Estimação do grau de habilidade em testes de múltipla-escolha

Gustavo Gerald Toja Frachia

Dissertação apresentada ao Instituto de Matemática
e Estatística da Universidade de São Paulo
para a obtenção do Grau de Mestre em Estatística

Área de Concentração: Estatística

Orientador: Prof. Dr. Carlos Alberto de Bragança Pereira

São Paulo, julho de 2001

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por Gustavo Gerald Toja Frachia
e aprovada pela comissão julgadora.

São Paulo, 6 de julho de 2001

Banca examinadora:

Prof. Dr. Carlos Alberto de Bragança Pereira (Orientador) - IME - USP

Prof. Dr. Flávio Wagner Rodrigues - IME -USP

Prof. Dr. José Afonso Mazzon - FEA - USP

Agradecimentos

"Gracias a la vida,
que me ha dado tanto..."
Violeta Parra, compositora chilena

Agradeço os ensinamentos dos meus pais que, desde a primeira infância, me mostraram que o esforço e a honestidade são dois dos principais ingredientes do caráter de uma pessoa. Creio tê-los incorporado à minha personalidade. Ensinaram-me também muitas outras coisas fundamentais para tornar-me um grande homem, mas, lamentavelmente, não as aprendi muito bem.

Agradeço ao meu primeiro professor de Estatística, Orual Andina, do curso de Psicologia da Universidade Católica do Uruguai, pela seriedade com a que encarou o programa da disciplina e que motivou-me a continuar a estudar Estatística.

Agradeço ao professor Carlos Alberto de Bragança Pereira por ter me dado uma *oportunidade de vida*. Poucos são os seres humanos que têm a chance ou a capacidade ou a vontade de dar *oportunidades de vida* aos outros.

Agradeço a minha esposa Ângela, pelo seu incentivo nestes três anos de Mestrado.

Finalmente, agradeço aos meus filhos Matilde, Felipe e Gabriel por serem o combustível que faz o motor da minha vida funcionar em altas rotações.

Abstract

The true level of proficiency of an individual who is tested using a multiple-choice test is studied in this work. Usually, the portion of answers that the respondent really knows is neglected, regarding only the total of correct answers.

Grades, in many tests alike Exame Nacional de Cursos, are fixed in an arbitrary way, taking into account the results of the hole set of grades of the individuals who were tested (standarization). This last aspect may produce great distortions, prizing bad performances or punishing good ones.

The proposal of this work is that grades must be obtained taking into account the intrinsical test characteristics and, also, the minimal individual skills required for that test.

Resumo

O verdadeiro grau de habilidade de um indivíduo que é submetido a um teste de múltipla-escolha é estudado neste trabalho. É comum observar que quem avalia o teste negligencie a porção de respostas que o respondente consegue acertar devido ao acaso, observando apenas o número de respostas corretas.

As notas de aprovação de muitos testes - como as do Exame Nacional de Cursos - são determinadas de forma arbitrária, levando em consideração os resultados do conjunto de indivíduos submetidos à prova (padronização das notas). Este último aspecto pode levar a sensíveis distorções na atribuição das notas, premiando maus resultados ou penalizando bons.

Este trabalho propõe que estas notas de coorte sejam determinadas levando em consideração as características do teste e também as habilidades mínimas que o respondente deveria possuir para ser considerado aprovado.

Sumário

Capítulo 1

Introdução

Propósito do Estudo

Uma das formas mais freqüentes de avaliação do rendimento escolar é o teste de múltipla-escolha. Tal teste consiste na administração de um conjunto de itens nos quais é feita a formulação de um enunciado ou pergunta e é proposta uma série de alternativas, das quais, geralmente, apenas uma resposta é correta.

A possibilidade de acerto casual geralmente disfarça o verdadeiro grau de conhecimento (habilidade) do indivíduo. Por este motivo, é relevante o propósito de investigar, dado o número de respostas corretas que o indivíduo obteve, qual a distribuição de probabilidades do número de respostas que o mesmo realmente sabe.

Vamos supor que temos uma caixa que contém itens do seguinte tipo:

Assinale um x na alternativa de resposta que você considera como correta para a seguinte pergunta:

A Olimpíada de 1956 foi realizada em qual cidade?

___ a) Tóquio.
___ b) Helsinski.
___ c) Atlanta.
___ d) Moscou.

Vamos supor também que a caixa contém outros itens como o de cima para cada Olimpíada realizada na Era Moderna.

Consideremos um indivíduo que irá responder perguntas desta natureza. É razoável supor que este indivíduo conhece um certo número de itens da nossa caixa imaginária.

Se desejássemos saber o número exato de itens que o indivíduo conhece, na referida caixa, nos encontraríamos em dificuldades. Isto porque existe uma chance de acerto por escolha casual. Por outro lado, pode descartar as alternativas que ele sabe que são incorretas. Por exemplo, ele pode saber que as Olimpíadas de Tóquio, Atlanta e Moscou **não** foram realizadas em 1956, de tal forma que sua única opção possível é a alternativa b. Esse conhecimento das opções incorretas ajuda a escolher a correta, aumentando a probabilidade de acerto.

É claro que essa dificuldade de determinar o número verdadeiro daqueles itens que ele sabe a resposta, não impede que procuremos estabelecer a nossa opinião sobre as possíveis opções para esse número. De fato, representaremos nossas opiniões por meio de probabilidades.

Se o número de itens da caixa for pequeno, poderemos fazer com que o indivíduo responda a todos eles. Assim, teríamos de estimar o número daqueles itens que tiveram resposta correta, mas por escolha casual.

Caso a análise exaustiva dos itens da caixa não seja recomendável pelo grande tamanho da caixa, podemos pensar em fazer a retirada de uma amostra aleatória, chamada de teste, do total de itens. Por teste consideramos qualquer conjunto finito de itens submetidos a um respondente. O número de respostas corretas ao teste, as probabilidades de acerto aos itens e o nosso conhecimento da história do indivíduo, servirão então para estimarmos o número de itens que o indivíduo conhece a resposta.

Os procedimentos para a construção de testes têm sido objeto de exaustivo estudo. Podemos destacar os trabalhos de Harris e Steward (1971), Glaser e Nitko (1971), Popham (1975), Millman (1974) e Hively et al. (1973). A proposta destes autores é a de definir apropriadamente o conjunto de itens que contém todas as questões possíveis sobre determinado assunto, similar àquele descrito na nossa caixa imaginária. Este conjunto é normalmente conhecido como "domínio". O "nível de proficiência" ou o "verdadeiro nível de habilidade" corresponde ao número de itens que o indivíduo conhece do domínio (Millman, 1972 e 1974).

Vários autores, dentre eles Novick e Lewis (1971), Kriewall (1972), Millman (1974), Hamilton (1975), Visco (1977) e Zúñiga (1978), propõem alternativas para a estimação do verdadeiro nível de habilidade de um certo indivíduo, levando em consideração o seu rendimento em um teste com itens extraídos aleatoriamente do domínio.

A principal divergência entre alguns destes autores está na forma em que a amostragem é realizada. Millman (1972 e 1974), Kriewall (1972) e Novick e Lewis (1974) desenvolveram teorias sob a suposição de que os itens do teste são extraídos *com* reposição do Domínio. Hambleton (1975) adicionou a suposição de que as respostas do respondente provêm de uma população distribuída normalmente.

Visco (1977) e Zúñiga (1978) preferem adotar o modelo hipergeométrico, isto é, assumem que os itens são escolhidos aleatoriamente do domínio, mas *sem* reposição. É claro que quando o número de itens da caixa é grande ou desconhecido, a suposição de reposição não é absurda. O modelo sem reposição parece ser mais realístico pois, para domínios relativamente pequenos, um determinado item teria alta chance de ser selecionado por mais de uma vez. Evidentemente só seria incluído no teste uma vez.

Na mesma linha de pesquisa, mas com uma modelagem matemática bastante diferente, a *Teoria da Resposta ao Item* utiliza técnicas "para representar a probabilidade de um indivíduo dar uma resposta certa a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente" (Andrade, Tavares & Valle, 2000). A *Teoria da Resposta ao Item* (TRI) se preocupa principalmente com as características do próprio item e com a habilidade que o respondente tem para esse item. O domínio de onde os itens são extraídos e o teste como um todo, não são os focos do modelo TRI. Sem dúvida, uma das grandes contribuições da TRI é o fato de podermos "comparar duas ou mais populações de respondentes, desde que tenham respondido itens comuns", ou ainda comparar "indivíduos da mesma população que tenham sido submetidos a testes diferentes" Andrade et al. (2000). Por afastar-se bastante do enfoque escolhido para o nosso estudo, não nos aprofundaremos nas peculiaridades da TRI. Recomendamos ao leitor a seguinte literatura sobre o tema: Lord (1980), Mislevy (1986), Goldstein e Wood (1989), Baker (1992), Linden e Hambleton (1997), Andrade e Valle (1998) e Andrade, Tavares e Valle (2000).

O suporte teórico principal do nosso estudo está baseado na análise bayesiana de dados categorizados, em especial no que se relaciona com o estudo do problema da não-resposta ou censura. Para o desenvolvimento do nosso trabalho iremos nos remeter a Basu & Pereira (1982) e Paulino & Pereira (1995).

Capítulo 2

Modelo Hipergeométrico

Neste capítulo apresentaremos a abordagem do problema de estimar o verdadeiro nível de habilidade de um indívíduo, segundo a suposição de não-reposição.

