésimo item. O critério de Owen seleciona o item por

$$i_k = \arg \min_m \{ |b_m - \hat{\theta}_{k-1}| \mid m \in R_k \}.$$

Uma das vantagens desse critério é seu enunciado simples, contudo existem críticas ao mesmo por não utilizar informações adicionais como, por exemplo, o parâmetro de discriminação (Van Der Linden, 2010).

3.4 Algoritmos ponderados pela distribuição *posteriori*

Nesta seção, introduzimos critérios que usam ponderação da distribuição a *posteriori* para selecionar itens. Os dois primeiros algoritmos de seleção propõem melhorias no critério da máxima Informação de Fisher apresentado na seção 3.2. O terceiro critério de seleção combina as motivações dos critérios apresentados na seção 3.4.3 e na seção 3.4.1 em um novo critério proposto por Van Der Linden (1998). O quarto algoritmo de seleção tenta minimizar a variância a *posteriori* (Owen, 1975).

3.4.1 Critério da Máxima Informação ponderada pela distribuição a *posteriori*

A motivação para este algoritmo de seleção é que no início do teste, quando temos pouca informação sobre o traço latente, pode ser aquém do ideal selecionar o item que tem o maior valor da Informação de Fisher na estimativa pontual interina da habilidade, pois valores próximos têm chance substancial de ser o valor real da habilidade e os ignoramos (Veerkamp & Berger, 1997). A proposta é selecionar o próximo item para maximizar o valor esperado da Informação de Fisher sobre a distribuição a *posteriori* de θ (Van Der Linden, 1998). Assim, escolhemos o primeiro item por

$$\begin{aligned}
 i_1 &= \arg \max_m \left\{ \int_{\mathbb{R}} I(\theta) \pi(\theta) d\theta \mid m \in R_0 \right\} \\
 &= \arg \max_m \left\{ \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta) d\theta \mid m \in R_0 \right\}
 \end{aligned}$$

e escolhemos o k -ésimo item após $k - 1$ itens respondidos através de

$$\begin{aligned}
i_k &= \arg \max_m \left\{ \int_{\mathbb{R}} I(\theta) \pi(\theta | \mathbf{u}) d\theta \mid m \in R_k \right\} \\
&= \arg \max_m \left\{ \sum_{j=1}^{k-1} \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\theta) \right)^2}{P_{i_j}(\theta)(1 - P_{i_j}(\theta))} \pi(\theta | \mathbf{u}) d\theta + \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}) d\theta \mid m \in R_k \right\}.
\end{aligned} \tag{3.2}$$

Note que (3.2) é equivalente a

$$i_k = \left\{ \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}) d\theta \mid m \in R_k \right\}$$

3.4.2 Distribuição preditiva em TAI

Em muitas aplicações, há interesse em realizar predições em variáveis não-observadas e TAI é uma destas aplicações. Estabelecemos os conceitos e notações de distribuição preditiva no contexto de TAI e os usaremos nas seções 3.4.3, 3.4.4 e 3.4.5. A notação apresentada nesta seção foi sugerida por Van Der Linden (2010).

Suponhamos que após $k - 1$ respostas ao teste estamos interessados na previsão da variável resposta U_{i_k} . Então, a distribuição preditiva U_{i_k} dados $\mathbf{u} = (u_{i_1}, \dots, u_{i_{k-1}})$ é obtida por

$$P(U_{i_k} = u_{i_k} | \mathbf{u}) = \int_{\mathbb{R}} P_{i_k}(\theta)^{u_{i_k}} (1 - P_{i_k}(\theta))^{1-u_{i_k}} \pi(\theta | \mathbf{u}) d\theta \tag{3.3}$$

Notemos que a probabilidade de acertar o item i_k , após $k - 1$ respostas ao teste é $P(U_{i_k} = 1 | \mathbf{u})$ e que um acerto a este item nos conduz a atualizar a distribuição a

posteriori e, conseqüentemente, a média e variância a *posteriori*. Ou seja,

- a distribuição a *posteriori* é atualizada para $\pi(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$. Quando não houver ambigüidade, denotaremos $\pi(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$ por $\pi(\theta \mid \mathbf{u}, U_{i_k} = 1)$;
- a média a *posteriori* é atualizada para $E(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$. Quando não houver ambigüidade, denotaremos $E(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$ por $E(\theta \mid \mathbf{u}, U_{i_k} = 1)$;
- a estimativa pontual da habilidade é atualizada para $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k}=1}$. Quando não houver ambigüidade, denotaremos $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k}=1}$ por $\hat{\theta}_{\mathbf{u}, U_{i_k}=1}$;
- a variância a *posteriori* é atualizada para $\text{Var}(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$. Quando não houver ambigüidade, denotaremos $\text{Var}(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 1)$ por $\text{Var}(\theta \mid \mathbf{u}, U_{i_k} = 1)$;
- denotaremos $P(U_{i_k} = 1 \mid \mathbf{u})$ por $P_{i_k}(1 \mid \mathbf{u})$.

De forma análoga, a probabilidade de errar o item i_k após $k - 1$ respostas ao teste é $P(U_{i_k} = 0 \mid \mathbf{u})$ e um erro a este item nos conduz a atualizar a distribuição a *posteriori* e, conseqüentemente, a média e variância a *posteriori*. Isto é,

- a distribuição a *posteriori* é atualizada para $\pi(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$. Quando não houver ambigüidade, denotaremos $\pi(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$ por $\pi(\theta \mid \mathbf{u}, U_{i_k} = 0)$;
- a média a *posteriori* é atualizada para $E(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$. Quando não houver ambigüidade, indicaremos $E(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$ por $E(\theta \mid \mathbf{u}, U_{i_k} = 0)$;
- a estimativa pontual da habilidade é atualizada para $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k}=0}$. Quando não houver ambigüidade, denotaremos $\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k}=0}$ por $\hat{\theta}_{\mathbf{u}, U_{i_k}=0}$;
- a variância a *posteriori* é atualizada para $\text{Var}(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$. Quando não houver ambigüidade, indicaremos $\text{Var}(\theta \mid u_{i_1}, \dots, u_{i_{k-1}}, U_{i_k} = 0)$ por $\text{Var}(\theta \mid \mathbf{u}, U_{i_k} = 0)$;
- denotaremos $P(U_{i_k} = 0 \mid \mathbf{u})$ por $P_{i_k}(0 \mid \mathbf{u})$.

Quando não há itens respondidos, substituímos a distribuição a *posteriori* na equação 3.3 pela distribuição a *priori* e obtemos

$$P(U_{i_1} = u_{i_1}) = \int_{\mathbb{R}} P_{i_1}(\theta)^{u_{i_1}} (1 - P_{i_1}(\theta))^{1-u_{i_1}} \pi(\theta) d\theta.$$

Então, a probabilidade de acerto ao item i_1 é $P(U_{i_1} = 1)$ e um acerto nos conduz a atualizar a distribuição a *posteriori* e, conseqüentemente, a média e a variância a *posteriori*. Ou seja,

- a distribuição a *posteriori* é atualizada para $\pi(\theta | U_{i_1} = 1)$;
- a média a *posteriori* é atualizada para $E(\theta | U_{i_1} = 1)$;
- a estimativa pontual da habilidade é atualizada para $\hat{\theta}_{U_{i_1}=1}$;
- a variância a *posteriori* é atualizada para $\text{Var}(\theta | U_{i_1} = 1)$.

De maneira análoga, a probabilidade de errar o item i_1 é $P(U_{i_1} = 0)$ e um erro nos conduz a atualizar a distribuição a *posteriori* e, conseqüentemente, a média e a variância a *posteriori*. Isto é,

- a distribuição a *posteriori* é atualizada para $\pi(\theta | U_{i_1} = 0)$;
- a média a *posteriori* é atualizada para $E(\theta | U_{i_1} = 0)$;
- a estimativa pontual da habilidade é atualizada para $\hat{\theta}_{U_{i_1}=0}$;
- a variância a *posteriori* é atualizada para $\text{Var}(\theta | U_{i_1} = 0)$.

3.4.3 Critério da máxima informação esperada

Van Der Linden (1998) propôs usar a distribuição preditiva apresentada na seção 3.4.2 para aprimorar o critério proposto na seção 3.2. Notemos que, conforme a seção 3.4.2,

a distribuição preditiva fornece a probabilidade de acerto (ou erro) a um item de acordo com a informação disponível. Assim, Van Der Linden (1998) propôs um algoritmo de seleção que determina o primeiro item por

$$\begin{aligned}
i_1 &= \arg \max_m \left\{ P(U_m = 1) I(\hat{\theta}_{U_m=1}) + P(U_m = 0) I(\hat{\theta}_{U_m=0}) \mid m \in R_0 \right\} \\
&= \arg \max_m \left\{ P(U_m = 1) \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{U_m=1}) \right)^2}{P_m(\hat{\theta}_{U_m=1})(1 - P_m(\hat{\theta}_{U_m=1}))} + P(U_m = 0) \cdot \right. \\
&\quad \left. \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{U_m=0}) \right)^2}{P_m(\hat{\theta}_{U_m=0})(1 - P_m(\hat{\theta}_{U_m=0}))} \mid m \in R_0 \right\},
\end{aligned}$$

e seleciona o k -ésimo item após $k - 1$ itens respondidos por

$$\begin{aligned}
i_k &= \arg \max_m \left\{ P_m(1 \mid \mathbf{u}) I(\hat{\theta}_{\mathbf{u}, U_m=1}) + P_m(0 \mid \mathbf{u}) I(\hat{\theta}_{\mathbf{u}, U_m=0}) \mid m \in R_k \right\} \\
&= \arg \max_m \left\{ P_m(1 \mid \mathbf{u}) \left[\sum_{j=1}^{k-1} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=1}) \right)^2}{P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=1})(1 - P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=1}))} + \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{\mathbf{u}, U_m=1}) \right)^2}{P_m(\hat{\theta}_{\mathbf{u}, U_m=1})(1 - P_m(\hat{\theta}_{\mathbf{u}, U_m=1}))} \right] + \right. \\
&\quad \left. + P_m(0 \mid \mathbf{u}) \left[\sum_{j=1}^{k-1} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=0}) \right)^2}{P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=0})(1 - P_{i_j}(\hat{\theta}_{\mathbf{u}, U_m=0}))} + \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{\mathbf{u}, U_m=0}) \right)^2}{P_m(\hat{\theta}_{\mathbf{u}, U_m=0})(1 - P_m(\hat{\theta}_{\mathbf{u}, U_m=0}))} \right] \right. \\
&\quad \left. \mid m \in R_k \right\}, \tag{3.4}
\end{aligned}$$

Notemos que (3.4) é equivalente a

$$i_k = \arg \max_m \left\{ P_m(1 | \mathbf{u}) \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{\mathbf{u}, U_m=1}) \right)^2}{P_m(\hat{\theta}_{\mathbf{u}, U_m=1})(1 - P_m(\hat{\theta}_{\mathbf{u}, U_m=1}))} + P_m(0 | \mathbf{u}) \cdot \frac{\left(\frac{\partial}{\partial \theta} P_m(\hat{\theta}_{\mathbf{u}, U_m=0}) \right)^2}{P_m(\hat{\theta}_{\mathbf{u}, U_m=0})(1 - P_m(\hat{\theta}_{\mathbf{u}, U_m=0}))} \mid m \in R_k \right\}.$$

3.4.4 Critério da Máxima Informação Esperada ponderada pela *posteriori*

Van Der Linden (1998) propões unir os conceitos apresentados na seção 3.4.2 e a motivação subjacente ao algoritmo de seleção exposto na seção 3.4.1 em um novo critério, descrito nesta seção. Este critério seleciona o primeiro item através de

$$\begin{aligned} i_1 &= \arg \max_m \left\{ P(U_m = 1) \int_{\mathbb{R}} I(\theta) \pi(\theta | U_m = 1) d\theta + P(U_m = 0) \int_{\mathbb{R}} I(\theta) \pi(\theta | U_m = 0) d\theta \mid m \in R_0 \right\} \\ &= \arg \max_m \left\{ P(U_m = 1) \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | U_m = 1) d\theta + P(U_m = 0) \cdot \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | U_m = 0) d\theta \mid m \in R_0 \right\} \end{aligned}$$

e seleciona o k -ésimo item após $k - 1$ itens respondidos por meio de

$$\begin{aligned}
i_k &= \arg \max_m \left\{ P_m(1 | \mathbf{u}) \int_{\mathbb{R}} I(\theta) \pi(\theta | \mathbf{u}, U_m = 1) d\theta + P_m(0 | \mathbf{u}) \cdot \right. \\
&\quad \left. \int_{\mathbb{R}} I(\theta) \pi(\theta | \mathbf{u}, U_m = 0) d\theta \mid m \in R_k \right\} \\
&= \arg \max_m \left\{ P_m(1 | \mathbf{u}) \left[\sum_{j=1}^{k-1} \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\theta) \right)^2}{P_{i_j}(\theta)(1 - P_{i_j}(\theta))} \pi(\theta | \mathbf{u}, U_m = 1) d\theta + \right. \right. \\
&\quad \left. \left. + \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}, U_m = 1) d\theta \right] + P_m(0 | \mathbf{u}) \cdot \right. \\
&\quad \left. \left[\sum_{j=1}^{k-1} \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_{i_j}(\theta) \right)^2}{P_{i_j}(\theta)(1 - P_{i_j}(\theta))} \pi(\theta | \mathbf{u}, U_m = 0) d\theta + \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}, U_m = 0) d\theta \right] \right. \\
&\quad \left. \mid m \in R_k \right\} \tag{3.5}
\end{aligned}$$

Notemos que (3.5) é equivalente a

$$\begin{aligned}
i_k &= \left\{ P_m(1 | \mathbf{m}) \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}, U_m = 1) d\theta + \right. \\
&\quad \left. + P_m(0 | \mathbf{m}) \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial \theta} P_m(\theta) \right)^2}{P_m(\theta)(1 - P_m(\theta))} \pi(\theta | \mathbf{u}, U_m = 0) d\theta \mid m \in R_k \right\}
\end{aligned}$$

3.4.5 Critério da Mínima Variância a *posteriori*

Como notado na seção 3.2, o uso da Informação de Fisher é questionável em critério de seleção, principalmente no início do teste, pois a motivação para o algoritmo de seleção da seção 3.2 é que a variância do estimador converge para a recíproca da Informação de

Fisher. Assim, expomos aqui um algoritmo de seleção que usa a variância do estimador sem aproximações proposto por Owen (1975). Na década de 70 havia sérias restrições computacionais que impediam a implementação desse algoritmo, contudo, segundo Van Der Linden (2010), tais restrições não existem mais. O critério seleciona o primeiro item através de

$$i_1 = \arg \max_m \{P(U_m = 1) \text{Var}(\theta | U_m = 1) + P(U_m = 0) \text{Var}(\theta | U_m = 0) | m \in R_0\},$$

e seleciona o k -ésimo item após $k - 1$ itens respondidos por meio de

$$i_k = \arg \max_m \{P_m(1 | \mathbf{u}) \text{Var}(\theta | \mathbf{u}, U_m = 1) + P_m(0 | \mathbf{u}) \text{Var}(\theta | \mathbf{u}, U_m = 0) | m \in R_k\}.$$

3.5 Algoritmo do Critério da Máxima Informação Global

Segundo Van Der Linden (2010), o algoritmo de seleção mais utilizado é o critério da máxima Informação de Fisher. Contudo, conforme já notado na seção 3.2, este critério tem por inspiração o fato de que a variância do estimador é, assintoticamente, a recíproca da Informação de Fisher e, quando não há um número grande de itens respondidos, a Informação de Fisher na estimativa pontual da habilidade pode não refletir a informação que o item contém (Chang & Ying, 1996). Além disto, Chang & Ying (1996) questionam se uma função univariada, como é o caso da Informação de Fisher, é suficiente pra capturar toda informação contida em um item. Dessa maneira, apresentamos um novo algoritmo de seleção, baseado na Informação de Kullback-Leibler ² proposta por Chang & Ying (1996).