2.1 O problema

Dados:

a) Um domínio de N itens de múltipla escolha, no qual cada item i ($i=1,\ldots,N$) contém m_i respostas alternativas.
b) Uma partição do domínio em subdomínios ou categorias, na qual cada categoria k está composta por itens permutáveis com respeito à probabilidade de acerto.
c) Um teste construído a partir da seleção de amostras de tamanho $n_{(i)}$ ($i= 1,\ldots,k$), de cada categoria do domínio. O tamanho do teste é

$$n = n_{(1)} + n_{(2)} + \ldots\ldots + n_{(k)}$$

O teste é administrado a um indivíduo que conhece Φ itens do domínio e desconhece $N - \Phi$ itens do mesmo domínio. O número de respostas corretas na categoria k será notado com $x_{(k)}$.
d) Para cada item do teste, o indivíduo escolherá a alternativa certa se ele souber a resposta e, caso não saiba a resposta, tentará o acerto com probabilidade, na hipótese de ignorância total, igual a $1/m_i$.

O nosso objetivo para este estudo consiste nos seguintes pontos:

1- Determinar a distribuição de probabilidades de $\mathbf{x_1}$ (ver tabela 2), o número de respostas que o indivíduo sabe no teste, como função do número de itens que ele acertou $\mathbf{x}$ e dos parâmetros de acerto casual e habilidade.

2- Determinar a distribuição de probabilidades de Φ (ver tabela 1), o número de respostas que o indivíduo sabe no domínio, em função do tamanho de respostas certas no teste $\mathbf{x}$ e dos parâmetros de acerto casual e habilidade.

3- Obter uma medida de informação para avaliar a precisão de um teste específico.

   A medida de informação a ser utilizada, que chamaremos função de incerteza, será a variância do estimador de Bayes:

$$\mathrm{Var}(\hat{\Phi}) = \mathrm{Var}(\Phi) - E\,\{\mathrm{Var}(\Phi) \mid X\}$$

Podemos representar por meio de uma tabela de dupla entrada a distribuição de freqüências dos valores de interesse, tanto no domínio, quanto no teste:

No **DOMÍNIO**

	CERTAS	ERRADAS	TOTAL
SABE	Φ	0	Φ
NÃO SABE	$\Theta - \Phi$	$N - \Theta$	$N - \Phi$
TOTAL	Θ	$N - \Theta$	N

Tabela 1

No **TESTE**

	CERTAS	ERRADAS	TOTAL
SABE	x_1	0	x_1
NÃO SABE	$x - x_1$	$n - x$	$n - x_1$
TOTAL	x	$n - x$	n

Tabela 2

O problema que estamos estudando refere-se a testes compostos por itens de múltipla-escolha o que não é uma limitação. Poderemos generalizar a solução também para os casos em que o acerto casual não é possível, ou seja onde tem-se probabilidade zero de acerto aleatório. Este extremo será apenas um caso limite das equações inferidas.

2.2 - Notação

N – Número de itens do domínio.

N_r – Número de itens da partição r.

P (.) – Probabilidade do evento ou expressão contida nos parênteses.

(*a* ; *b*) – Representa o número de conjuntos que podem ser formados com *b* elementos, a partir de um conjunto de *a* elementos, as chamadas *Combinações*, que são calculadas assim:

$$(a\ ;\ b) = \frac{a!}{b!\ (a-b)!}$$

$\perp\!\!\!\perp$ - Símbolo que representa a independência entre duas variáveis.

Π – Probabilidade de responder corretamente um item cuja resposta correta é desconhecida pelo indivíduo que está sendo testado.

x – Número de respostas corretas no teste.

x_k – Número de respostas corretas para a categoria K do teste.

n – Número de itens no teste.

n_k - Número de itens na categoria k

n - x – Número de respostas incorretas no teste.

$n_k - x_k$ – Número de respostas incorretas para a categoria K do teste.

Notação adicional será introduzida ao longo do presente estudo e definida oportunamente.

2.3 Definições sobre as distribuições de probabilidades relevantes a este estudo

Definição 1 – Uma quantidade aleatória p tem uma função de probabilidades Beta com parâmetros reais positivos α e β e escrevemos p ~ Be(α,β), se sua função densidade de probabilidade for:

$$f(p) = \frac{p^{\alpha-1}\ (1-p)^{\beta-1}}{B(\alpha,\ \beta)}$$

Com $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

A generalização da distribuição Beta para um vetor k-dimensional positivo **p** = (p_1, ..., p_k) tal que $\sum_{i=1}^{k} p_i = 1$ é denominada distribuição de **Dirichlet** de parâmetros reais não-negativos (α_1, α_2, ..., α_k).

Definição 2 – Sejam α e β números reais positivos e n um inteiro positivo. Uma quantidade aleatória x, $0 \leq x \leq n$, tem função de probabilidades Beta-binomial com parâmetros n, α e β, se sua função densidade de probabilidade for:

$$f(x) = (n\ ;\ x)\ \frac{B(\alpha + x,\ \beta + n - x)}{B(\alpha,\ \beta)}$$

A generalização da distribuição Beta-binomial para um vetor de inteiros não-negativos $\mathbf{x} = (x_1, ..., x_k)$ é chamada de distribuição de Dirichlet – Multinomial, cuja função de probabilidades é dada por:

$$f(x) = \frac{n!\,\Gamma(\alpha)}{\Gamma(\alpha+n)} \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + x_i)}{x_i!\,\Gamma(\alpha_i)} \qquad (2.1)$$

onde $\alpha = \sum_{i=1}^{k} \alpha_i$.

2.4 Distribuição de Probabilidades para o número de questões que o indivíduo sabe (x_1) dado o número de respostas certas (x)

Seja $\mathbf{X} = (x_1,.....,x_k)$ um vetor de dados distribuído como Dirichlet Multinomial com parâmetros $(n; \alpha_1, ..., \alpha_k)$.

Proposição 1. Se $(i_1, ..., i_k)$ é uma permutação de $(1, ..., k)$ então $(x_{i1}, ...x_{ik}) \sim DM(n; \alpha_{i1}, ..., \alpha_{ik})$.

Proposição 2. Se $m \in \{1,2,......,k\}$ é fixado, então para $\beta = \sum_{i=1}^{m} \alpha_i$

$$(x_1,.....,x_m, n - \sum_{i=1}^{m} x_i) \sim DM(n; \alpha_1, ..., \alpha_m, \alpha - \beta)$$

e

$$x = \sum_{i=1}^{m} x_i \sim BB(n; \beta, \alpha - \beta)$$

Estes dois resultados são conseqüências imediatas de propriedades análogas das distribuições Multinomial e Dirichlet.

Proposição 3. Para m e x definidos como acima, temos que:

$$(x_1,.....,x_m) \mid x \sim DM(x; \alpha_1, ..., \alpha_m)$$

Prova – Observe-se que a distribuição de probabilidades condicional $(x_1,\ldots,x_m) \mid x$ é obtida pela divisão da função de probabilidades de $(x_1,\ldots,x_m, n - x)$ e a função de probabilidades de x, a qual fornece a distribuição de probabilidades de uma DM $(x; \alpha_1, \ldots, \alpha_m)$. □

O resultado seguinte é uma importante caracterização da Distribuição Dirichlet-Multinomial.

Seja $(x_1,\ldots,x_k)$ um vetor aleatório de inteiros não-negativos com $\sum_{i=1}^{k} x_i = n$ fixo. Escolha um inteiro m ε {2,......,k-1} e denote $x = \sum_{i=1}^{m} x_i$ e $y = n - x$. Considere agora o seguinte conjunto de condições:

$$(x_1,\ldots,x_m) \perp\!\!\!\perp (x_{m+1},\ldots,x_k) \mid x$$

$$(x_1,\ldots,x_m) \mid x \sim \mathrm{DM}\,(x;\ \alpha_1,\ \ldots,\ \alpha_m) \qquad (2.2)(i)$$

$$(x_{m+1},\ldots,x_k) \mid x \sim \mathrm{DM}\,(y;\ \alpha_{m+1},\ \ldots,\ \alpha_k), \qquad (2.2)(ii)$$

e

$$x \sim \mathrm{BB}(n;\ \sum_{i=1}^{m} \alpha_i,\ \alpha - \sum_{i=1}^{m} \alpha_i) \qquad (2.2)(iii)$$

Teorema 1. O conjunto de condições (2.2) é necessário e suficiente para

$$(x_1,\ldots,x_k) \sim \mathrm{DM}\,(n;\ \alpha_1,\ \ldots,\ \alpha_k) \qquad (iv)$$

Prova – Pelas proposições 1, 2 e 3, (iv) implica (ii) e (iii). Para provar as outras implicações precisamos notar apenas que (2.1) pode ser fatorado como:

$$f(x) = \frac{n!\,\Gamma(\alpha)}{\Gamma(\alpha+n)} \quad \frac{\Gamma(\beta+x)\,\Gamma(\alpha-\beta+y)}{x!\,y!\,\Gamma(\beta)\,\Gamma(\alpha-\beta)} \quad \mathrm{x}$$

$$\mathrm{x} \quad \frac{x!\,\Gamma(\beta)}{\Gamma(\beta+x)} \prod_{i=1}^{m} \frac{\Gamma(\alpha_i+x_i)}{x_i!\,\Gamma(\alpha_i)} \quad \mathrm{x}$$

$$\mathrm{x} \quad \frac{y!\,\Gamma(\alpha-\beta)}{\Gamma(\alpha-\beta+y)} \prod_{i=m+1}^{k} \frac{\Gamma(\alpha_i+x_i)}{x_i!\,\Gamma(\alpha_i)}$$

Onde, como antes, $\alpha = \sum_{i=1}^{k} \alpha_i$ e $\beta = \sum_{i=1}^{m} \alpha_i$. O primeiro fator é a função de probabilidades de uma BB(n; β, $\alpha - \beta$), a segunda é uma função de probabilidades de uma DM(x; α_1, ..., α_m) e, finalmente, a terceira é a função de probabilidades de uma DM(y; α_{m+1}, ..., α_k). □

Aplicando os resultados anteriores ao nosso problema, observamos que **a distribuição de probabilidades do número de respostas corretas que o indivíduo sabe (x_1) dado o número de respostas corretas (x) é uma BB (x; α_1, α_2).**

2.5 A Estimação do Grau de Habilidade em Testes de Múltipla-Escolha

Vamos voltar à idéia de que um certo *domínio* de itens de múltipla-escolha, é passível de ser particionado em categorias de acordo com, por exemplo, itens de diferentes tipos de dificuldade, de diferentes disciplinas ou qualquer outro critério que o divida em categorias disjuntas.

Para cada categoria, associaremos dois valores: o primeiro corresponde à probabilidade de o indivíduo acertar o item quando não sabe (escolha aleatória da resposta) e o outro corresponde a **nossa opinião** sobre o nível de habilidade que o indivíduo tem para esta categoria.

Assim, o tratamento dado à habilidade não é dicotômico, isto é, não partimos da base de que o indivíduo possui ou não habilidade, mas admitimos que ele tem um certo **grau de habilidade**, que representaremos por uma proporção entre 0 (ausência total de habilidade, devido a qual o indivíduo optaria, fatalmente, pela escolha aleatória da resposta) e 1 (habilidade plena, graças a qual o indivíduo acertaria, sempre, a resposta correta).

Em relação ao conceito do Domínio acima citado, estabeleceremos que existe um conjunto finito de possíveis perguntas que poderão ser utilizadas para confeccionar um teste. Este conjunto de perguntas pode ser particionado em subconjuntos (subdomínios ou categorias) com uma ou mais características comuns, como por exemplo, perguntas mais ou menos difíceis ou perguntas sobre determinado assunto ou perguntas cujos conteúdos foram bem trabalhados ou não em sala de aula, etc.. O importante é que cada subdomínio seja mutuamente exclusivo em relação aos outros.

Podemos representar, esquematicamente, para uma certa categoria ***k***, como seriam divididas as respostas nesse subdomínio, na tabela 3:

categoria ***k***	CERTAS	ERRADAS	TOTAL
SABE	Φ	0	Φ
NÃO SABE	$\Theta - \Phi$	$N - \Theta$	$N - \Phi$
TOTAL	Θ	$N - \Theta$	N

TABELA 3

Lembremos que aqui, Φ representa o número de questões no subdomínio que o indivíduo realmente sabe e que responderia corretamente com probabilidade 1, caso alguma pergunta deste subdomínio seja escolhida para compor o teste.

O valor 0 (zero) indica que se o indivíduo souber a resposta, ele não errará.

O valor $\Theta - \Phi$ representa o número de respostas que o indivíduo não sabe mas que poderia responder corretamente, quando escolhe aleatoriamente uma alternativa do item.

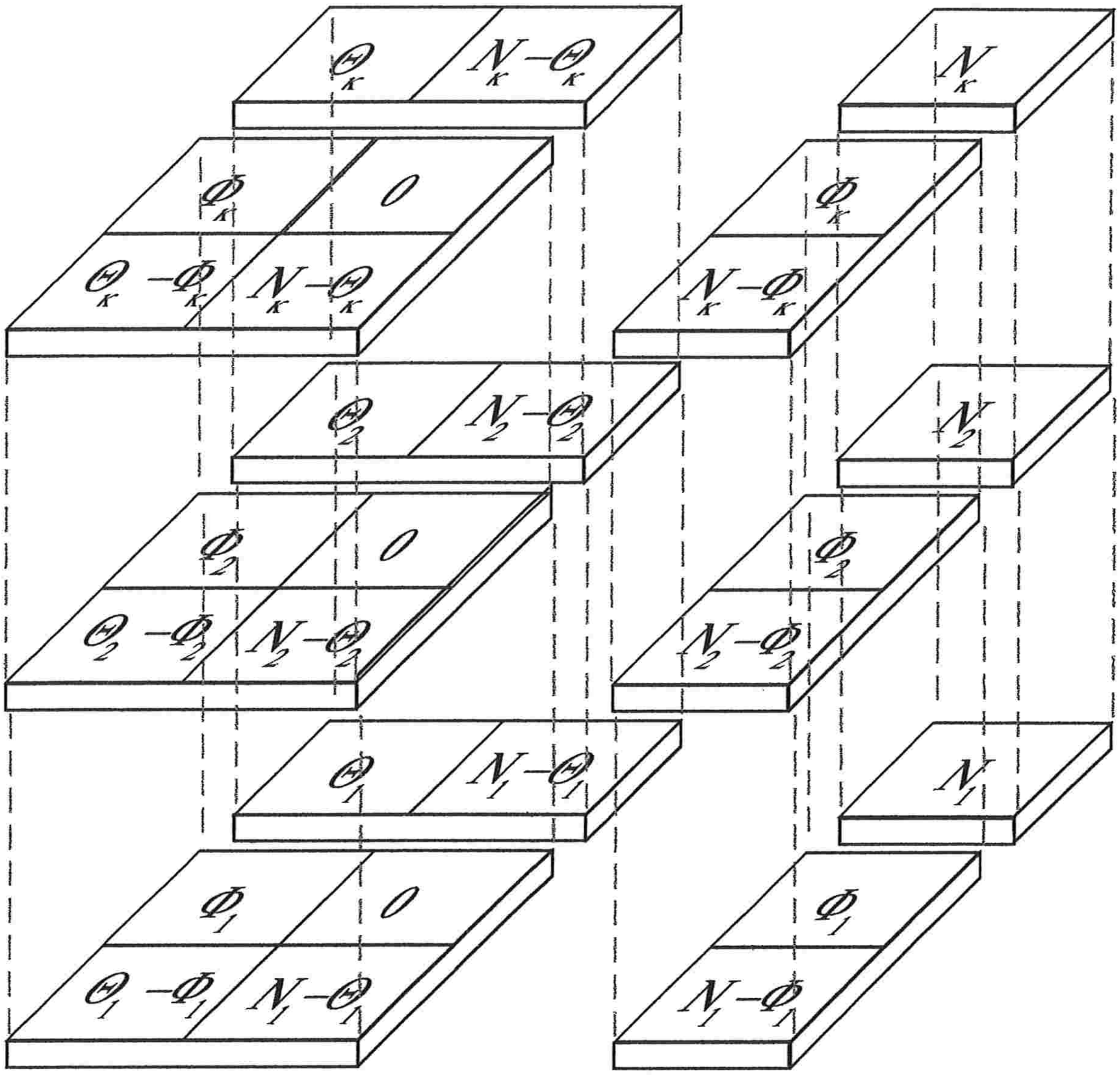
Finalmente, o valor $N - \Theta$ está vinculado ao número de respostas que o indivíduo não sabe e que também não acertaria na escolha casual.

Φ é, portanto, o parâmetro de interesse, cuja estimativa é objeto do nosso estudo.

Θ representa o número de questões que o indivíduo responderia corretamente.

N é o tamanho do subdomínio.

Podemos imaginar que o domínio é uma espécie de prédio no qual, em cada andar, estão as perguntas com os mesmos valores de probabilidade de acerto casual e habilidade. Cada pavimento deste prédio é um subdomínio e pode ser descrito com uma tabela similar à mostrada acima.



Para confeccionar o teste, são feitas retiradas sem reposição de perguntas de cada subdomínio. Se a partição foi feita, por exemplo, em ***k*** subdomínios, um teste de tamanho **n**, estará composto pelo vetor $\{n_1, n_2, ...,n_k\}$; com

$$n = n_1 + n_2 + ...+n_k.$$

$n_{\cdot}$ é a amostra extraída da categoria ..

A escolha dos valores n_1, n_2, ...,n_k faz parte da construção do teste e é, portanto, arbitrária.

O seguinte quadro (tabela 4) mostra a distribuição das respostas do indivíduo no teste, para a categoria j.

Categoria j	CERTAS	ERRADAS
SABE	x	0
NÃO SABE		n - x
TOTAL	x	n - x

TABELA 4

O número de questões que o indivíduo sabe não é observável, nem o número de questões que ele não sabe quando as acerta, mas podemos representar as suas freqüências como um vetor $\mathbf{x} = (x_1, x_2)$. Com esta modificação a tabela 4 teria este aspecto:

Categoria j	CERTAS	ERRADAS
SABE	x_1	0
NÃO SABE	x_2	n - x
TOTAL	x	n - x

TABELA 5

O que nós observamos no teste são as quantidades x e n - x. Naturalmente, também saberemos quantas questões da categoria 1, 2,.., j,..,k, foram respondidas corretamente. Assim, uma tabela análoga à de cima poderá ser feita para cada categoria.

Então o parâmetro de interesse passa a ser $\Phi = \{ \Phi_1, \Phi_2, \ldots, \Phi_k \}$ e a função de verossimilhança associada ao fato de acertar x dentre n questões, tem distribuição hipergeométrica com parâmetros (Θ, N, n):

$$L = \frac{(\Phi; x_1)\,(\Theta - \Phi; x_2)}{(\Theta\,; x)} \quad \frac{(\Theta; x)\,(N - \Theta; n - x)}{(N\,; n)} \qquad (2.3)$$

Considerando o vetor de parâmetros como $(\Theta, \Phi, N - \Theta)$, a priori pode ser escrita da seguinte forma equivalente:

$$\begin{gathered} \Theta \sim BB\,(N;\, \alpha_1 + \alpha_2,\, \alpha_4) \\ \Phi \mid \Theta \sim BB\,(\Theta;\, \alpha_1,\, \alpha_2) \\ (\Theta - \Phi) \mid \Theta \sim BB\,(\Theta;\, \alpha_2,\, \alpha_1) \end{gathered} \tag{2.4}$$

Observamos também que $\Phi \mid (\Theta, x) \sim \Phi \mid \Theta$ pois x não traz qualquer informação sobre Φ.