²Os conceitos e notações desta seção foram propostos por Chang & Ying (1996).

3.5.1 Informação de Kullback-Leibler

Apresentaremos a Informação de Kullback-Leibler primeiro para um item e depois para um número qualquer de itens.

Definição 3.5.1 (Informação de Kullback-Leibler para um item). *Seja θ_v o verdadeiro valor do habilidade/traço latente e $\theta \in \Theta$, em que Θ representa o conjunto de todos os valores possíveis, em nosso caso $\Theta = \mathbb{R}$. A Informação de Kullback-Leibler do k -ésimo item (U_{i_k}) é definido por*

$$K_{i_k}(\theta|\theta_v) = E_{\theta_v} \log \left(\frac{L(\theta_v | U_{i_k} = u_{i_k})}{L(\theta | U_{i_k} = u_{i_k})} \right),$$

em que E_{θ_v} denota a esperança sobre U_{i_k} dado que o verdadeiro valor do traço latente é θ_v e

$$L(\theta | U_{i_k} = u_{i_k}) = P_{i_k}(\theta)^{u_{i_k}} (1 - P_{i_k}(\theta))^{1-u_{i_k}},$$

é a verossimilhança para o item U_{i_k} .

Notemos que

$$K_{i_k}(\theta|\theta_v) = P_{i_k}(\theta_v) \log \left(\frac{P_{i_k}(\theta_v)}{P_{i_k}(\theta)} \right) + (1 - P_{i_k}(\theta_v)) \log \left(\frac{1 - P_{i_k}(\theta_v)}{1 - P_{i_k}(\theta)} \right).$$

Algumas propriedades para a Informação de Kullback-Leibler são válidas:

i) $K_{i_k}(\theta_v|\theta_v) = 0$

ii) $K_{i_k}(\theta|\theta_v) \geq 0$.

A verificação da validade da propriedade (i) é imediata, basta notarmos que $\log\left(\frac{P_{i_k}(\theta_v)}{P_{i_k}(\theta_v)}\right) = \log\left(\frac{1 - P_{i_k}(\theta_v)}{1 - P_{i_k}(\theta_v)}\right) = 0$.

Para demonstrar o item (ii), vamos utilizar dois lemas.

Lema 3.5.2. *A seguinte desigualdade é verdadeira*

$$\log(x) \leq x - 1, \quad x > 0. \quad (3.6)$$

Demonstração. Primeiro consideremos a função $f(x) = \log(x) + 1 - x$, $x > 0$. As seguintes afirmações são verdadeiras

- i) $f'(x) = \frac{1}{x} - 1$;
- ii) $f'(x) > 0$, para $0 < x < 1$;
- iii) $f'(x) \leq 0$, para $x \geq 1$;
- iv) $f(1) = 0$.

Concluimos então que $f(x) \leq 0$ para $x > 0$ e, conseqüentemente, $\log(x) \leq x - 1$, $x > 0$.

□

Lema 3.5.3. *Seja q_1, \dots, q_n e p_1, \dots, p_n , com $p_i, q_i > 0$ para $\forall i \in \{1, \dots, n\}$, $\sum_{i=1}^n p_i = 1$ e $\sum_{i=1}^n q_i = 1$. Então, $\sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right) \geq 0$.*

Demonstração. Pelo lema 3.5.2, temos que

$$\begin{aligned}
\sum_{i=1}^n p_i \log \left(\frac{q_i}{p_i} \right) &\leq \sum_{i=1}^n p_i \left(\frac{q_i}{p_i} - 1 \right) \\
&= \sum_{i=1}^n (q_i - p_i) \\
&= 0,
\end{aligned}$$

ou seja, $\sum_{i=1}^n p_i \log \left(\frac{q_i}{p_i} \right) \leq 0$. Se multiplicarmos esta última desigualdade por -1 , obtemos $\sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right) \geq 0$.

□

Agora, considere $n=2$, $p_1 = P_{i_k}(\theta_v)$, $p_2 = 1 - P_{i_k}(\theta_v)$, $q_1 = P_{i_k}(\theta)$ e $q_2 = 1 - P_{i_k}(\theta)$.

Pelo lema 3.5.3, concluímos que $K_{i_k}(\theta || \theta_v) \geq 0$, como queríamos.

Temos definido a Informação de Kullback-Leibler para apenas um item, vamos agora generalizar para n itens.

Definição 3.5.4 (Informação de Kullback-Leibler). *Seja θ_v o verdadeiro valor do traço latente. Para algum θ , a Informação de Kullback-Leibler para um indivíduo que respondeu k itens é definida como*

$$K(\theta || \theta_v) = E_{\theta_v} \log \left(\frac{L(\theta_v | U_{i_1}, \dots, U_{i_k})}{L(\theta | U_{i_1}, \dots, U_{i_k})} \right)$$

em que U_{i_1}, \dots, U_{i_k} são variáveis respostas (vide Seção 3.1 para definição de variável resposta).

Uma propriedade importante da Informação de Kullback-Leibler é a aditividade.

Proposição 3.5.5 (Aditividade da Informação de Kullback-Leibler). *Seja θ_v o verdadeiro valor do traço latente e $\theta \in \Theta$, em que Θ é o conjunto de todos os possíveis valores da*

habilidade e , conforme a definição 3.5.1, $\Theta = \mathbb{R}$. Então vale a seguinte igualdade para a Informação de Kullback-Leibler

$$\begin{aligned} K(\theta || \theta_v) &= E_{\theta_v} \log \left(\frac{L(\theta_v | U_1, \dots, U_n)}{L(\theta | U_1, \dots, U_n)} \right) \\ &= \sum_{i=1}^n E_{\theta_v} \log \left(\frac{L(\theta_v | U_i)}{L(\theta | U_i)} \right). \end{aligned}$$

A demonstração da propriedade 3.5.5 é uma consequência imediata das propriedades de logaritmo e das propriedades de esperança.

Análogo à Informação de Fisher, a proposição 3.5.5 diz que a contribuição de cada item à informação é aditiva. Portanto, a quantidade total de informação para um conjunto de itens respondidos pelo examinado pode ser facilmente determinada. Segundo Van Der Linden (2010), esta característica é altamente recomendável em TAI's, porque possibilita calcular essas informações separadamente para cada item. Outra característica da Informação de Kullback-Leibler é o fato dela ser bivariada, isto é, a Informação de Kullback-Leibler representa o poder de discriminação do item em dois níveis (o verdadeiro valor do traço latente e outro valor que é de interesse). Além disso, não requer que um número suficientemente grande de itens respondidos para que possamos usá-la de modo seguro, como é o caso da Informação de Fisher (Chang & Ying, 1996). Devido ao fato de não precisarmos ter respondido um número grande de itens, Chang & Ying (1996) a chamaram de “Informação Global”.

Nesse contexto, apresentamos a seguir o Critério da Máxima Informação Global.

3.5.2 Critério da Máxima Informação Global

Agora, vamos definir um critério de seleção que utiliza a Informação de Kullback-Leibler, que indicaremos de forma abreviada como KL.

Notemos que não conhecemos o valor do traço latente antes do TAI ser aplicado.

Mas, neste trabalho, supomos que a habilidade tem distribuição a *priori* $\pi(\theta)$ ³. Assim, a probabilidade do valor real da habilidade θ_v ser maior que $\gamma > 0$ ou menor que $-\gamma$, por exemplo com $\gamma = 10$, é pequena. Dessa forma, podemos assumir que $\theta_v \in [-\gamma, \gamma]$ e $\theta_v \sim U(-\gamma, \gamma)$. Então, em vez de usarmos KL para escolher um item, usamos sua esperança sobre θ_v . Chang & Ying (1996) propuseram usar um intervalo na estimativa pontual interina da habilidade e diminuir este intervalo quando mais itens forem respondidos. Assim, o critério seleciona o primeiro item por

$$\begin{aligned} i_1 &= \arg \max_m \left\{ \mathbb{E}_{\theta_v} \left[K_{\theta_v}(\hat{\theta}_0 | \theta_v) \right] \mid m \in R_0 \right\} \\ &= \arg \max_m \left\{ \int_{\hat{\theta}_0 - \delta_k}^{\hat{\theta}_0 + \delta_1} K(\hat{\theta}_0 | \theta_v) d\theta_v \mid m \in R_0 \right\} \end{aligned}$$

como $\hat{\theta}_0 = 0$, pois é a média da distribuição a *priori*, temos que

$$i_1 = \arg \max_m \left\{ \int_{-\delta_1}^{\delta_1} K(\hat{\theta}_0 | \theta_v) d\theta_v \mid m \in R_0 \right\}.$$

E escolhemos o k -ésimo item após $k - 1$ itens respondidos por

$$\begin{aligned} i_k &= \arg \max_m \left\{ \mathbb{E}_{\theta_v} \left[K(\hat{\theta}_{k-1} | \theta_v) \right] \mid m \in R_k \right\} \\ &= \arg \max_m \left\{ \int_{\hat{\theta}_{k-1} - \delta_k}^{\hat{\theta}_{k-1} + \delta_k} K(\hat{\theta}_{k-1} | \theta_v) d\theta_v \mid m \in R_k \right\} \end{aligned}$$

em que δ_k é uma seqüência não-crescente.

No próximo capítulo, apresentaremos os estudos de simulação comparando os algoritmos de seleção já expostos.

³Neste trabalho, a distribuição a *priori* é normal padrão.

Capítulo 4

Resultados

Até o momento, descrevemos as abordagens mais importantes para construir um Teste Adaptativo Informatizado. Contudo, não sabemos qual tem melhor performance. Apresentamos aqui quatro medidas para comparar os algoritmos de seleção propostos no Capítulo 3 e realizamos simulações para calcular essas medidas.

4.1 Medidas para comparar algoritmos de seleção

Para analisar os algoritmos de seleção propostos no capítulo 3, poderíamos usar duas estatísticas: o Erro Quadrático Médio

$$EQM = E \left(\theta - \hat{\theta} \right)^2$$

e o vício

$$b(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Contudo, notemos que é difícil estabelecer uma expressão matemática pra $\hat{\theta}$, a estimativa do traço latente obtida através do TAI, pois o número de itens respondidos e os itens escolhidos não são fixos. Assim, utilizamos duas aproximações numéricas para essas medidas:

i) Para o EQM, utilizamos a seguinte medida

$$EQM_P = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta_i)^2 \quad (4.1)$$

ii) Para o vício, usamos a seguinte medida

$$b(\theta)_P = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta_i). \quad (4.2)$$

Notemos que para calcular as medidas em (4.1) e em (4.2), precisamos de uma amostra de M valores da habilidade e precisamos também aplicar o TAI para cada um dos valores da amostra. Na seção 4.2, apresentamos como obtivemos essa amostra e como aplicamos o TAI para um dos valores da amostra.

Além do Erro Quadrático Médio e do Vício pequenos, no contexto de TAI é desejável que o tempo entre a resposta do aluno e a apresentação de um novo item seja curto, ou seja, é desejável que o critério de seleção seja rápido em decidir qual o próximo item. Chamaremos esse tempo de decisão do algoritmo de seleção simplesmente de tempo de seleção ¹ e notemos que essa medida é calculada em segundos aqui e depende da máquina onde foi calculada. Além disso, um dos objetivos do TAI é diminuir o número de itens que o examinando precisa responder, logo usaremos também o tamanho médio de um teste para avaliar os critérios de seleção. Por uma questão de simplicidade denominaremos o tamanho médio de um teste por $E[N]$.

¹O tempo de seleção foi medido em segundos.

4.2 Estudos de simulação

Explicaremos agora como obtivemos as quatro medidas descritas na seção 4.1 dadas por EQM_P , o $b(\hat{\theta})_P$, o $E(N)$ e o tempo de seleção.