De (2.3) e (2.4) obtemos: $\Theta - x \mid x \sim BB\,(N - n;\, \alpha_1 + \alpha_2 + x,\, \alpha_4 + n - x)$ (2.5)

O nosso propósito é determinar agora a distribuição a posteriori de $\Phi \mid x$:

$$P\{\Phi = k \mid x\} = \sum_{m=x}^{N-n+x} P\{\Theta = m \mid x\} \,.\, P\{\Phi = k \mid (x,\, \Theta = m)\} =$$

$$\sum_{m-x=0}^{N-n} P\{\Theta - x = m - x \mid x\} \,.\, P\{\Phi = k \mid (\Theta = m)\}$$

Agora, todas as distribuições envolvidas nessa composição são familiares (ver 2.4 e 2.5).

$$\sum_{m-x=0}^{N-n} (N-n, m-x) \frac{\Gamma\,(\alpha_1 + \alpha_2 + m)\,\Gamma\,(\alpha_4 + N - m) \quad \Gamma\,(\alpha_1 + \alpha_2 + \alpha_4 + n)}{\Gamma\,(\alpha_1 + \alpha_2 + \alpha_4 + N) \quad \Gamma\,(\alpha_1 + \alpha_2 + x)\,\Gamma\,(\alpha_4 + n - x)}$$

$$.\,(m, k) \quad \frac{\Gamma\,(\alpha_1 + k)\,\Gamma\,(\alpha_2 + m - k)}{\Gamma\,(\alpha_1 + \alpha_2 + m)} \quad \frac{\Gamma\,(\alpha_1 + \alpha_2)}{\Gamma\,(\alpha_1)\,\Gamma\,(\alpha_2)}$$

2.6 Estimação do parâmetro de interesse Φ e determinação da medida de informação por meio da variância do estimador de Bayes.

$$\hat{\Phi} = E\{\Phi \mid x\} = E\{E\{\Phi \mid \Theta, x\} \mid x\} = E\{E\{\Phi \mid \Theta\} \mid x\} =$$
$$E\{\Theta\ [\alpha_1 / (\alpha_1+\alpha_2)] \mid x\}$$

$$\hat{\Phi} = E\{\Phi \mid x\} = \frac{\alpha_1}{(\alpha_1+\alpha_2)} [(N-n) \frac{(\alpha_1+\alpha_2+x)}{(\alpha+n)} + x] \qquad (2.6)$$

Com $\alpha = \alpha_1+\alpha_2+\alpha_4$

$$Var\{\Phi \mid x\} = E\{ Var\{\Phi \mid \Theta, x\} \mid x\} + Var\{E\{\Phi \mid \Theta, x\} \mid x\} =$$

$$= E\{Var\{\Phi \mid \Theta\} \mid x\} + Var\{E\{\Phi \mid \Theta\} \mid x\} =$$

$$= E\{\Theta \frac{\alpha_1}{(\alpha_1+\alpha_2)} \frac{\alpha_2}{(\alpha_1+\alpha_2)} \frac{(\alpha_1+\alpha_2+\Theta)}{(\alpha_1+\alpha_2+1)} \mid x\} + Var\{\Theta \frac{\alpha_1}{(\alpha_1+\alpha_2)} \mid x\} =$$

$$= \frac{\alpha_1}{(\alpha_1+\alpha_2)} \frac{\alpha_2}{(\alpha_1+\alpha_2)} \frac{1}{(\alpha_1+\alpha_2+1)} E\{(\alpha_1+\alpha_2+\Theta)\Theta \mid x\} + \frac{\alpha_1^2}{(\alpha_1+\alpha_2)^2} Var\{\Theta \mid x\}$$

Ainda,

$$E\{(\alpha_1+\alpha_2+\Theta)\Theta \mid x\} = (\alpha_1+\alpha_2) E\{\Theta \mid x\} + E\{\Theta^2 \mid x\} =$$

$$= (\alpha_1+\alpha_2) E\{\Theta \mid x\} + Var\{\Theta \mid x\} + E^2\{\Theta \mid x\}$$

e

$$Var\{\Theta \mid x\} = Var\{\Theta - x \mid x\} = (N-n) \frac{\alpha_1+\alpha_2+x}{\alpha+n} (\alpha_4+n-x) \frac{\alpha+N}{\alpha+n+1}$$

De tal forma que:

$$\boldsymbol{Var\,\{\Phi \mid x\}} = \frac{\alpha_1\,\alpha_2}{(\alpha_1+\alpha_2)^2}\;\frac{1}{\alpha_1+\alpha_2+1}\;\{(\alpha_1+\alpha_2)[(N-n)\;\frac{\alpha_1+\alpha_2+x}{\alpha+n}\;+x]\;+$$

$$+\,(N-n)\,\frac{(\alpha_1+\alpha_2+x)\;(\alpha+N)}{(\alpha+n)\;(\alpha+n+1)}\;(\alpha_4+n-x)\;+\;\{(N-n)\;\frac{\alpha_1+\alpha_2+x}{\alpha+n}\;+x\,\}^{2}\}+$$

$$+\;\frac{\alpha_1^{2}}{(\alpha_1+\alpha_2)^2}\;\frac{(\alpha_1+\alpha_2+x)\;(\alpha+N)}{(\alpha+n)(\alpha+n+1)}\;(N-n)\;(\alpha_4+n-x)$$

A função de incerteza (F.I.) é dada por:

$$F.I. = Var\,(\hat{\Phi}) = Var(\Phi) - E\,\{Var\{\Phi \mid X\}\} = Var\,\{E\{\Phi \mid X\}\}$$

$$\boldsymbol{Var\,\{E\{\Phi \mid X\}\}} = Var\,\{\;\frac{\alpha_1}{\alpha_1+\alpha_2}\;[(N-n)\;\frac{\alpha_1+\alpha_2+x}{\alpha+n}+x]\,\} =$$

$$= \frac{\alpha_1^{2}}{(\alpha_1+\alpha_2)^2}\;Var\,\{\,[\;\frac{(N-n)(\alpha_1+\alpha_2)}{(\alpha+n)} + \frac{(N-n)\,x}{(\alpha+n)} + x\,]\,\} =$$

$$= \frac{\alpha_1^{2}}{(\alpha_1+\alpha_2)^2}\,Var\,\{\,[\;\frac{N-n}{\alpha+n}\;+1\,]\,x\,\} =$$

$$= \frac{\alpha_1^{2}}{(\alpha_1+\alpha_2)^2}\,[\,\frac{N-n}{\alpha+n}\;+1\,]^{2}\,Var\,\{\,x\,\}$$

Agora,

$$x \sim BB(n;\ \alpha_1+\alpha_2,\ \alpha_4) \qquad\qquad \text{ver } (2.2)(iii)$$

Portanto,

$$Var\,\{E\{\Phi \mid x\}\} = n\ \frac{\alpha_1^2\,\alpha_4\,(\alpha+n)}{(\alpha_1+\alpha_2)\alpha^2\,(\alpha+1)}\ \left[\frac{(N-n)}{(\alpha+n)} + 1\right]^2 \qquad (2.7)$$

Definimos o ganho relativo de informação como:

$I = 100 \times F.\ I.\ /\ Var\,\{\Phi\}$

Vamos determinar agora a Var { Φ }.

Lembremos que $\Phi \sim BB\ (N;\ \alpha_1,\ \alpha_2+\alpha_4)$

$$Var\,\{\Phi\} = N\ \frac{\alpha_1}{\alpha}\ \frac{(\alpha_2+\alpha_4)}{\alpha}\ \frac{(\alpha+N)}{(\alpha+1)} \qquad (2.8)$$

Então, o ganho de informação é determinado por:

$$I = \frac{n}{N}\ \frac{\alpha_1\,\alpha_4\,(\alpha+n)(\alpha+N)}{(\alpha_1+\alpha_2)(\alpha+1)^2\,(\alpha_2+\alpha_4)}\ \left[\frac{N-n}{\alpha+n} + 1\right]^2 \qquad (2.9)$$

2.7 Ilustração

A solução computacional para a aplicação dos resultados acima apresentados foi conseguida com o auxílio da planilha eletrônica Excel(R) da Microsoft Inc..

Com pequenas adaptações, mostramos qual seria a visualização dos resultados, usando a planilha desenvolvida.

Quando um teste múltipla-escolha é aplicado a um indivíduo, esperamos que os acertos em algumas das questões sejam devidos ao acaso e não ao grau de conhecimento ou habilidade que ele possui.

Por essa razão, seria interessante poder estimar o real valor das questões que ele conhece no teste e no universo do qual as perguntas são tiradas.

Se considerarmos este universo ou DOMÍNIO pequeno em relação ao número de perguntas do teste, será apropriado que representemos nosso modelo de escolha aleatória das perguntas como hipergeométrico.

Para uma certa categoria, a distribuição de freqüências certas ou erradas que o indivíduo sabe ou não, pode ser representada pela seguinte tabela:

	CERTAS	ERRADAS	TOTAL
SABE	Φ	0	Φ
NÃO SABE	$\Theta - \Phi$	$N - \Theta$	$N - \Phi$
TOTAL	Θ	$N - \Theta$	**N**

Φ *é o parâmetro de interesse*

Uma vez aplicado o teste de tamanho n, o resultado do indivíduo pode ser representado da seguinte forma:

	CERTAS	ERRADAS	
SABE	x	0	
NÃO SABE		n – x	
TOTAL	x	n – x	**n**

Utilizaremos a informação de que dispomos, o número de respostas certas x, para estimar o número de respostas que ele conhece tanto no Teste quando no Domínio.