Neste trabalho, estamos supondo que o traço latente tem distribuição normal padrão. Assim, obtivemos uma amostra com tamanho $M = 1000$ da distribuição normal com média zero e variância um. Além disso, calculamos o EQM_P , o $b(\hat{\theta})_P$, o $E(N)$ e o tempo de seleção para quatro critérios de parada:

- i) ou o erro padrão do estimador da habilidade fica menor que 0,01 ou o examinado respondeu 10 itens, o que ocorrer primeiro;
- ii) ou o erro padrão do estimador da habilidade fica menor que 0,01 ou o examinado respondeu 20 itens, o que ocorrer primeiro;
- iii) ou o erro padrão do estimador da habilidade fica menor que 0,01 ou o examinado respondeu 30 itens, o que ocorrer primeiro;
- iv) ou o erro padrão do estimador da habilidade fica menor que 0,01 ou o examinado respondeu 40 itens, o que ocorrer primeiro.

Notemos que temos $M = 1000$ valores θ_j da amostra do traço latente, mas não temos indivíduos com habilidade θ_j pra responder o TAI. Então, a princípio, não há possibilidade de obter as respostas para examinandos com habilidade θ_j para obter $\hat{\theta}_j$ através do TAI. Contudo, conhecendo o valor da habilidade θ_j , a variável resposta U_{i_k} tem distribuição bernoulli com probabilidade de acerto $P_{i_k}(\theta_j)$. Deste modo, podemos mimetizar a resposta de um examinando com habilidade θ_j gerando um valor aleatório de $U_{i_k} \sim \text{Bernoulli}(P_{i_k}(\theta_j))$ e consideramos esse valor como a resposta do indivíduo com habilidade θ_j .

Destacamos também que utilizamos um “banco aleatório” com 50, 300 e 700 itens fictícios, ou seja, criamos um banco com 50, 300 e 700 itens. Os valores dos parâmetros

de discriminação, dificuldade e o parâmetro de acerto casual deste banco foram obtidos através de:

- i) os valores do parâmetro de discriminação foram gerados da distribuição log-normal com média zero e variância um;
- ii) os valores do parâmetro de acerto casual foram gerados da distribuição beta com $\alpha = 1$ e $\beta = 8^2$;
- iii) os valores do parâmetro de dificuldade foram gerados da distribuição normal padrão.

Utilizamos a linguagem C e a biblioteca GSL (Galassi *et al.*, 2007) para implementar as simulações. Usamos o Cluster Intel Linux Puma ³ do LCCA - USP para executá-las.

Na seção 4.4 apresentamos um método de reportar ao examinado o seu desempenho no teste e na seção 4.3 apresentamos os resultados obtidos da simulação descrito nesta seção.

4.3 Resultados da simulação

Expomos agora os resultados obtidos da simulação descrita na seção 4.2 e apresentamos os valores obtidos do EQM_P , do $b(\theta)_P$, do $E(N)$ e do tempo de seleção. Por simplicidade, usamos as seguintes abreviações:

- o critério de parada i) apresentado na seção 4.2 é denotado por **10**;
- o critério de parada ii) apresentado na seção 4.2 é denotado por **20**;
- o critério de parada iii) apresentado na seção 4.2 é denotado por **30**;
- o critério de parada iv) apresentado na seção 4.2 é denotado por **40**;
- o Critério de Owen é chamado por CO;

²A distribuição beta com parâmetros α e β tem densidade $f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

³O Cluster PUMA é constituído por 59 servidores DELL PowerEdge 1950, com 2 Xeon 5430 (8 cores de 2,66 GHz), 16 GB de RAM DDR2-FBDIMM 667 MHz, e com 300 GB de disco SAS cada.

- o Critério da Máxima Informação de Fisher é indicado por CMIF;
- o Critério da Máxima Informação Ponderada pela Distribuição a *Posteriori* é indicado por CMIP;
- o Critério da Máxima Informação Esperada é denotado por CMIE;
- o Critério da Máxima Informação Esperada Ponderada pela *Posteriori* é chamado por CMIEPP;
- o Critério da Mínima Variância a *Posteriori* é denotado por CMVP;
- o Critério da Máxima Informação Global é indicado por CMIG.

Para cada algoritmo de seleção apresentado no Capítulo 3 e para cada critério de parada da seção 4.2 apresentamos os resultados de simulação nas tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 e nas figuras 4.1, 4.2 e 4.3. Nas tabelas 4.1, 4.2 e 4.3 mostramos os valores obtidos para o EQM_P , $b(\theta)_P$ e $E(N)$ para os bancos simulados com 50, 300 e 700 itens. Nas tabelas 4.4, 4.5 e 4.6 apresentamos o mínimo, o máximo, a média, o primeiro quartil, o terceiro quartil e o intervalo interquartil para o tempo de seleção. As figuras 4.1, 4.2 e 4.3 contêm *boxplots* para o erro padrão do estimador para os critérios estudados e para os bancos com 50, 300 e 700 itens que são mostradas com intuito de estudar o comportamento dos critérios de parada.

TABELA 4.1: EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 50 itens.

Número máximo de itens								
Estatística	10	20	30	40	10	20	30	40
	CO				CMIG			
EQM	0,5172	0,4457	0,4201	0,4169	0,5661	0,4319	0,4115	0,3683
Vicio	0,0201	0,0074	-0,0011	0,0253	0,0202	0,0310	-0,0075	0,0135
E	10,0000	20,0000	30,0000	40,0000	10,0000	20,0000	30,0000	40,0000
	CMIE				CMIP			
EQM	0,4064	0,3339	0,3009	0,2912	0,5725	0,4296	0,3678	0,3485
Vicio	-0,0117	-0,0012	-0,0097	-0,0043	-0,0015	-0,0241	-0,0110	-0,0168
E	10,0000	20,0000	30,0000	40,0000	10,0000	20,0000	30,0000	40,0000
	CMVP				CMIF			
EQM	0,4380	0,3972	0,3904	0,3545	0,4235	0,3692	0,3516	0,3365
Vicio	-0,0146	-0,0343	-0,0054	0,0107	0,0302	0,0203	0,0137	0,0054
E	10,0000	20,0000	29,9920	39,9290	10,0000	20,0000	30,0000	40,0000
	CMIEPP							
EQM	0,5725	0,4296	0,3678	0,3485				
Vicio	-0,0015	-0,0241	-0,0110	-0,0168				
E	10,0000	20,0000	30,0000	40,0000				

TABELA 4.2: EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 300 itens.

Número máximo de itens								
Estadística	10	20	30	40	10	20	30	40
	CO				CMIG			
EQM	0,5586	0,4057	0,3294	0,2841	0,5819	0,4480	0,3872	0,3514
Vicio	0,0322	0,0008	0,0137	-0,0007	0,0541	0,0078	0,0105	-0,0072
E	10,0000	20,0000	30,0000	40,0000	10,0000	20,0000	30,0000	40,0000
	CMIF							
EQM	0,3156	0,2388	0,2234	0,2027				
Vicio	0,0096	0,0145	0,0093	-0,0099				
E	10,0000	20,0000	30,0000	40,0000				

TABELA 4.3: EQM_P , $b(\theta)_P$ e $E(N)$ para o banco de 700 itens.

Número máximo de itens								
Estadística	10	20	30	40	10	20	30	40
	CO				CMIG			
EQM	0,5495	0,3940	0,3544	0,2990	0,5107	0,4471	0,3723	0,3498
Vicio	0,0262	0,0135	0,0089	-0,0037	0,0187	0,0172	0,0164	-0,0094
E	10,0000	20,0000	30,0000	40,0000	10,0000	20,0000	30,0000	40,0000
	CMIF							
EQM	0,2998	0,2115	0,1942	0,1765				
Vicio	-0,0010	0,0021	0,0009	-0,0066				
E	10,0000	20,0000	30,0000	40,0000				

TABELA 4.4: Resultados da simulação para o tempo de seleção em segundos para um banco de 50 itens

Estatística	Número máximo de itens							
	10	20	30	40	10	20	30	40
	CO				CMIG			
MIN	0,6	1,2	2,2	3,0	0,7	1,8	3,2	4,5
MAX	1,9	7,0	13,6	18,0	2,2	7,1	16,0	23,9
MEDIA	1,0	3,1	5,8	8,3	1,2	3,7	7,4	12,3
Q1	0,8	2,3	5,0	7,2	1,0	3,0	5,9	9,6
Q3	1,2	3,5	6,6	9,4	1,5	4,3	8,4	14,0
Q3-Q1	0,3	1,2	1,6	2,2	0,4	1,3	2,6	4,4
	CMIE				CMIP			
MIN	15,0	53,7	101,5	167,5	8,5	32,1	69,8	113,8
MAX	26,4	123,9	296,9	539,4	12,0	52,1	111,8	188,2
MEDIA	19,5	81,5	168,1	278,8	10,4	41,8	87,5	144,3
Q1	17,3	63,7	128,1	213,7	9,7	37,6	78,4	129,8
Q3	20,6	98,5	208,4	335,9	11,1	45,6	95,6	156,9
Q3-Q1	3,3	34,8	80,3	122,1	1,4	8,1	17,2	27,0
	CMVP				CMIF			
MIN	29,9	87,9	162,1	219,9	0,6	1,4	2,1	3,8
MAX	48,5	311,9	538,2	819,3	1,9	8,4	16,8	29,6
MEDIA	38,7	153,2	317,4	513,5	1,0	3,5	6,9	12,4
Q1	35,0	122,6	246,7	384,6	0,8	2,7	5,7	10,3
Q3	41,6	177,4	394,2	643,5	1,2	3,9	8,1	14,9
Q3-Q1	6,5	54,9	147,6	258,8	0,4	1,2	2,4	4,6
	CMIEPP							
MIN	16,7	57,6	129,1	209,5				
MAX	24,9	105,8	259,2	385,7				
MEDIA	19,3	80,3	171,3	280,6				
Q1	16,9	70,3	153,6	252,6				
Q3	23,6	90,3	188,2	306,1				
Q3-Q1	6,7	19,9	34,6	53,5				

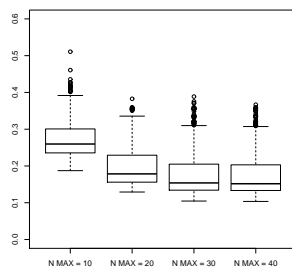
TABELA 4.5: Resultados da simulação para o tempo de seleção em segundos para um banco de 300 *itens*

Estatística	Número máximo de itens							
	10	20	30	40	10	20	30	40
	CO				CMIG			
MIN	0,6	1,6	2,5	3,7	1,5	3,6	5,8	8,6
MAX	2,8	9,0	17,5	30,8	4,2	12,5	24,8	38,1
MEDIA	1,2	4,4	8,7	16,6	2,5	7,7	12,4	18,6
Q1	0,9	3,3	6,2	12,2	2,0	6,3	10,2	14,8
Q3	1,5	5,0	10,4	20,6	2,9	8,7	14,0	21,7
Q3-Q1	0,5	1,7	4,2	8,4	0,9	2,4	3,9	6,9
	CMIF							
MIN	0,7	1,8	2,8	5,8				
MAX	3,8	15,3	30,8	43,6				
MEDIA	1,8	6,4	12,9	22,7				
Q1	1,3	4,5	9,9	17,5				
Q3	2,2	8,3	16,4	28,7				
Q3-Q1	0,9	3,8	6,6	11,2				

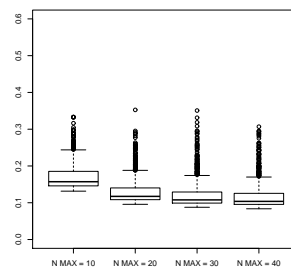
TABELA 4.6: Resultados da simulação para o tempo de seleção em segundos para um banco de 700 itens

Estatística	Número máximo de itens							
	10	20	30	40	10	20	30	40
	CO				CMIG			
MIN	0,6	1,5	2,4	3,3	3,1	6,3	8,9	11,9
MAX	2,0	6,9	17,4	29,0	5,2	12,0	27,9	47,7
MEDIA	1,0	3,1	6,7	12,3	3,6	8,2	14,4	24,7
Q1	0,9	2,4	4,9	8,4	3,3	7,5	12,1	20,5
Q3	1,2	3,4	8,1	15,9	3,9	8,7	16,3	28,3
Q3-Q1	0,3	1,0	3,2	7,5	0,5	1,3	4,3	7,8
	CMIF							
MIN	1,0	2,2	4,2	8,4				
MAX	3,1	10,6	23,3	45,8				
MEDIA	1,7	5,2	11,8	21,6				
Q1	1,3	4,1	9,0	16,1				
Q3	2,1	6,5	14,1	26,5				
Q3-Q1	0,8	2,4	5,1	10,4				

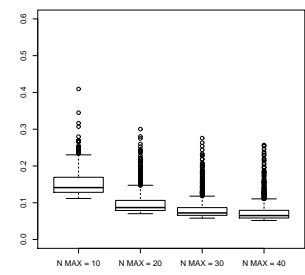
FIGURA 4.1: Erro padrão do estimador para um banco com 50 itens no final do TAI