Por meio de um exemplo, veremos os principais resultados deste estudo.

A probabilidade de acerto casual é, normalmente, o inverso de número de alternativas do teste. Nada impede que você considere valores maiores ou menores para esta probabilidade.

O "grau de conhecimento do indivíduo" é a sua avaliação subjetiva de quanto ele sabe para cada categoria, por exemplo, se voce considerar que ele sabe a metade do conteúdo dessa categoria, então o grau de conhecimento é 0,5.

Nesta ilustração, trabalharemos com um teste composto por 40 perguntas tiradas, sem reposição, de um domínio de 100 questões, divididas em 3 categorias de dificuldade, com probabilidades de acerto casual iguais a 0,2; 0,25 e 0,3 respectivamente, e habilidades iguais a 0,9; 0,7 e 0,5 respectivamente.

Probabilidade de acerto casual

Categoria 1	0,20
Categoria 2	0,25
Categoria 3	0,30

Grau de conhecimento do indivíduo

Categoria 1	0,90
Categoria 2	0,70
Categoria 3	0,50

Para esta ilustração, consideraremos os seguintes tamanhos de categoria (no domínio): 50, 30 e 20 questões, respectivamente, por categoria.

Qual o tamanho da categoria 1?	50
Qual o tamanho da categoria 2?	30
Qual o tamanho da categoria 3?	20

DOMÍNIO	**100**

E de cada categoria serão escolhidas, aleatoriamente, 20, 12 e 8 perguntas, respectivamente, por categoria.

Quantos itens da categoria 1 há no Teste?	20
Quantos itens da categoria 2 há no Teste?	12
Quantos itens da categoria 3 há no Teste?	8
TESTE	**40**

Uma vez aplicado o teste, suponhamos que o indíviduo acertou 12, 6 e 3 perguntas em cada categoria, respectivamente.

Da categoria 1, quantos itens o indivíduo acertou?	12
Da categoria 2, quantos itens o indivíduo acertou?	6
Da categoria 3, quantos itens o indivíduo acertou?	3

PONTUAÇÃO BRUTA	**21**

Estabeleceremos agora um valor para Alfa priori. Quanto maior for o valor de alfa, menor será a variância da distribuição (que é uma Dirichlet Multinomial). Um valor neutro para Alfa priori é 2 (distribuição uniforme).

Alfa Priori	2

A seguir apresentamos os resultados para os alfas de cada célula da tabela de dupla entrada mostrada acima.

Observe que alfa3 é zero, pois o valor associado a esta célula é zero.

Parâmetros para a Categoria 1*	
alfa1	1,80
alfa2	0,04
alfa4	0,16

Parâmetros para a Categoria 2	
alfa1	1,40
alfa2	0,15
alfa4	0,45

Parâmetros para a Categoria 3	
alfa1	1,00
alfa2	0,30
alfa4	0,70

* alfa1 = Grau de habilidade x alfa
alfa2 = Probabilidade de acerto casual * (alfa - alfa1)
alfa4 = alfa - (alfa1 + alfa2)

A partir desses dados, determinaremos a esperança e variância do número de itens que o indivíduo sabe no teste, dado o número de itens que ele acertou.

Espera-se que o individuo saiba 17,64 perguntas das 40 que compõem o teste. A esperança e variância do número de itens que o indivíduo sabe no teste dado o número de acertos são obtidas por meio de (2.2) (iii)

Esperança do nº de itens que sabe no teste dado o nº de acertos	17,64
Variância do nº de itens que sabe no teste dado o nº de acertos	8,05

Naturalmente, o nosso objetivo é estimar o número de questões que o indivíduo sabe no Domínio, dado o número de acertos no teste. Para determinar este valor, utilizaremos o estimador de Bayes.

Assim, estimamos que conhece 50,67 questões do total de 100 que há no Domínio. Ver (2.6).

Estimador de Bayes do nº de itens que o indivíduo conhece no Domínio

50,67

A seguinte tabela, traz informações adicionais, sobre o comportamento do estimador de Bayes de Φ para cada categoria, dependendo do número de respostas que o indivíduo acerta no teste.

Depois você poderá observar o gráfico da distribuição de probabilidades do número de itens que o indivíduo sabe no teste, dado o número de acertos que conseguiu.

Categoria 1

Nº de acertos	$\hat{\Phi}$
0	2,45
1	4,77
2	7,08
3	9,39
4	11,70
5	14,02
6	16,33
7	18,64
8	20,95
9	23,26
10	25,58
11	27,89
12	30,20
13	32,51
14	34,83
15	37,14
16	39,45
17	41,76
18	44,08
19	46,39
20	48,70

Categoria 2

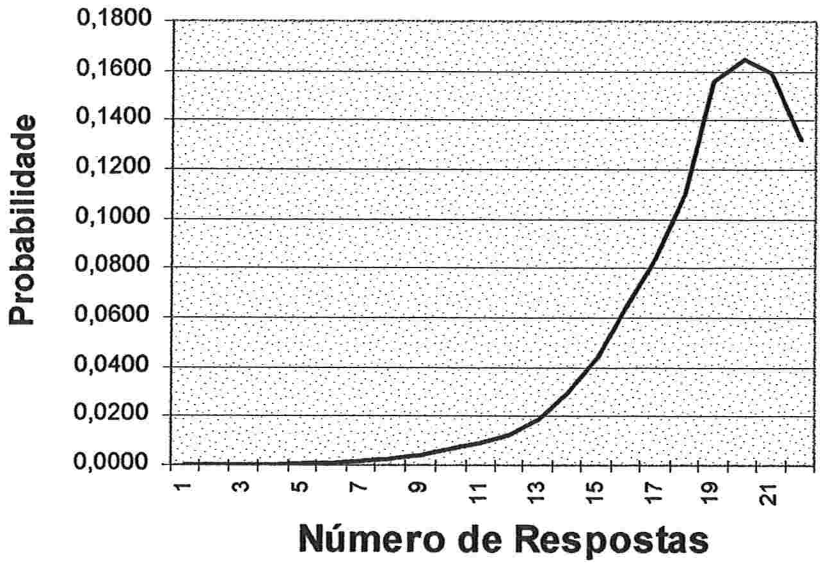
Nº de acertos	$\hat{\Phi}$
0	1,80
1	3,86
2	5,93
3	7,99
4	10,06
5	12,12
6	14,19
7	16,25
8	18,32
9	20,38
10	22,45
11	24,51
12	26,57

Categoria 3

Nº de acertos	$\hat{\Phi}$
0	1,20
1	2,89
2	4,58
3	6,28
4	7,97
5	9,66
6	11,35
7	13,05
8	14,74

O estimador de Bayes do número de itens que o indivíduo conhece no Domínio foi obtido somando os valores correspondentes de Φ para o número de acertos que o respondente obteve no teste: 30,20 + 14,19 + 6,28 = 50,67

Gráfico da distribuição de probabilidades do número de itens que o indivíduo sabe no teste, dado o número de itens corretos.



Agora apresentamos a distribuição de probabilidade do número de questões que o indivíduo sabe no teste, dado o número de respostas corretas.

Logo depois, você terá informações importantes sobre o parâmetro de interesse.

Respostas que sabe	*Probabilidade*	*Acumulado*	*Acum. Inverso*
0	0,0000	0,0000	1,0000
1	0,0000	0,0000	1,0000
2	0,0001	0,0001	1,0000
3	0,0002	0,0004	0,9999
4	0,0005	0,0009	0,9996
5	0,0009	0,0018	0,9991
6	0,0017	0,0035	0,9982
7	0,0028	0,0063	0,9965
8	0,0044	0,0107	0,9937
9	0,0065	0,0171	0,9893
10	0,0090	0,0261	0,9829
11	0,0123	0,0384	0,9739
12	0,0188	0,0572	0,9616
13	0,0289	0,0861	0,9428
14	0,0435	0,1295	0,9139
15	0,0644	0,1939	0,8705
16	0,0848	0,2786	0,8061
17	0,1104	0,3890	0,7214
18	0,1554	0,5444	0,6110
19	0,1643	0,7086	0,4556
20	0,1588	0,8674	0,2914
21	0,1326	1,0000	0,1326

O objetivo principal deste estudo é o de obter informações sobre o parâmetro Φ e medir qual é o ganho de informação ao considerar a porção não observável do teste, isto é, o número de itens que o indivíduo realmente sabe.

Definimos a função de incerteza como:

$$\boxed{\text{F.I.} = \text{Função de incerteza} = \text{Variância do estimador de } \Phi}$$

E o ganho relativo de informação como:

$$\boxed{\text{Ganho relativo de informação} = 100 \times \text{F.I.} \times Var^{-1}(\Phi)}$$

Veja os resultados:

Esperança de Φ dado o número de acertos no Teste	50,67	*Ver (2.6)*
Variância de Φ dado o número de acertos no Teste	50,12	
Variância do estimador de Φ para a cat. 1 (F. I. 1)	57,7138	*Ver (2.7)*
Variância do estimador de Φ para a cat. 2 (F. I. 2)	41,6206	
Variância do estimador de Φ para a cat. 3 (F. I. 3)	17,3744	
Variância de Φ para a categoria 1	78,0000	*Ver(2.8)*
Variância de Φ para a categoria 2	67,2000	
Variância de Φ para a categoria 3	36,6667	
Ganho relativo de Informação para a categoria 1	73,99%	*Ver (2.9)*
Ganho relativo de Informação para a categoria 2	61,94%	
Ganho relativo de Informação para a categoria 3	47,38%	

Você poderá observar agora como o ganho relativo de informação varia de acordo com as mudanças no tamanho do Domínio, no tamanho do teste, a probabilidade de acerto casual, a habilidade e o valor de alfa.

Você pode observar o que acontece quando, por exemplo, o valor de alfa é aumentado, mantendo a probabilidade de acerto casual a habilidade e os tamanhos do domínio e teste constantes.

Quando alfa aumenta, a variância, em geral diminui. Mas, a variância do estimador decresce mais rapidamente em relação à variância do parâmetro, o que significa uma diminuição do ganho relativo de informação. Como comentam Basu & Pereira (1981) quanto maior a variância do estimador, maior é o montante de informação no experimento.

Vejamos um exemplo: mantendo a probabilidade de acerto casual igual a 0,2, a habilidade em 0,7 , o tamanho do teste em 20 e o tamanho do domínio em 50 observemos que acontece quando alfa vale: 2, 5, 10, 20, 50 e 100.

Alfa	*2*	*5*	*10*	*20*	*50*	*100*
Ganho	*69,67%*	*64,84%*	*58,95%*	*51,58%*	*42,11%*	*36,84%*

O ganho estabiliza-se em 29,47%, quando alfa cresce indefinidamente.

Se agora decidíssemos fazer a probabilidade variar, mantendo alfa, a habilidade e os tamanhos do teste e do domínio constantes, observaremos quequanto maior a probabilidade de acerto casual, menor será o ganho relativo de informação. Isto é natural, pois uma maior probabilidade de acerto casual nos fará perder informação sobre o real grau de conhecimento do indivíduo.

Vejamos um exemplo: mantendo alfa constante é igual a 2, a habilidade em 0,7 e o tamanho do teste em 20 perguntas, observemos *como o ganho relativo de informação se modifica para os seguintes valores de probabilidade de acerto casual: 0,001 (quase zero!); 0,1; 0,2; 0,25; 0,33; 0,5.*

Prob.	*0,001*	*0,1*	*0,2*	*0,25*	*0,33*	*0,5*
Ganho	*94,41%*	*81,59%*	*69,67%*	*64,05%*	*55,50%*	*38,93%*

Vejamos que acontece quando o tamanho do teste é modificado. Para valores muito baixos de perguntas no teste o ganho de informação é pequeno. Isto se explica pelo fato de que a probabilidade de acerto casual tem peso maior quando o teste tem poucas perguntas. Isto significa que se eu fizer mais perguntas, terei mais chances de forçar o indivíduo a mostrar conhecimentos. Mas, o ganho de informação se estabiliza em um certo patamar (em torno de 70%, para 21 perguntas), a partir do qual qualquer aumento no tamanho do teste não representa ganho relativo de informação. Exemplo: para uma probabilidade de acerto casual igual a 0,2 (5 alternativas), um alfa de 2, a habilidade igual a 0,7 , o domínio igual a 50 e valores de tamanho de teste diversos: 2, 5, 10, 30, 40, 50.

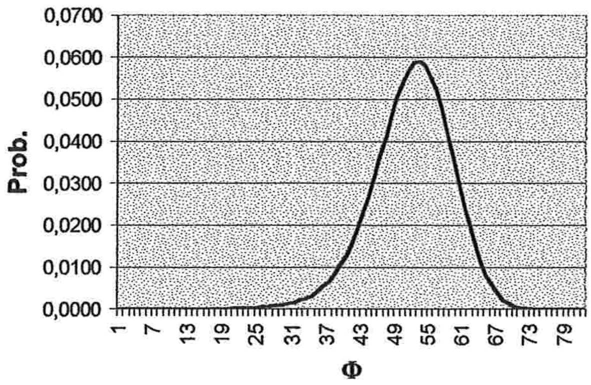
n	*2*	*5*	*10*	*30*	*40*	*50*
Ganho	*38,32%*	*54,74%*	*63,86%*	*71,84%*	*72,98%*	*73,68%*

Por último, vejamos que acontece quando o tamanho do Domínio cresce:

N	*20*	*100*	*200*	*500*	*1000*	*Infinito*
Ganho	*73,68%*	*68,33%*	*67,66%*	*67,25%*	*67,12%*	*66,99%*

Observemos que o ganho relativo de informação decresce até estabilizar-se em 66,99%, que corresponde ao valor obtido quando o modelo multinomial é considerado. Como esperávamos, os modelos multinomial e hipergeométrico coincidem no limite, quando o tamanho *do domínio cresce indefinidamente.*

Graficamente, mostraremos agora o comportamento da distribuição de probabilidades do números de itens que o indivíduo conhece no Domínio, dado o número de respostas corretas no Teste. O gráfico é baseado nos dados iniciais do problema.



Para os nossos dados iniciais a tabela da distribuição de probabilidades do número de itens que o indivíduo sabe no Domínio, dado o número de itens corretos no teste é a seguinte:

Respostas que sabe	*Probabilidade*	*Acumulado*	*Acum. Inverso*
0	0,0000	0,0000	1,0000
1	0,0000	0,000	1,0000
2	0,0000	0,0000	1,0000
3	0,0000	0,0000	1,0000
4	0,0000	0,0000	1,0000
5	0,0000	0,0000	1,0000
6	0,0000	0,0000	1,0000
7	0,0000	0,0000	1,0000
8	0,0000	0,0000	1,0000
9	0,0000	0,0000	1,0000
10	0,0000	0,0000	1,0000
11	0,0000	0,0000	1,0000
12	0,0000	0,0001	1,0000

Respostas que sabe	*Probabilidade*	*Acumulado*	*Acum. Inverso*
13	0,0000	0,0001	0,9999
14	0,0000	0,0001	0,9999
15	0,0000	0,0002	0,9999
16	0,0001	0,0002	0,9998
17	0,0001	0,0003	0,9998
18	0,0001	0,0004	0,9997
19	0,0001	0,0006	0,9996
20	0,0002	0,0008	0,9994
21	0,0002	0,0010	0,9992
22	0,0003	0,0013	0,9990
23	0,0004	0,0016	0,9987
24	0,0004	0,0021	0,9984
25	0,0005	0,0026	0,9979
26	0,0007	0,0033	0,9974
27	0,0008	0,0042	0,9967
28	0,0010	0,0052	0,9958
29	0,0013	0,0065	0,9948
30	0,0016	0,0081	0,9935
31	0,0020	0,0102	0,9919
32	0,0026	0,0127	0,9898
33	0,0033	0,0160	0,9873
34	0,0041	0,0201	0,9840
35	0,0052	0,0254	0,9799
36	0,0066	0,0319	0,9746
37	0,0082	0,0402	0,9681
38	0,0103	0,0504	0,9598
39	0,0127	0,0632	0,9496
40	0,0156	0,0788	0,9368
41	0,0190	0,0977	0,9212
42	0,0228	0,1205	0,9023
43	0,0270	0,1476	0,8795
44	0,0317	0,1792	0,8524
45	0,0365	0,2157	0,8208
46	0,0414	0,2571	0,7843
47	0,0461	0,3033	0,7429
48	0,0505	0,3538	0,6967
49	0,0542	0,4079	0,6462
50	0,0569	0,4649	0,5921
51	0,0586	0,5235	0,5351
52	0,0590	0,5824	0,4765
53	0,0580	0,6405	0,4176
54	0,0557	0,6962	0,3595
55	0,0522	0,7484	0,3038
56	0,0477	0,7961	0,2516
57	0,0424	0,8385	0,2039
58	0,0367	0,8752	0,1615

Respostas que sabe	*Probabilidade*	*Acumulado*	*Acum. Inverso*
59	0,0309	0,9061	0,1248
60	0,0252	0,9313	0,0939
61	0,0199	0,9512	0,0687
62	0,0152	0,9664	0,0488
63	0,0113	0,9777	0,0336
64	0,0080	0,9857	0,0223
65	0,0055	0,9912	0,0143
66	0,0036	0,9948	0,0088
67	0,0023	0,9971	0,0052
68	0,0014	0,9984	0,0029
69	0,0008	0,9992	0,0016
70	0,0004	0,9996	0,0008
71	0,0002	0,9998	0,0004
72	0,0001	0,9999	0,0002
73	0,0000	1,0000	0,0001
74	0,0000	1,0000	0,0000
75	0,0000	1,0000	0,0000
76	0,0000	1,0000	0,0000
77	0,0000	1,0000	0,0000
78	0,0000	1,0000	0,0000
79	0,0000	1,0000	0,0000
80	0,0000	1,0000	0,0000
81	0,0000	1,0000	0,0000

Capítulo 3

Modelo Multinomial

Neste capítulo apresentaremos a abordagem do problema de estimar o verdadeiro nível de habilidade de um indíviduo, segundo a suposição de reposição.

3.1 Apresentação do problema

Neste modelo, estipularemos que o tamanho do domínio é suficientemente grande como para ser considerado infinito. Desta forma, para denotar as características do Domínio, iremos trabalhar com proporções, em vez de números absolutos. Estas proporções populacionais são mostradas na seguinte tabela 2 x 2:

categoria ***k***	CERTAS	ERRADAS	
SABE	p_1	0	
NÃO SABE	p_2	p_4	
TOTAL	$1 - p_4$	p_4	1

TABELA 6

Uma vez aplicado o teste, os resultados podem ser resumidos assim:

Categoria ***k***	CERTAS	ERRADAS	
SABE	x	0	
NÃO SABE		x_4	
TOTAL	$n - x_4$	x_4	n

TABELA 7

A verossimilhança é descrita assim:

$$L = (1 - p_4)^{n - x_4} \; p_4^{x_4} \tag{3.1}$$

Vamos admitir que

$$p = (1 - p_4, p_4) \sim B(\alpha_1 + \alpha_2, \alpha_4) \qquad \text{a priori} \qquad (3.2)$$

Então, a distribuição a posteriori será

$$B\,(\alpha_1 + \alpha_2 + n - x_4,\ \alpha_4 + x_4)$$

A seguinte reparametrização vai ser útil para introduzirmos uma solução Bayesiana para o caso das respostas com censura (Basu & Pereira, 1981):

$$q = \frac{p_1}{p_1 + p_2} = \frac{p_1}{1 - p_4}$$

$$q_1 = p_1 + p_2$$

com transformação reversa sendo:

$$p_1 = q.\, q_1$$
$$p_2 = q_1\,(1 - q)$$
$$p_4 = 1 - q_1$$

3.2 Importante resultado teórico e sua aplicação à resolução do problema

O seguinte resultado teórico para as distribuições Dirichlet é o caminho para a solução do problema. Seja m ε {2, ..., k-1) fixo.