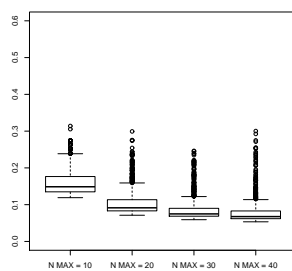
(a) CO



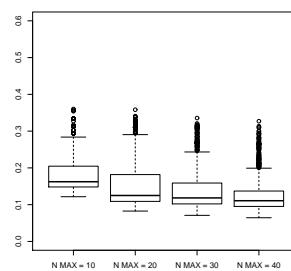
(b) CMIF



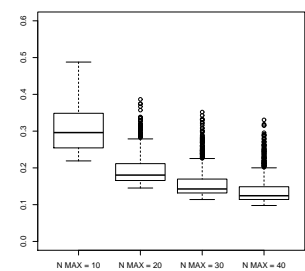
(c) CMIP



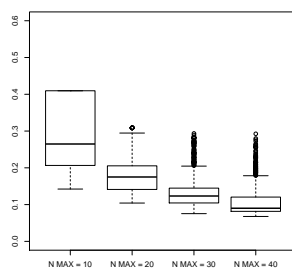
(d) CMIE



(e) CMVP

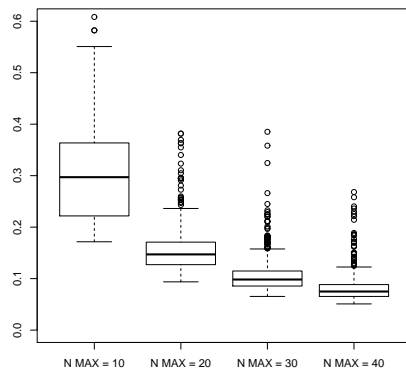


(f) CMIG

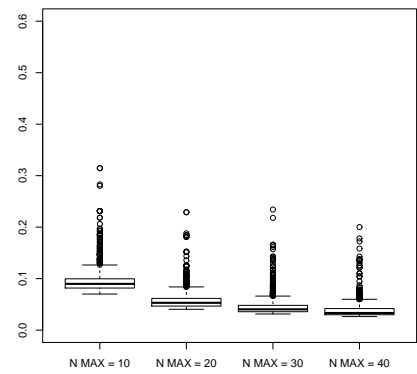


(g) CMIEPP

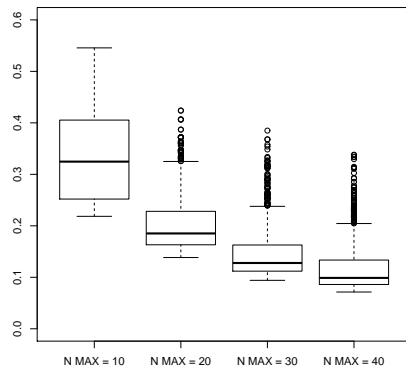
FIGURA 4.2: Erro padrão do estimador para um banco com 300 itens no final do TAI



(a) CO

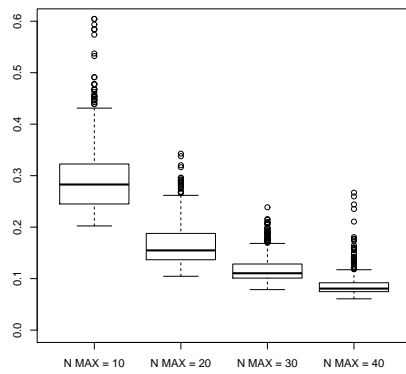


(b) CMIF

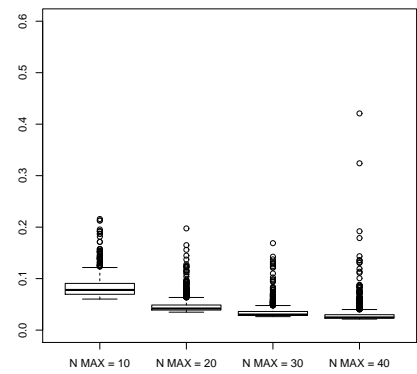


(c) CMIG

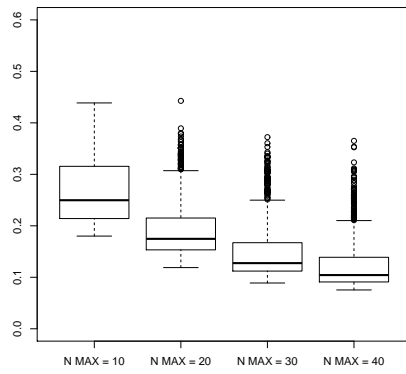
FIGURA 4.3: Erro padrão do estimador para um banco com 700 itens no final do TAI



(a) CO



(b) CMIF



(c) CMIG

Observando as figuras 4.1, 4.2 e 4.3 e as tabelas 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6, podemos notar que:

- i) os algoritmos de seleção CMVP, CMIF e CMIE têm os menores valores de EQM;
- ii) os critérios de seleção CMIE, CMIP, CMVPM, CMIEPP e CO têm os menores valores de vício;
- iii) os critérios mais rápidos em selecionar item são CO, CMIF e CMIG;
- iv) somente para o critério CMVP houve casos do TAI parar porque o erro padrão do estimador ficou menor que 0,01, contudo podemos notar pelas figuras 4.1, 4.2 e 4.3 que quanto mais itens são respondidos, menor é o erro padrão do estimador da habilidade em média;
- v) entre os algoritmos de seleção CO, CMIF e CMIG, o critério de seleção com o menor EQM e com o menor vício é o CMIF;
- vi) para os algoritmos de seleção CMIP, CMIE, CMVP e CMIEPP não foram realizadas simulações para os bancos com 300 e 700 itens, pois o tempo médio para selecionar um item foi grande ⁴;
- vii) apesar dos critérios CMVP, CMIEPP, CMIE e CMIP demorarem para selecionar um item, eles apresentam EQM e vício menores. Considerando isso, uma abordagem possível é utilizar os critérios CMVP, CMIEPP, CMIE e CMIP no começo do teste, quando há pouco informação sobre a habilidade e depois passar a utilizar os critério CO, CMIG e CMIF.

Neste trabalho, escolhemos utilizar um único algoritmo durante toda seção do TAI e escolhemos os critério CO, CMIG e CMIF em nossa implementação. Na Seção 4.4 apresentamos a abordagem utilizada para reportar o desempenho ao aluno e na Seção 4.5 explicamos como foi construída o TAI resultante deste trabalho.

⁴Para os critério CMIP, CMIE, CMVP e CMIEPP demorou mais de dois minutos para selecionar um item no início do TAI.

4.4 Apresentação de resultado

Explicar as sutilezas sobre o valor da estimativa da habilidade de uma maneira direta, concisa e sem entrar nos detalhes da TRI pode ser uma tarefa difícil. Nesse contexto, Stocking (1996) propôs apresentar ao examinando o seu desempenho através de uma transformação da estimativa do traço latente. Notemos em primeiro lugar que testes tradicionais apresentam resultado como o número de acertos. Agora suponha que a estimativa definitiva da habilidade seja $\hat{\theta}$. Stocking (1996) sugere que utilizemos o seguinte valor para reportar o desempenho ao examinando

$$\begin{aligned} E &= \sum_{i=1}^I E_{\hat{\theta}}(U_i) \\ &= \sum_{i=1}^I P_i(\hat{\theta}), \end{aligned}$$

ou seja, o valor esperado de respostas corretas de um teste composto por todos os itens do TAI supondo que $\hat{\theta}$ é o valor verdadeiro do traço latente do examinando.

4.5 Implementação

Usando a abordagem de Stocking (1996) para reportar o resultado, implementamos os seguintes algoritmos de seleção: Critério da Máxima Informação de Fisher, Critério de Owen e o Critério da Máxima Informação Global. Utilizamos uma interface entre o Excel/VBA e o *software* R (RDevelopment, 2008) construída por Baier & Neuwirth (2007). Em nossa abordagem, o Excel é usado para o usuário fornecer dados ao R, que realiza todos os cálculos necessários em TAI. Nomeamos o aplicativo de TAI2U e algumas de suas características merecem ser citadas:

- i) O usuário pode fornecer o próprio banco de itens. É necessário fornecer os parâmetros de dificuldade, o parâmetro de discriminação, o parâmetro de acerto casual e o enunciado

do item;

- ii) O usuário pode escolher o algoritmo de seleção. Há três alternativas: Critério da Máxima Informação de Fisher, Critério de Owen e Critério da Máxima Informação Global;
- iii) O usuário pode fornecer o critério de parada, definindo o número máximo que um indivíduo pode responder e o valor mínimo para o erro padrão do estimador;
- iv) Por padrão, o algoritmo de seleção escolhido é o Critério da Máxima Informação de Fisher;
- v) Por padrão, o critério de parada é definido por:
 - o número máximo de itens que o indivíduo pode responder é 2;
 - o algoritmos de seleção não selecionará mais itens quando o erro padrão ficar menor que 0.0001;

Notamos que melhorias podem ser realizadas no TAI2U. A primeira delas é utilizar modelos multimencionais no lugar de modelos unidimensionais. Além disso, verificamos utilizar o Excel/VBA em parceria com *software* R não é uma abordagem ótima, pois verificamos que o aplicativo não é rápido como gostaríamos. Dessa maneira, uma melhoria seria usar uma linguagem compilada, como a linguagem C, para implementar TAI. Outras melhorias incluem: controlar de taxa de exposição dos itens para termos um uso mais uniforme do banco de itens conforme citado no Capítulo 1 e utilizar dois ou mais critérios em uma seção de TAI conforme notado na Seção 4.3. Apesar da velocidade de nosso aplicativo estar aquém do ideal, escolhemos usar o Excel neste trabalho por ser um *software* de uso bastante difundido e por apresentar uma interface *user-friendly*. No Capítulo 5 mostraremos como instalar o TAI2U.

Capítulo 5

Instalando o TAI2U

Neste capítulo, expomos como instalar e utilizar o *software* de TAI desenvolvido neste projeto, que denominaremos simplesmente por TAI2U. A implementação do TAI2U foi realizada através do *software* VBA-Excel com uma interface com o *software* R (RDevelopment, 2008) proposta por Baier & Neuwirth (2007), sendo o MS-Excel a interface com o usuário e o *software* R responsável por realizar as computações. Em nosso aplicativo, optamos por implementar os seguintes algoritmos de seleção: critério da máxima informação de Fisher (vide seção 3.2), critério da máxima informação Global (vide seção 3.5) e critério de Owen (vide seção 3.3). O aplicativo permite escolher entre os algoritmos de seleção disponíveis, permite estabelecer o critério de parada e apresenta o resultado imediatamente usando o método de escoragem proposto por Stocking (1996) e explanado na seção 4.4. Organizamos o capítulo da seguinte forma:

- na seção 5.1 ensinamos como instalar o aplicativo;
- na seção 5.2 ensinamos como inserir um banco de itens no TAI, escolher o algoritmo de seleção e estabelecer o critério de parada, ou seja, ensinamos como configurar o aplicativo TAI2U desenvolvido neste trabalho;
- na seção 5.3 explicamos como utilizar o aplicativo.

5.1 Como Instalar

Nesta seção, vamos conduzi-lo no processo de instalação do aplicativo TAI2U. Antes de iniciar a instalação de TAI2U, é necessário ter o *software* R instalado na versão 1.12.0 (ou numa versão mais atual). Notemos que o aplicativo foi desenvolvido com versão 2003 do MS-Excel e não há garantias de funcionamentos em outras versões. Notemos também que o aplicativo foi desenvolvido para o sistema operacional *Windows* e supomos que o leitor é usuário do *software* R e do MS-Excel. A instalação do aplicativo é realizada em duas fases apresentadas nesta seção. Por simplicidade, denominaremos por R o *software* R e por Excel o MS-Excel.

5.1.1 Primeira fase

Abra o R e siga os passos abaixo:

Passo 1) instale o pacote “rcom”;

Passo 2) instale o pacote “rscproxy”;

Passo 3) digite “installconnDCOM()” no R. Este comando irá abrir uma janela de diálogo.

Siga as seguintes instruções:

1. Clique no lugar indicado pela seta preta, conforme figura a seguir

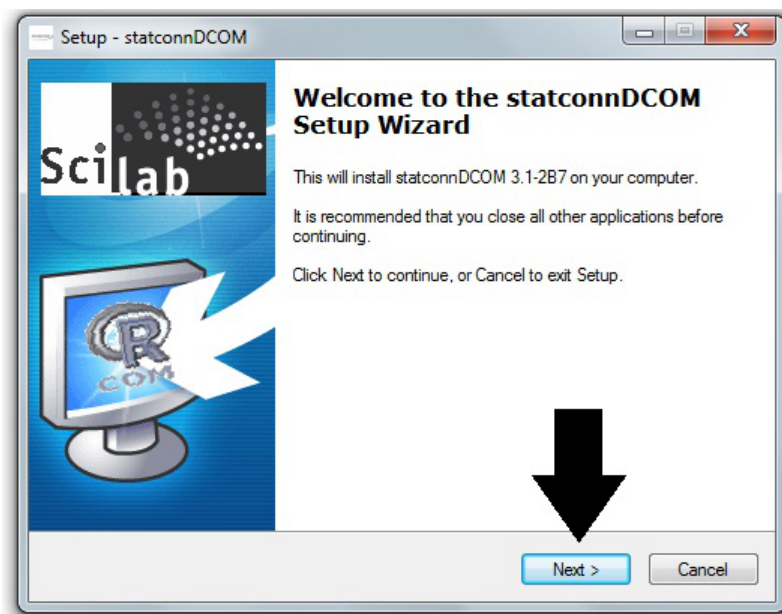


FIGURA 5.1: Passo 3) - Instrução número 1.

2. Aceite os termos de utilização, clicando no lugar indicado pela seta

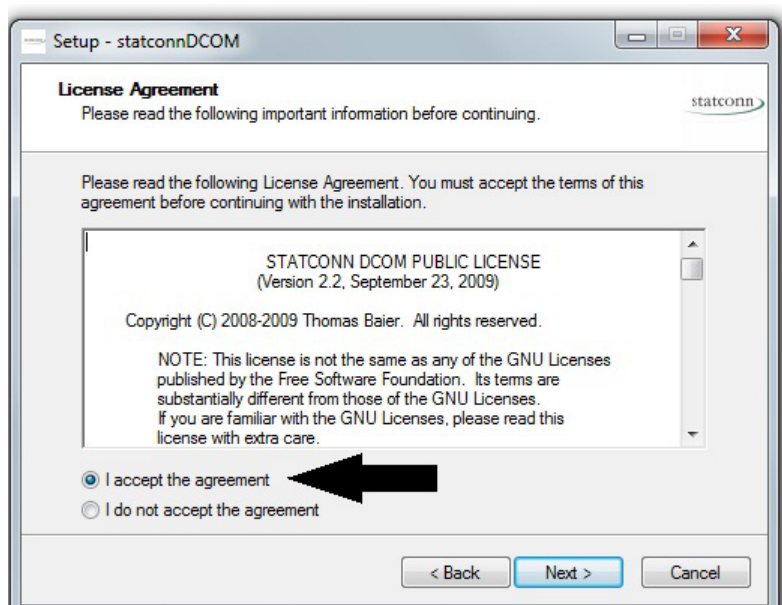


FIGURA 5.2: Passo 3) - Instrução número 2.

3. Esta janela pede para o usuário aceitar os termos de uso. Clique em *next*, conforme indica a seta preta

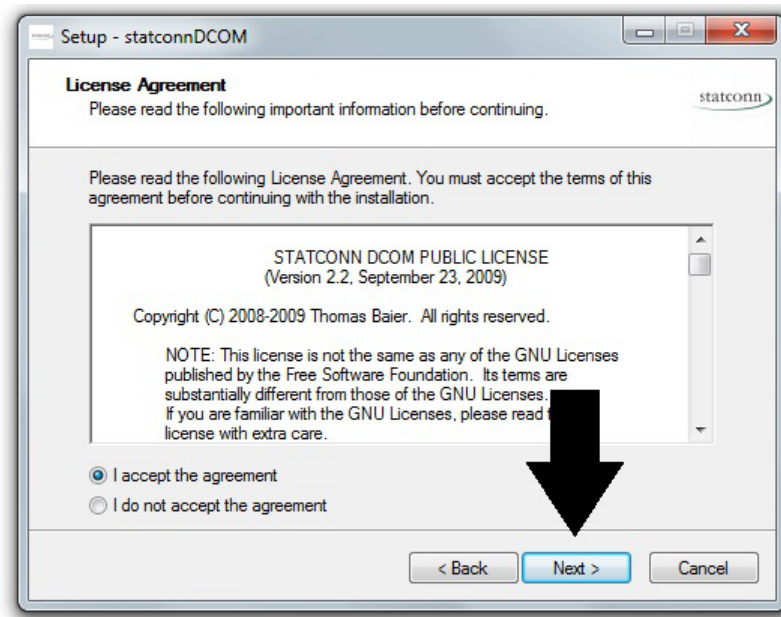


FIGURA 5.3: Passo 3) - Instrução número 3.

4. Esta janela alerta o usuário que é necessário ter os pacotes “rcom” e “rscproxy” instalados. No **Passo 1)** e no **Passo 2)** instalamos estes dois pacotes. Assim, clique em *next*, conforme indicado pela seta preta

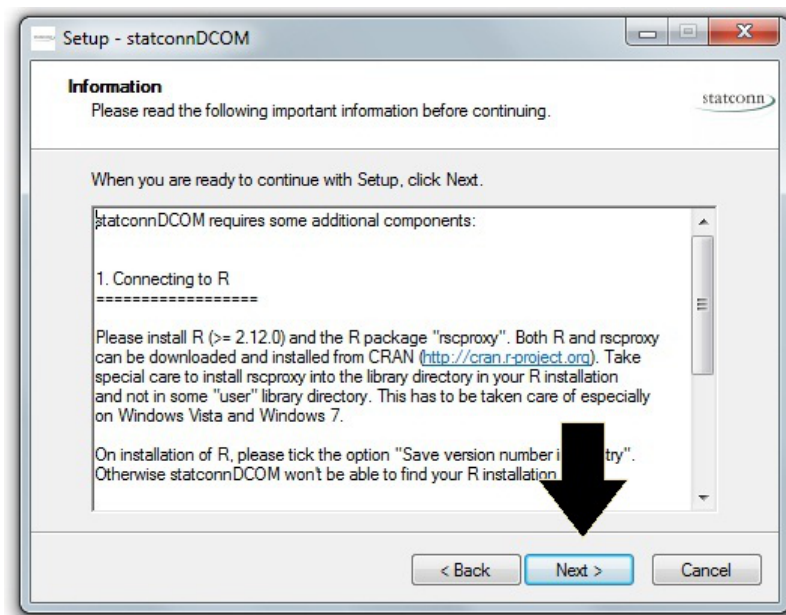


FIGURA 5.4: Passo 3) - Instrução número 4.

5. Esta janela pede para o usuário fornecer o diretório que deseja instalar o aplicativo DCOM usado na interface proposta por Baier & Neuwirth (2007). É recomendável usar o diretório padrão fornecido. Assim clique em *next*, conforme indicado pela seta preta

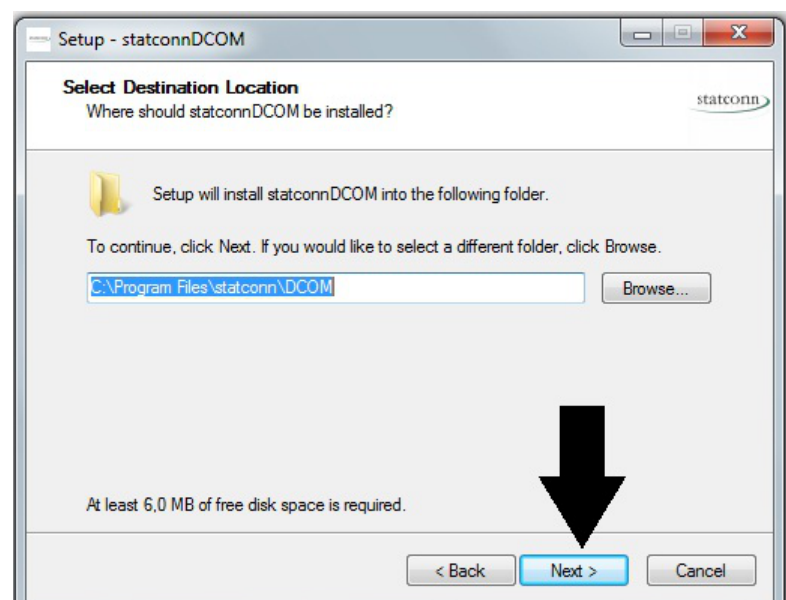


FIGURA 5.5: Passo 3) - Instrução número 5.

6. Esta janela apresenta os componentes disponíveis no instalador, escolha os mesmos itens escolhidos na figura abaixo e depois clique em *next* no lugar

indicado pela indicado pela seta preta

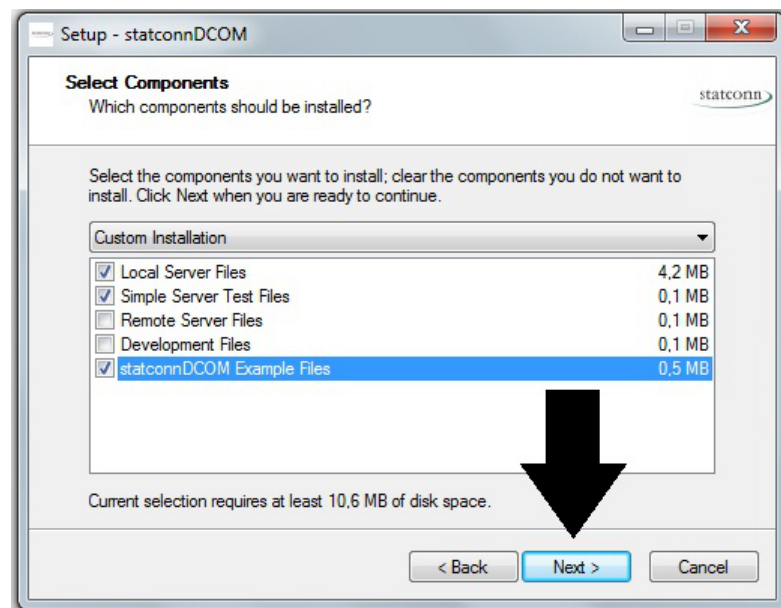


FIGURA 5.6: Passo 3) - Instrução número 6.

7. Esta janela pede ao usuário para fornecer o nome do atalho no menu iniciar para DCOM que sendo instalado. É recomendável usar o nome padrão fornecido, assim clique em *next*

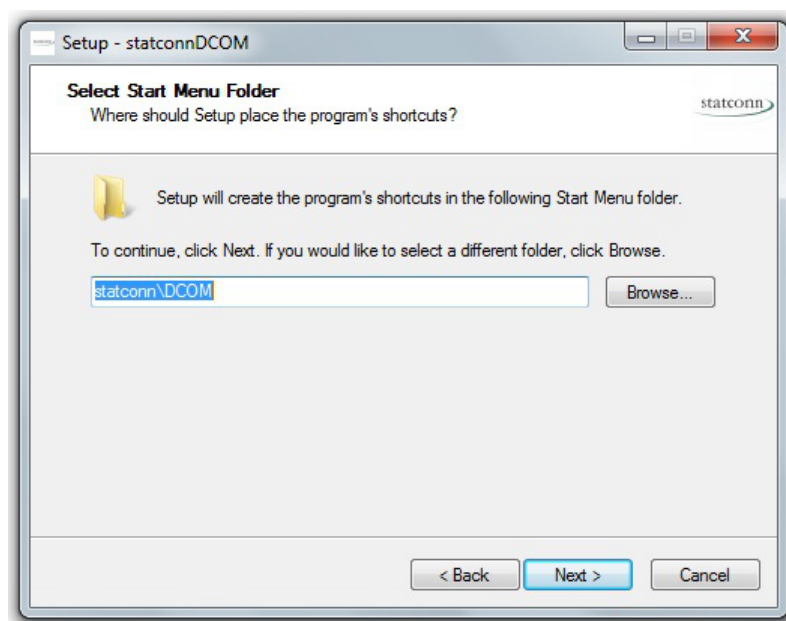


FIGURA 5.7: Passo 3) - Instrução número 7.

8. Estamos próximo de finalizar esse passo de instalação. Clique em *install* no botão indicado pela seta preta

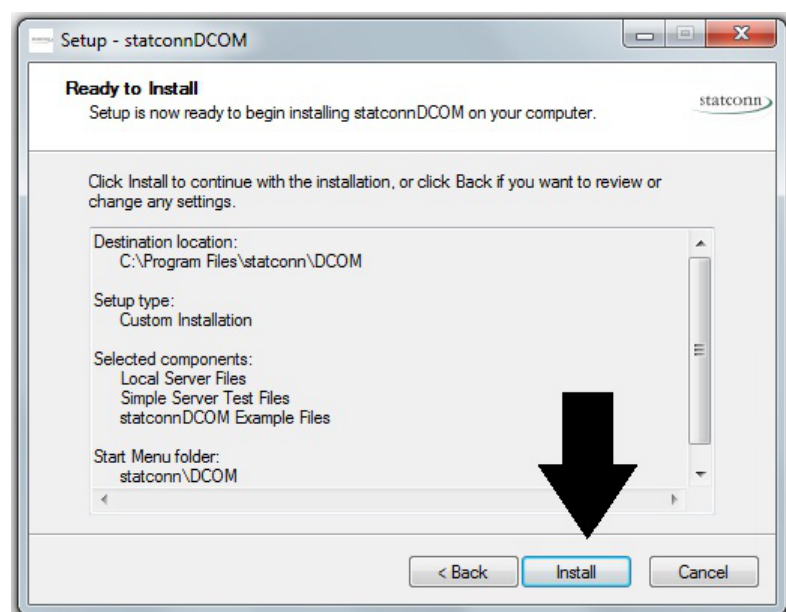


FIGURA 5.8: Passo 3) - Instrução número 8.

9. Desmarque a opção indicada pela seta preta e depois clique em *Finish*. Com esta janela terminamos esta etapa de instalação da primeira fase.

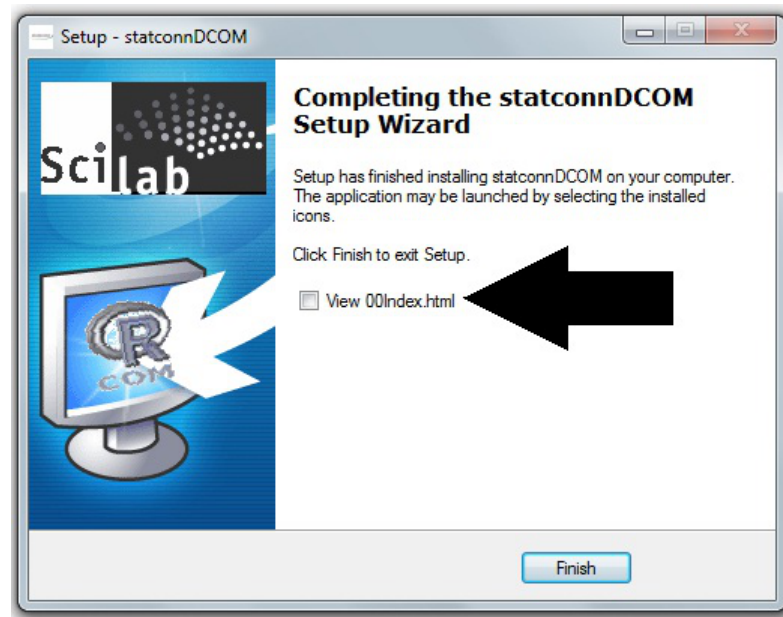


FIGURA 5.9: Passo 3) - Instrução número 9.

Passo 4) digite o seguinte comando no *software* R: “install.packages(“RExcelInstaller”)”

Passo 5) carregue o pacote através do comando: “library(RExcelInstaller)”

Passo 6) digite o seguinte comando no R: “installRExcel()” e siga os passos descritos abaixo:

1. Clique no botão *next* indicado pela seta preta



FIGURA 5.10: Passo 6) - Instrução número 1.

2. Aceite os termos de uso, assinalando a opção indicada pela seta negra

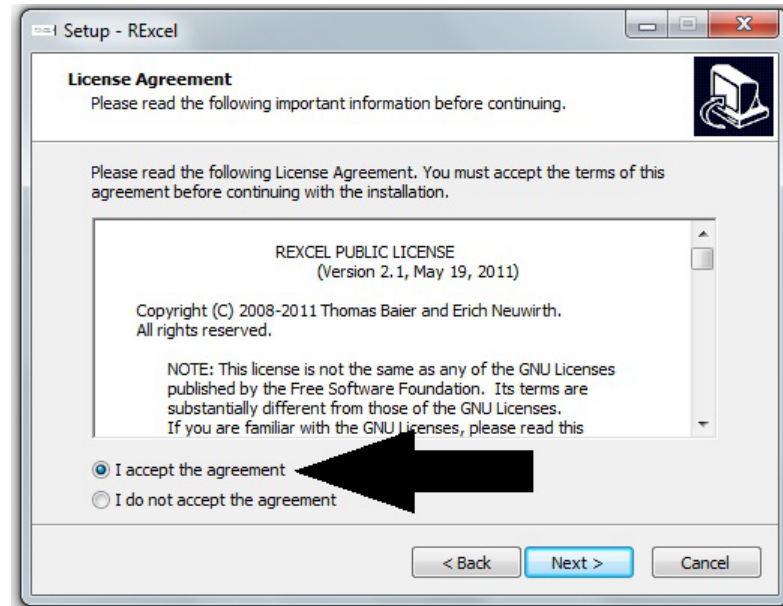


FIGURA 5.11: Passo 6) - Instrução número 2.

3. Clique no botão *next* indicado pela seta preta

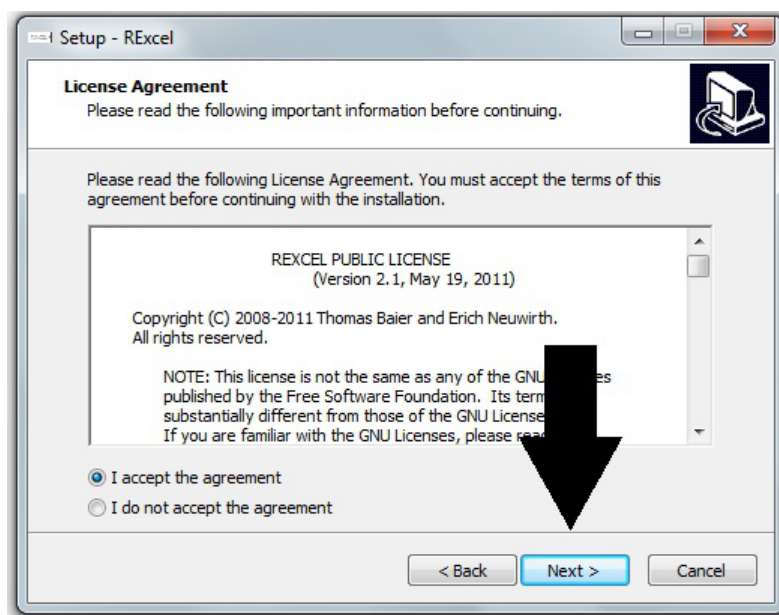


FIGURA 5.12: Passo 6) - Instrução número 3.

4. Esta janela fornece várias informações sobre o RExcel (a interface entre o R e o Excel construída por Baier & Neuwirth (2007)). Um fato importante é que garanta-se o funcionamento do RExcel para as seguintes versões do Excel: Excel 2002, Excel 2003 e Excel 2007. Em outras versões do Excel pode ocorrer bugs. Após uma leitura rápida, clique no botão *next* indicado pela seta preta na figura a seguir

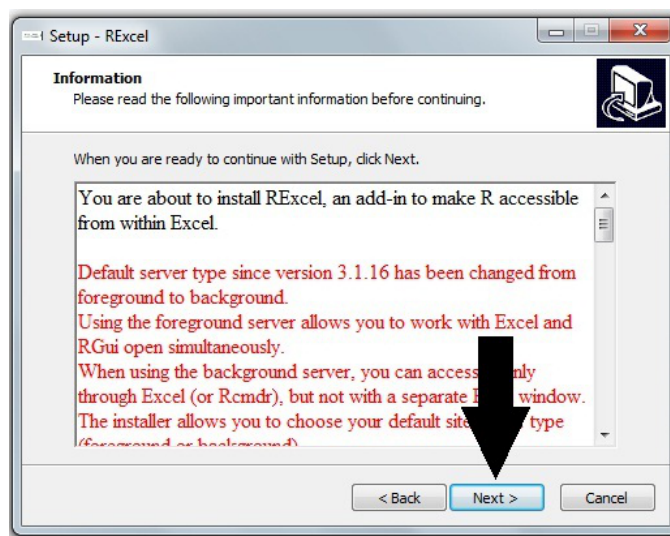


FIGURA 5.13: Passo 6) - Instrução número 4.

5. Esta janela pede ao funcionário para fornecer o diretório que deseja instalar o RExcel: é recomendável usar o padrão fornecido, mas é possível escolher um diretório da preferência do usuário. Após, escolher o diretório, clique no botão *next* indicado pela seta preta

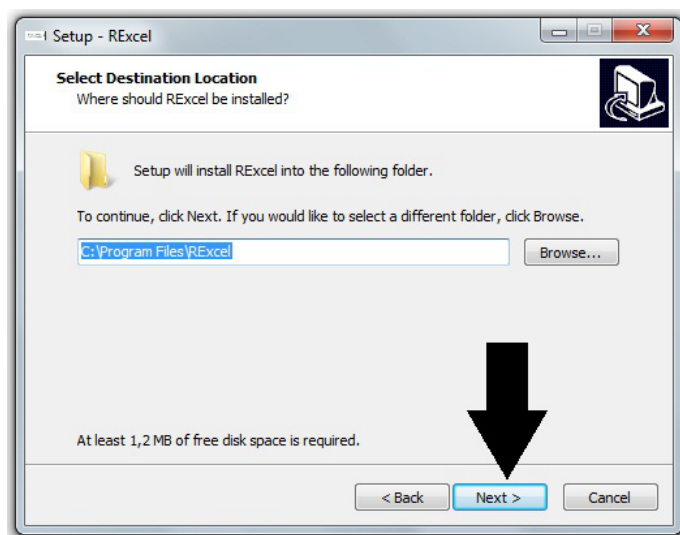


FIGURA 5.14: Passo 6) - Instrução número 5.

6. A seguinte janela pede ao usuário para escolher os componentes do REXcel que deseja instalar. Selecione as opções marcadas abaixo. Após, clique no botão *next* indicado pela seta preta

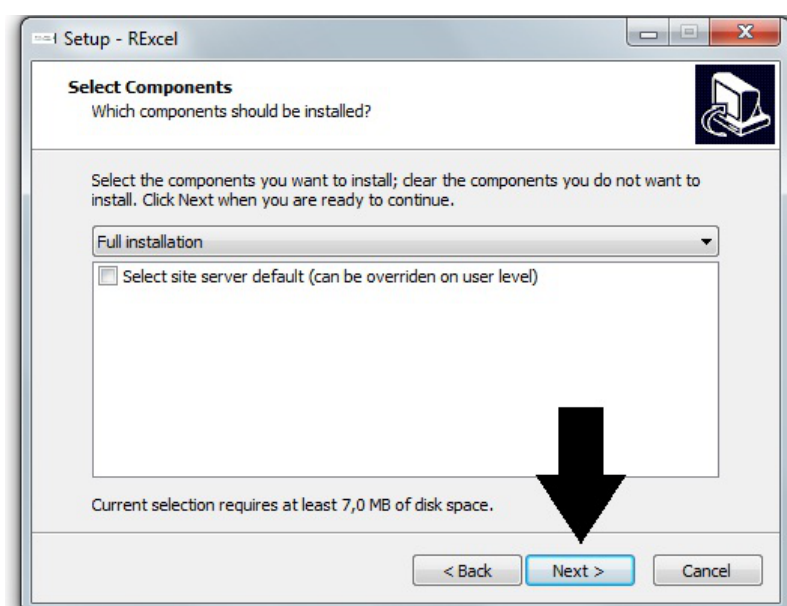


FIGURA 5.15: Passo 6) - Instrução número 6.

7. Estamos perto de finalizar esse passo de instalação. Clique no botão *install* indicado pela seta preta. Aparecerá várias mensagens até que seja mostrada a janela da figura 5.1: em todas clique em *ok*.

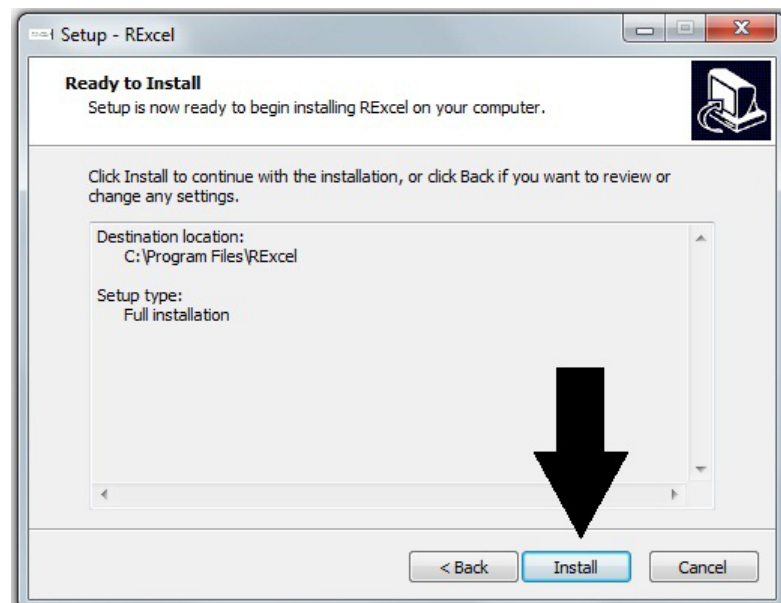


FIGURA 5.16: Passo 6) - Instrução número 7.

8. Se a seguinte mensagem aparecer, o usuário seguiu todos os passos de instalação corretamente e o RExcel foi instalado com sucesso. Clique no botão *Finish*. Com essa etapa finalizada, realizamos a primeira fase de instalação. Na seção 5.1.2 expomos a segunda fase de instalação.

5.1.2 Segunda fase

Nesta seção, apresentamos como instalar a aplicativo desenvolvido neste trabalho, chamado de TAI2U, depois que as configurações da seção 5.1.1 foram realizadas. Para efetuar essa segunda fase de instalação com sucesso, siga as instruções a seguir na ordem apresentada

Passo 1) descompacte o arquivo “TAI.zip”;

Passo 2) escolha um diretório e salve a pasta “TAI” neste diretório, recomendamos que a pasta seja salva em “C:\”;

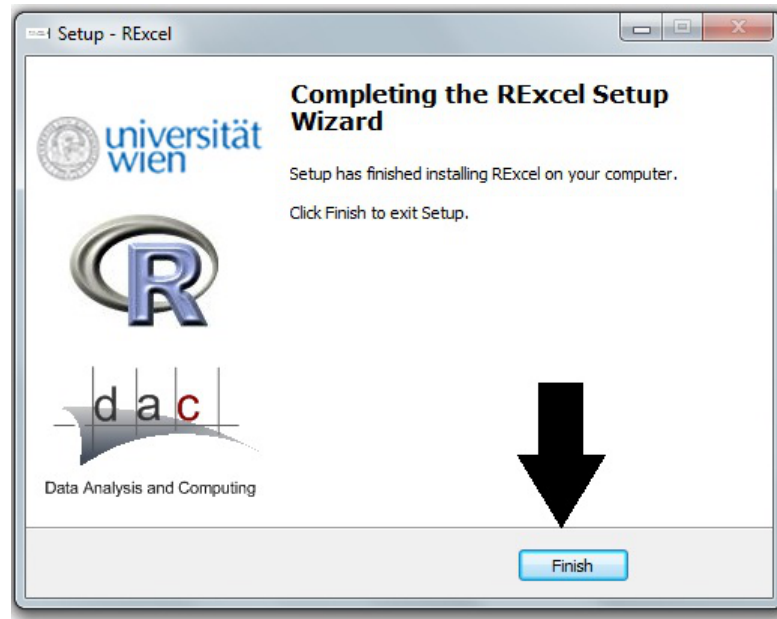


FIGURA 5.17: RExcel instalado com sucesso.

Passo 3) abre o arquivo "TAI2U.xls" e clique em instalação, conforme indicado na figura a seguir

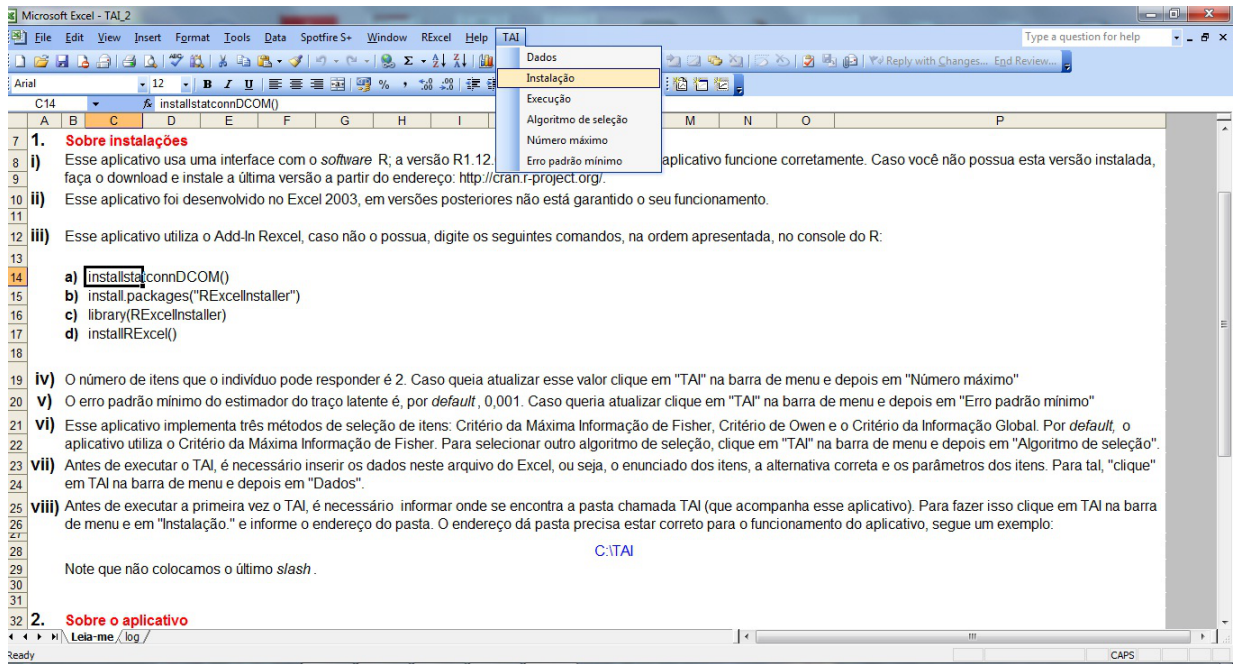


FIGURA 5.18: Instalando o TAI - Instrução número 1.

Passo 4) depois de efetuado o **Passo 3)**, a janela de diálogo da figura a seguir é apresentada. Forneça o diretório que foi salvo a pasta no lugar indicado pela seta vermelha e após clique em *ok*. Por exemplo, na figura abaixo, a pasta foi salva em "C:\\" e devemos digitar "C:\TAI" no campo apontado pela seta vermelha. Notemos que o diretório padrão do aplicativo é "C:\", ou seja, deveríamos digitar "C:\TAI" no lugar indicado na figura, mas esse é o endereço já aparece na janela de diálogo: por isso a recomendação de salvar a pasta em "C:\\" do **Passo 3)**

Após a segunda fase de instalação, o TAI2U já está pronto para uso. O TAI2U vem pronto para uso, ou seja: acompanha o TAI2U um banco de itens ¹, um algoritmo de seleção está selecionado e um critério de parada é estabelecido. Para usar o ML3 no BDI, dicotomizamos as respostas dos itens da seguinte forma: se o indivíduo tem o sintoma mais grave atribuímos 1, caso contrário atribuímos 0. Nas seções a seguir, mostramos como o usuário pode fornecer o banco de itens, estabelecer o critério de parada, escolher o algoritmo de seleção e como pode executar um Teste Adaptativo através do TAI2U.

¹Os itens são questões do Inventário de Depressão de Beck, ou abreviadamente BDI (a abreviação utilizada na literatura é oriunda do inglês).

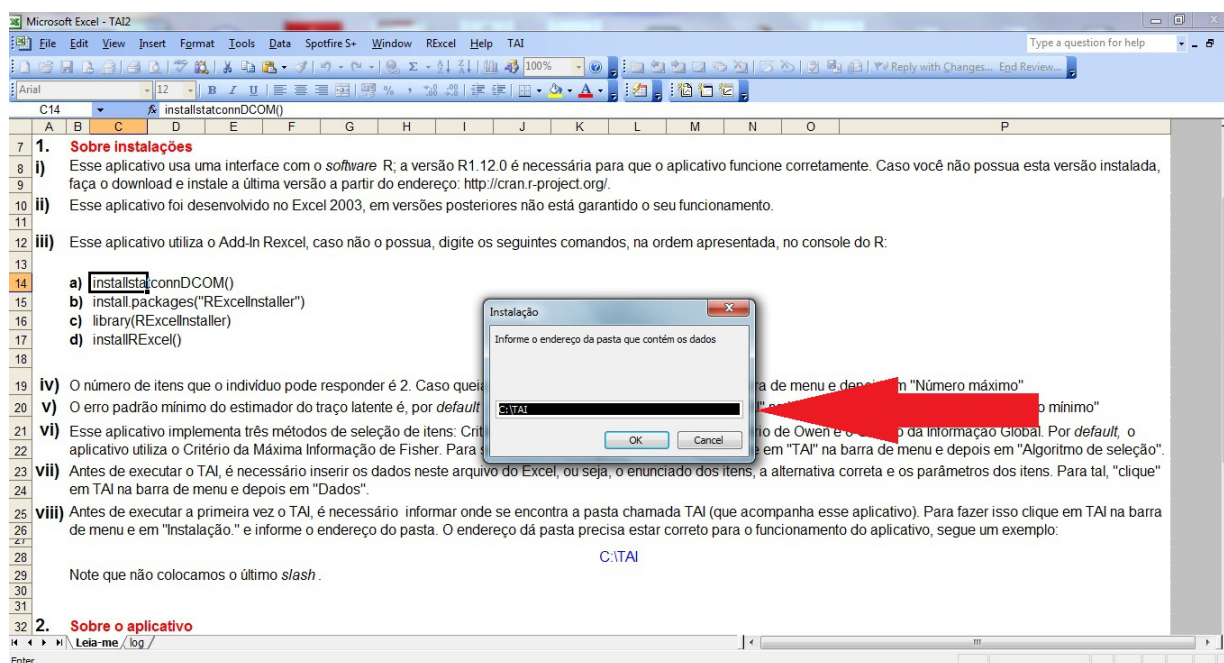


FIGURA 5.19: Instalando o TAI - Instrução número 2.

5.2 Como configurar

Nesta seção, mostramos como o usuário pode fornecer banco de itens, algoritmo de seleção e critério de parada, ou seja, explicamos como realizar a configuração do TAI2U. Após iniciar a configuração, é necessário terminá-la. O TAI2U obriga que o usuário forneça todas as informações, assim configure somente quando tiver banco de itens, tiver escolhido o algoritmo de seleção e tiver estabelecido o critério de parada. Através dos passos a seguir, explicamos como configurar o TAI2U.

Passo 1) Clique em “TAI” na barra de menu e depois em “Dados”, conforme indicado na figura abaixo

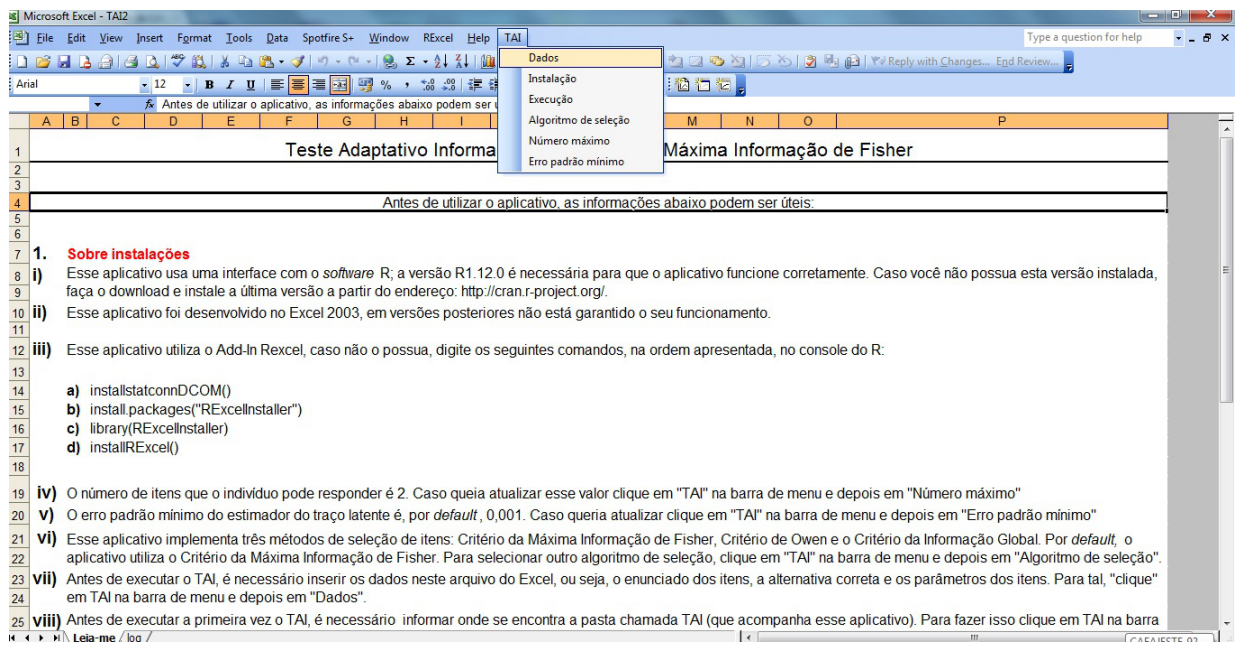


FIGURA 5.20: Configurando o TAI - Instrução número 1.

Passo 2) Na figura abaixo, no local indicado pelo círculo vermelho, informe o número de itens que o usuário fornecerá nesta configuração. Notemos que este número deve ser inteiro. No local indicado pelo círculo azul, clique no algoritmo de seleção escolhido. Após, clique no botão "Ok" indicado pela seta marrom

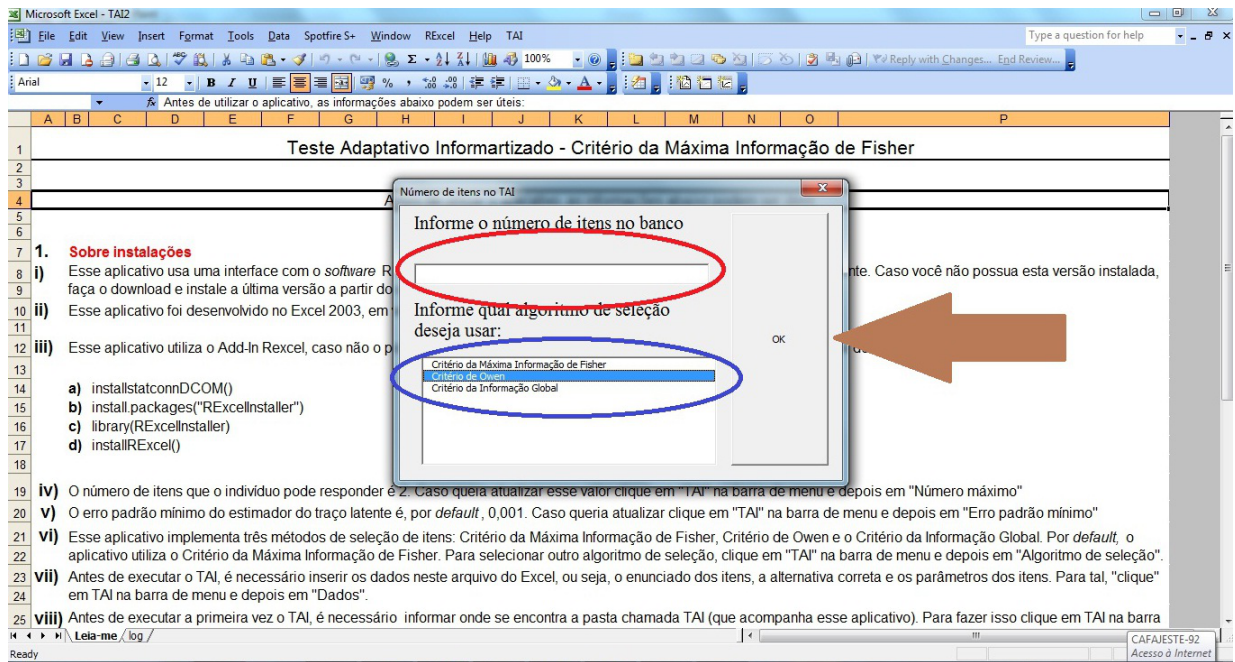


FIGURA 5.21: Configurando o TAI - Instrução número 2.

Passo 3) Neste passo, estabelecemos o critério de parada. No local indicado pelo círculo vermelho, indique o número máximo que um indivíduo poderá responder. No lugar indicado pelo círculo azul indique o valor mínimo do erro padrão do estimador da habilidade. Atenção neste ponto para a casa decimal do sistema operacional: verifique se as casas decimais são separadas por ponto ou por vírgula . Após, clique no botão “Ok” indicado pela seta preta

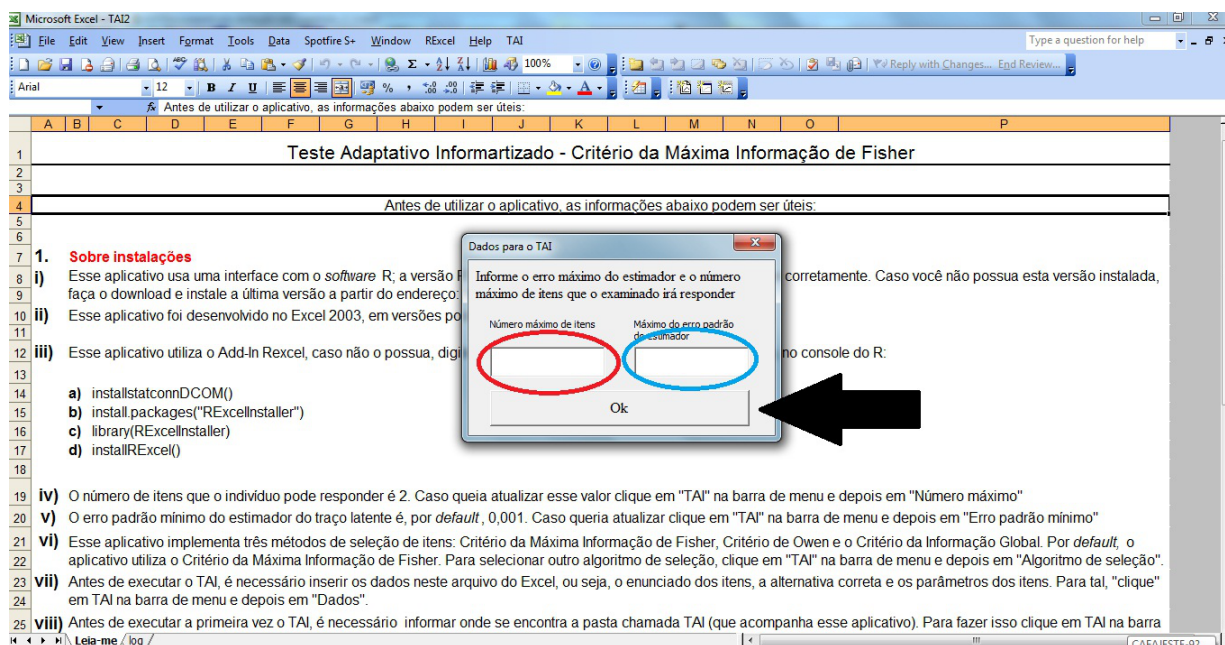


FIGURA 5.22: Configurando o TAI - Instrução número 3.

Passo 4) Com esta etapa encerramos a configuração do TAI2U. Neste passo, incluímos no TAI2U o banco de itens, ou seja, para todos os itens do TAI fornecemos o enunciado, as alternativas, a dificuldade, a discriminação, o parâmetro de acerto casual. Notemos que é necessário prover N itens, em que N foi fornecido no **Passo 2)**. Na figura a seguir: insira o enunciado do item no local indicado pelo círculo preto; insira as alternativas no local indicado pelo círculo amarelo; insira a dificuldade, a discriminação e o parâmetro de “chute” no local indicado pelo círculo azul; clique na alternativa correta no local indicado pelo círculo marrom; após todas as inclusões, clique no botão “ok” no lugar indicado pelo seta negra

5.3 Como executar

Nesta última seção deste trabalho, mostramos como utilizar o TAI2U após a instalação (vide 5.1) e configurações (vide 5.2). O uso do TAI é simples e é dividido em três passos:

Passo 1) Abra o “TAI2U”;

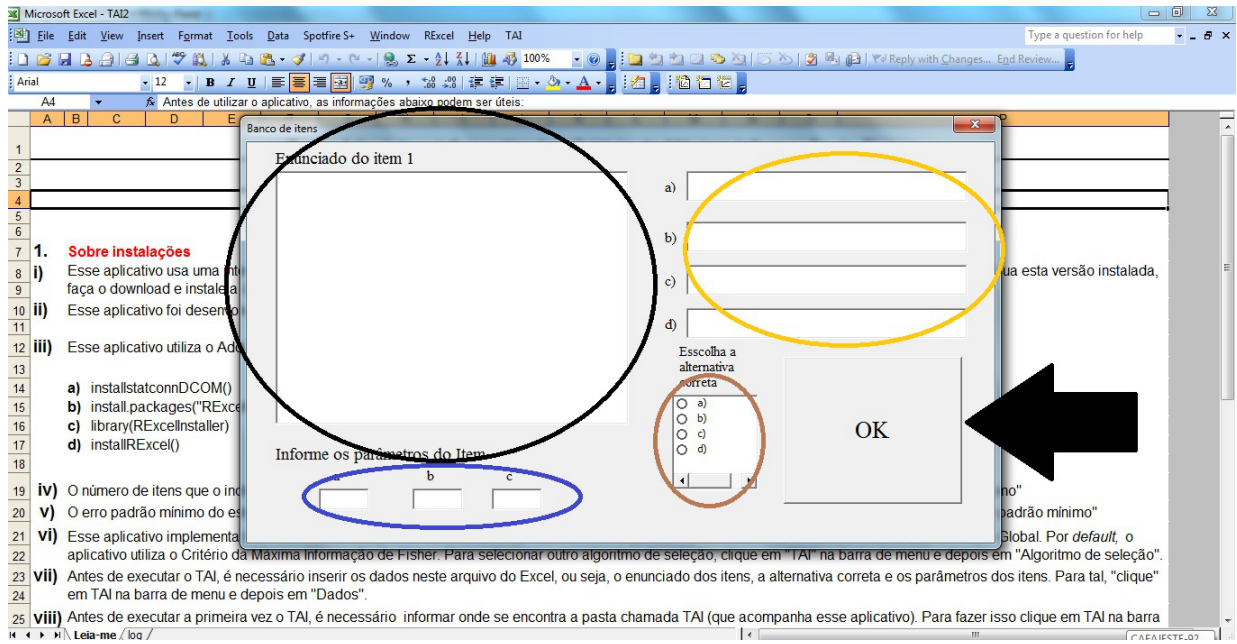


FIGURA 5.23: Configurando o TAI - Instrução número 4.

Passo 2) Clique em "TAI" na barra de menu e depois em "Execução" conforme a figura abaixo

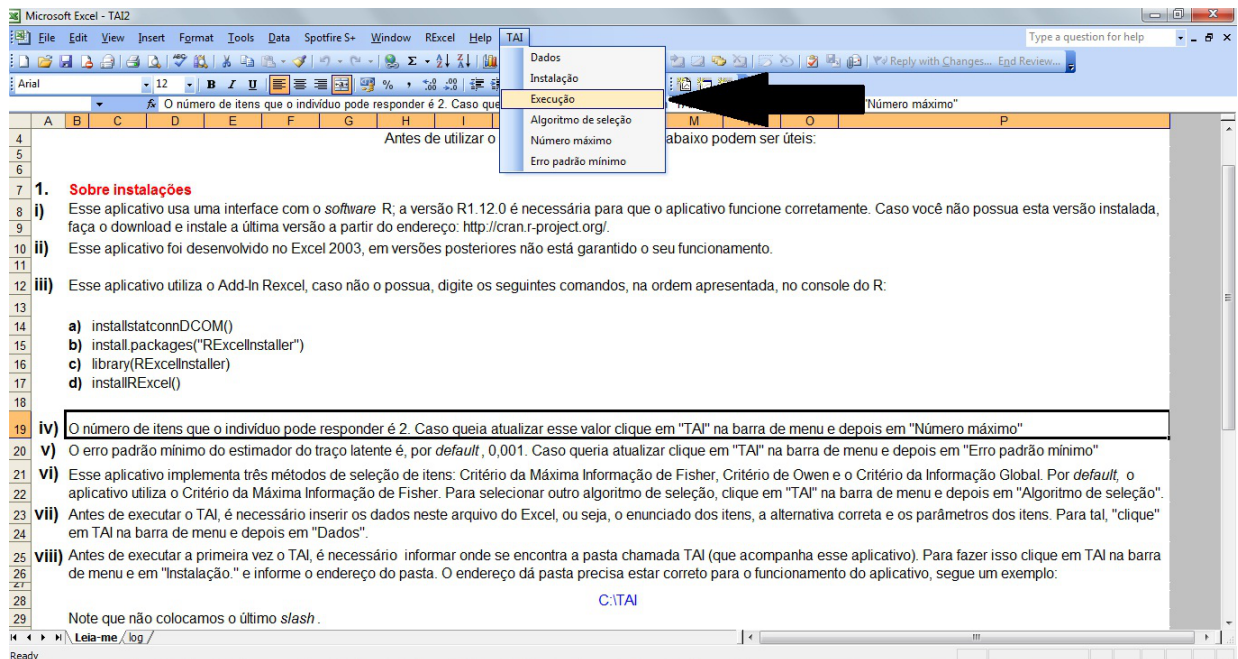


FIGURA 5.24: Executando o TAI - Instrução número 1.

Passo 3) Execute as seguintes etapas até que o critério de parada seja satisfeito: leia o enunciado no local indicado quadrado azul na figura abaixo e clique na alternativa correta no local indicado pelo quadrado amarelo na figura a seguir e, depois, clique no botão “ok” indicado pela seta preta

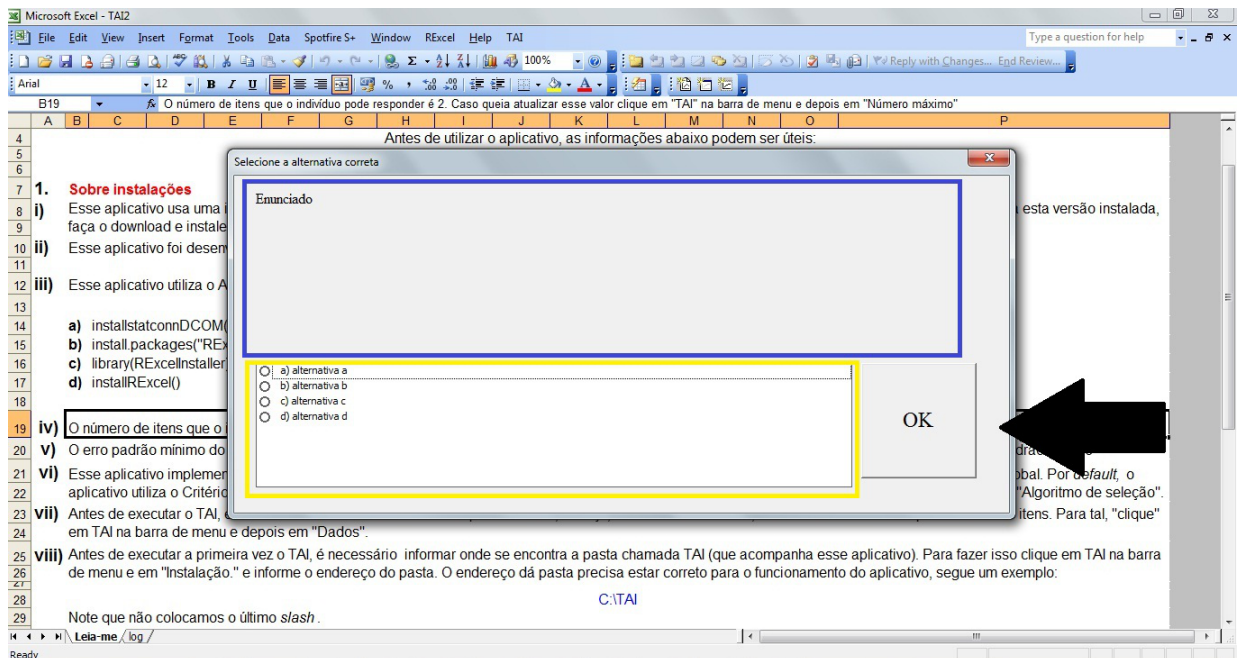


FIGURA 5.25: Executando o TAI - Instrução número 2.

Referências

- Andrade, D., Tavares, H. & da Cunha Valle, R. (2000). Teoria da Resposta ao Item: conceitos e aplicações. *ABE, Sao Paulo*.
- Baier, T. & Neuwirth, E. (2007). Excel:: Com. *Computational Statistics*, **22**(1), 91–108.
- Baker, F. (2001). *The basics of item response theory*. For full text: <http://ericae.net/irt/baker>. ISBN 1886047030.
- Baker, F. (2004). *Item response theory: Parameter estimation techniques*, volume 176. CRC.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, pages 397–479.
- Boekkooi-Timminga, E. (1991). A method for designing rasch model based item banks. In *annual meeting of the Psychometric Society, Princeton, NJ*.
- Casella, G. & Berger, R. (2001). *Statistical inference*. Duxbury Press.
- Chang, H. & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, **20**(3), 213.
- Chang, H. & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, **23**(3), 211.
- Cox, D. (2006). *Principles of statistical inference*. Cambridge Univ Pr.
- Davey, T. & Nering, M. (1998). Controlling item exposure and maintaining item security. *Paper presented at the ETS-sponsored colloquium entitled Computer-based testing: Building the foundations for future assessments, Philadelphia PA*.
- DeMars, C. (2010). *Item response theory*. Oxford Univ Pr.
- Flaughner, R. (1990). Item pools. *Computerized adaptive testing: a primer*, pages 41–64.
- Fragoso, T. M. (2010). *Modelos multidimensionais da teoria de resposta ao item*. ICMC - USP.
- Galassi, M., Gough, B., Jungman, G., Theiler, J., Davies, J., Booth, M. & Rossi, F. (2007). The gnu scientific library reference manual, 2007. URL <http://www.gnu.org/software/gsl>. ISBN, **954161734**.

- Hambleton, R. & Swaminathan, H. (1989). *Item response theory*. Kluwer-Nijhoff.
- Kingsbury, G. & Zara, A. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, **2**(4), 359–375.
- Linden, W. & Pashley, P. (2010). Item selection and ability estimation in adaptive testing. *Elements of Adaptive Testing*, pages 3–30.
- Lord, F., Novick, M. & Birnbaum, A. (1968). Statistical theories of mental test scores.
- McBride, J. & Martin, J. (1983). Reliability and validity of adaptive ability tests in a military setting. *New horizons in testing*, pages 223–226.
- Meijer, R. & Nering, M. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, **23**(3), 187.
- Ostini, R. & Nering, M. (2006). *Polytomous item response theory models*. Number 144. Sage Publications, Inc.
- Owen, R. (1969). A bayesian approach to tailored testing. *Research Bulletin*, pages 69–92.
- Owen, R. (1975). A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, **70**(350), 351–356.
- RDevelopment, C. (2008). R: A language and environment for statistical computing. URL <http://www.R-project.org>.
- Stocking, M. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics*, **21**(4), 365.
- Sympton, J. & Hetter, R. (1985). Controlling item-exposure rates in computerized adaptive testing. In *Proceedings of the 27th annual meeting of the Military Testing Association*, pages 973–977.
- Van Der Linden, W. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, **63**(2), 201–216.
- Van Der Linden, W. (2000). *Computerized adaptive testing: Theory and practice*. Springer Netherlands. ISBN 0792364252.
- Van Der Linden, W. (2010). *Elements of adaptive testing*. Springer Verlag.
- Van Der Linden, W. & Veldkamp, B. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, **29**(3), 273.
- Van Der Linden, W. & Veldkamp, B. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics*, **32**(4), 398.
- Van Der Linden, W., Veldkamp, B. & Reese, L. (2000). An integer programming approach to item bank design. *Applied Psychological Measurement*, **24**(2), 139.

-
- Veerkamp, W. & Berger, M. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, **22**(2), 203.
- Wainer, H. (2000). Computerized adaptive testing. *Corsini Encyclopedia of Psychology*.
- Wainer, H. (2001). Computer-adaptive testing: a primer. *Language Learning & Technology*, **5**(2), 23–27.
- Wainer, H. & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, **24**(3), 185–201.
- Weiss, D. & Yoes, M. (1991). *Item response theory*. Kluwer Academic/Plenum Publishers.