Lema – Se $(p_1, p_2, ..., p_k) \sim D\,(\alpha_1, \alpha_2, ...,\alpha_k)$, então o seguinte conjunto de afirmações pode ser verificado:

$$y = \sum_{i=1}^{m} p_i \sim B\,(\sum_{i=1}^{m} \alpha_i,\ \sum_{i=m+1}^{k} \alpha_i) \qquad (i)$$

$$\frac{1}{y}(p_1, p_2, \ldots\ldots, p_m) \sim D(\alpha_1, \alpha_2, \ldots\ldots\ldots, \alpha_m) \qquad (ii)$$

e

$$y \perp\!\!\!\perp \frac{1}{y}(p_1, p_2, \ldots\ldots, p_m) \qquad (iii)$$

Prova – Para o nosso problema é suficiente limitarmos a demonstração para $y = p_1 + p_2$ e p_1 / y . Ver apêndice.

Suponha que, a priori, (3.2) é considerada. Pelo Lema, isto é equivalente a:

$$q \sim B(\alpha_1, \alpha_2) \quad , \quad q_1 \sim (1 - p_4) \sim B(\alpha_1 + \alpha_2, \alpha_4) \quad e \quad q \perp\!\!\!\perp q_1$$

A reparametrização modifica a verossimilhança (3.1) para:

$$L = q^{n - x_4} (1 - q)^{x_4} \qquad (3.4)$$

De (3.3) e (3.4), derivamos as distribuições a posteriori de $(p_4 | x_4)$ e $(q | x_4)$:

$$q \perp\!\!\!\perp p_4 | x_4, \qquad (3.5)(i)$$

$$p_4 | x_4 \sim B(\alpha_4 + x_4, \alpha_1 + \alpha_2 + n - x_4) \quad e \qquad (3.5)(ii)$$

$$q | x_4 \sim q \sim B(\alpha_1, \alpha_2) \qquad (3.5)(iii)$$

Pois x_4 é não informativo para q.

3.3 Estimação do parâmetro de interesse, determinação da Função de Incerteza e do Ganho de Informação.

Para estimar p_1, o parâmetro de interesse, lembremos que:

$$p_1 = q\,(1- p_4)$$

e que

$$q \perp\!\!\!\perp p_4 \mid x_4$$

$$E\{p_1 \mid x_4\} = E\{q\,(1-p_4) \mid x_4\} = E\{E\{q(1-p_4) \mid x_4,\, p_4\} \mid x_4\} =$$

$$= E\{(1-p_4)\, E\{q \mid x_4, p_4\} \mid x_4\} =$$

$$= E\{(1-p_4)\, E\{q \mid p_4\} \mid x_4\} = E\{(1-p_4)\, E\{q\} \mid x_4\} =$$

$$= \frac{\alpha_1}{\alpha_1 + \alpha_2} E\{(1-p_4) \mid x_4\} = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{\alpha_1 + \alpha_2 + n - x_4}{\alpha + n}$$

Então,

$$E\{p_1 \mid x_4\} = \frac{\alpha_1}{(\alpha_1 + \alpha_2)} \frac{(\alpha_1 + \alpha_2 + n - x_4)}{(\alpha + n)}$$

e

$$Var\{p_1 \mid x_4\} = Var\{q\,(1-p_4) \mid x_4\} =$$

$$= E\{Var\{q\,(1-p_4) \mid p_4, x_4\} \mid x_4\} + Var\{E\{q\,(1-p_4) \mid p_4, x_4\} \mid x_4\} =$$

$$= E\{(1-p_4)^2\, Var\{q \mid x_4\} \mid x_4\} + Var\{(1-p_4)\, E\{q \mid x_4\} \mid x_4\} =$$

$$= \frac{\alpha_1\ \alpha_2}{(\alpha_1 + \alpha_2)^2\ (\alpha_1 + \alpha_2 + 1)}\ E\{(1-p_4)^2 \mid x_4\} + \frac{\alpha_1^2}{(\alpha_1 + \alpha_2)^2} Var\{(1-p_4) \mid x_4\} =$$

$$= \frac{\alpha_1 \ \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)} [Var\{(1-p_4) \mid x_4\} + E^2\{(1-p_4)^2 \mid x_4\}] +$$

$$+ \frac{\alpha_1^2}{(\alpha_1 + \alpha_2)^2} Var\{(1-p_4) \mid x_4\} =$$

$$= \frac{\alpha_1 \ \alpha_2}{(\alpha_1 + \alpha_2)^2 \ (\alpha_1 + \alpha_2 + 1)} \left[\frac{(\alpha_1+\alpha_2+ n - x_4)\ (\alpha_4 + x_4)}{(\alpha + n + 1)\ (\alpha + n)^2} + \frac{(\alpha_1+\alpha_2+ n - x_4)^2}{(\alpha + n)^2} \right]$$

$$+ \frac{\alpha_1^2 \ (\alpha_1+\alpha_2+ n - x_4)\ (\alpha_4 + x_4)}{(\alpha_1 + \alpha_2)^2 \ (\alpha + n + 1)\ (\alpha + n)^2}$$

A função de incerteza (F.I.) é dada por:

$$\mathbf{F.\,I. = Var\,(\hat{p}_1) = Var\,(p_1) - E\,\{Var\,\{\,p_1 \mid x_4\}\} \ = Var\,\{E\,\{\,p_1 \mid x_4\}\}}$$

e

$$\mathbf{Var\,\{E\,\{\,p_1 \mid x_4\}\}} = Var \left\{ \frac{\alpha_1 (\ \alpha_1 + \alpha_2 + n\ )}{(\alpha_1 + \alpha_2)\,(\alpha + n)} - \frac{\alpha_1 \ x_4}{(\alpha_1 + \alpha_2)\,(\alpha + n)} \right\} =$$

$$= \frac{\alpha_1^2}{(\alpha_1 + \alpha_2)^2 (\alpha + n)^2} \ Var\{x_4\}$$

Mas como $\mathbf{p_4 \sim B\,(\alpha_4,\ \alpha_1 + \alpha_2)}$ e $\mathbf{x_4 \mid p_4 \sim Bin\,(n,\ p_4)}$

então

$x_4 \sim BB\,(\,n,\ \alpha_4,\ \alpha_1 + \alpha_2\,)$ ver (2.2) (iii)

Portanto,

$$\mathbf{Var\ \{E\ \{\ p_1\ |\ x_4\}\}} = \frac{n\ \ \alpha_1^2\ \ \alpha_4}{(\alpha_1 + \alpha_2)\ (\alpha\ + n)\ (\alpha\ + 1)\ \alpha^{\ 2}}$$

Como vimos no caso hipergeométrico, o ganho relativo de informação como:

$$\mathbf{Ganho = 100\ x\ F.I.\ /\ Var\ \{\ p_1\}}$$

Vamos determinar agora $Var\ \{p_1\}$.
Lembremos que $p_1 \sim B\ (\alpha_1,\ \alpha_2+\alpha_4)$

$$\mathbf{Var\ \{\ p_1\}} = \frac{\alpha_1\ (\alpha_2 + \alpha_4)}{\alpha^2\ \ (\alpha + 1)}$$

Portanto o ganho relativo de informação é igual a:

$$Ganho = 100\ x\ \frac{n\ .\ \alpha_1\ \alpha_4}{(\alpha_1 + \alpha_2)\ (\alpha_2+\alpha_4)\ (\alpha + n)}$$

3.4 Determinação da função densidade de probabilidade de $p_1|x_4$

Vamos determinar agora a densidade de probabilidade de $p_1|x_4$:

Lembremos que:

$p_4|x_4 \sim B\ (\alpha_4 + x_4,\ \alpha_1 + \alpha_2 + n - x_4)$

$q\ |\ x_4 \sim q\ \sim B(\alpha_1,\ \alpha_2)$

$$q = \frac{p_1}{p_1 + p_2} = \frac{p_1}{1 - p_4}$$

$$q \perp\!\!\!\perp p_4 \mid x_4$$

$$f(q, p_4 \mid x_4) \overset{\text{ind.}}{=} f(q \mid x_4) \,.\, f(p_4 \mid x_4) = f(q) \,.\, f(p_4 \mid x_4) =$$

$$= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)}\, q^{\alpha_1 - 1} (1-q)^{\alpha_2 - 1} \frac{\Gamma(\alpha + n)}{\Gamma(\alpha_1 + \alpha_2 + n - x_4)\Gamma(\alpha_4 + x_4)} \quad x$$

$$x \quad p_4^{\alpha_4 + x_4 - 1} (1-p_4)^{\alpha_1 + \alpha_2 + n - x_4 - 1}$$

Fazendo $\mathbf{C} = B^{-1}(\alpha_1, \alpha_2)\; B^{-1}(\alpha_4 + x_4, \alpha_1 + \alpha_2 + n - x_4)$

$$= \mathbf{C} \; \frac{p_1^{\alpha_1 - 1}}{(1-p_4)^{\alpha_1 - 1}} \; (1 - \frac{p_1}{1 - p_4})^{\alpha_2 - 1} \; p_4^{\alpha_4 + x_4 - 1} (1-p_4)^{\alpha_1 + \alpha_2 + n - x_4 - 1} =$$

$$= \mathbf{C} \quad p_1^{\alpha_1 - 1} \,[(1 - p_1) - p_4\,]^{\alpha_2 - 1} \quad p_4^{\alpha_4 + x_4 - 1} \; (1-p_4)^{n - x_4 + 1} =$$

$$= \mathbf{C} \quad p_1^{\alpha_1 - 1} (1 - p_1)^{\alpha_2 - 1} [1 - \frac{p_4}{(1 - p_1)}]^{\alpha_2 - 1} p_4^{\alpha_4 + x_4 - 1} (1-p_4)^{n - x_4 + 1} =$$

$$= f(p_1, p_4 \mid x_4)$$

Vamos desenvolver o binômio $[1 - \frac{p_4}{(1 - p_1)}]^{\alpha_2 - 1}$ em séries de potências:

Caso 1: $\alpha_2 < 1$

$$[1 - \frac{p_4}{(1 - p_1)}]^{\alpha_2 - 1} = 1 + (1 - \alpha_2)\frac{p_4}{(1 - p_1)} + \frac{(1 - \alpha_2)(2 - \alpha_2)}{2!} \frac{p_4^2}{(1 - p_1)^2} +$$

$$+ \frac{(1-\alpha_2)(2-\alpha_2)(3-\alpha_2)}{3!} \frac{p_4^3}{(1-p_1)^3} + \cdots\cdots\cdots\cdots$$

$$f(p_1, p_4 \mid x_4) = C\, p_1^{\alpha_1-1}(1-p_1)^{\alpha_2-1}\{1 + (1-\alpha_2)\frac{p_4}{(1-p_1)} +$$

$$+ \frac{(1-\alpha_2)(2-\alpha_2)}{2!} \frac{p_4^2}{(1-p_1)^2} \cdots\cdots \}\, p_4^{\alpha_4+x_4-1}(1-p_4)^{n-x_4+1} =$$

$$= C\, p_1^{\alpha_1-1}(1-p_1)^{\alpha_2-1}\{p_4^{\alpha_4+x_4-1}(1-p_4)^{n-x_4+1} + (1-\alpha_2)\frac{p_4^{\alpha_4+x_4}(1-p_4)^{n-x_4+1}}{(1-p_1)} +$$

$$+ \frac{(1-\alpha_2)(2-\alpha_2)\, p_4^{\alpha_4+x_4+1}(1-p_4)^{n-x_4+1}}{2!\,(1-p_1)^2} +$$

$$+ \frac{(1-\alpha_2)(2-\alpha_2)(3-\alpha_2)\, p_4^{\alpha_4+x_4+2}(1-p_4)^{n-x_4+1}}{3!\,(1-p_1)^3} \ldots\}$$

Integrando $f(p_1,p_4 \mid x_4)$ no intervalo $(0\,,\,1)$ em p_4, *termo a termo*:

$$I_1 = \int_0^1 p_4^{\alpha_4+x_4-1}(1-p_4)^{n-x_4+1}\, dp_4 = \frac{\Gamma(\alpha_4+x_4)\,\Gamma(n-x_4+2)}{\Gamma(\alpha_4+n+2)} = B(\alpha_4+x_4\,,\, n-x_4+2)$$

$$I_2 = \int_0^1 \frac{(1-\alpha_2)}{(1-p_1)}\, p_4^{\alpha_4+x_4}(1-p_4)^{n-x_4+1}\, dp_4 = \frac{(1-\alpha_2)}{(1-p_1)} B(\alpha_4+x_4+1,\, n-x_4+2)$$

$$I_3 = \int_0^1 \frac{(1-\alpha_2)(2-\alpha_2)}{2!\,(1-p_1)^2}\, p_4^{\alpha_4+x_4+1}(1-p_4)^{n-x_4+1} dp_4 = \frac{(1-\alpha_2)(2-\alpha_2)}{2!\,(1-p_1)^2} B(\alpha_4+x_4+2, n-x_4+2)$$

....

....

$$I_K = \frac{(1-\alpha_2)\,(2-\alpha_2)\ldots(k-1-\alpha_2)}{(k-1)!\,(1-p_1)^{k-1}}\, B(\alpha_4+x_4+k-1,\ n-x_4+2)$$

.....

.....

Portanto,

$$f(p_1 \mid x_4) = \frac{p_1^{\alpha_1-1}\,(1-p_1)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)B(\alpha_4+x_4,\alpha_1+\alpha_2+n-x_4)} \{B(\alpha_4+x_4, n-x_4+2) +$$

$$+ \frac{(1-\alpha_2)}{(1-p_1)} B(\alpha_4+x_4+1, n-x_4+2) + \frac{(1-\alpha_2)(2-\alpha_2)}{2!\,(1-p_1)^2} B(\alpha_4+x_4+2, n-x_4+2) + \ldots\ldots\}$$

Finalmente,

$$f(p_1|x_4) = \frac{p_1^{\alpha_1-1}\,(1-p_1)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)B(\alpha_4+x_4,\alpha_1+\alpha_2+n-x_4)} \sum_{k=1}^{\infty} \frac{(1-\alpha_2)\,(2-\alpha_2)\ldots(k-1-\alpha_2)}{(k-1)!\,(1-p_1)^{k-1}} B(\alpha_4+x_4+k-1, n-x_4+2)$$

Caso 2: $\alpha_2 > 1$

Operando analogamente ao Caso 1, chegamos à seguinte expressão:

$$f(p_1|x_4) = \frac{p_1^{\alpha_1-1}\,(1-p_1)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)B(\alpha_4+x_4,\alpha_1+\alpha_2+n-x_4)} \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(\alpha_2-1)(\alpha_2-2)\ldots(\alpha_2-k+1)}{(k-1)!\,(1-p_1)^{k-1}} B(\alpha_4+x_4+k-1, n-x_4+2)$$

3.5 Ilustração dos resultados deste capítulo e uma aplicação prática

Novamente, a solução computacional para a aplicação dos resultados acima apresentados foi obtida com o auxílio da planilha eletrônica Excel(R) da Microsoft(R).

Com pequenas adaptações, mostramos qual seria a visualização dos resultados, usando a planilha desenvolvida.

Quando o tamanho do Domínio puder ser considerado infinito em relação ao tamanho do teste, o modelo apropriado será o Multinomial. Isto é, trabalharemos com um modelo de escolha aleatória das perguntas com reposição.

Para explicitar as características do Domínio iremos trabalhar com proporções. Estas proporções populacionais podem ser mostradas na seguinte tabela:

	CERTAS	ERRADAS	TOTAL
SABE	p_1	0	p_1
NÃO SABE	p_2	p_4	$1 - p_1$
TOTAL	$1 - p_4$	p_4	1

p_1 *é o parâmetro de interesse*

Uma vez aplicado o teste, os resultados podem ser apresentados na seguinte tabela:

	CERTAS	ERRADAS	
SABE	$n - x_4$	**0**	
NÃO SABE		$\mathbf{x_4}$	
TOTAL	$n - x_4$	$\mathbf{x_4}$	**n**

Utilizaremos o número de respostas incorretas x_4 para estimar a proporção de respostas que ele conhece no Domínio.

Por meio de um exemplo, veremos os principais resultados deste estudo.

A probabilidade de acerto casual é, normalmente, o inverso do número de alternativas do teste. Porém, nada impede que você considere valores maiores ou menores para esta probabilidade, caso possua alguma informação adicional.

Aqui, consideraremos esta probabilidade de forma livre.

O grau de conhecimento do indivíduo é a sua avaliação subjetiva de quanto ele sabe para cada categoria, por exemplo, se voce considerar que ele sabe a metade do conteúdo dessa categoria, então o grau de conhecimento é 0,5.

Nesta ilustração, trabalharemos com um teste composto por 40 perguntas, divididas em 3 categorias de dificuldade, com probabilidades de acerto casual iguais a 0,2; 0,25 e 0,3 respectivamente, e habilidades iguais a 0,9; 0,7 e 0,5 respectivamente.

Probabilidade de acerto casual

Categoria 1	0,20
Categoria 2	0,25
Categoria 3	0,30

Grau de conhecimento do indivíduo

Categoria 1	0,90
Categoria 2	0,70
Categoria 3	0,50

Suponhamos que o teste contém, 20 perguntas da categoria 1, 12 perguntas da categoria 2 e 8 perguntas da categoria 3.

Quantos itens da categoria **1** há no Teste?	20
Quantos itens da categoria **2** há no Teste?	12
Quantos itens da categoria **3** há no Teste?	8
Tamanho do Teste	40

Uma vez aplicado o teste, constatamos que o indivíduo acertou, respectivamente, 12, 6 e 3 perguntas.

Da categoria **1**, quantos itens o indivíduo acertou?	12
Da categoria **2**, quantos itens o indivíduo acertou?	6
Da categoria **3**, quantos itens o indivíduo acertou?	3
Pontuação Bruta	21

Por último, estabeleceremos um valor para Alfa priori. Um valor neutro para Alfa priori é 2 (distribuição uniforme).

Alfa Priori	2

A continuação aparecem os alfas para cada célula.

Observe que alfa3 é zero, pois o valor associado a esta célula é zero.

Parâmetros para a **Categoria 1**

alfa1	1,800
alfa2	0,040
alfa4	0,160

Parâmetros para a **Categoria 2**

alfa1	1,400
alfa2	0,150
alfa4	0,450

Parâmetros para a **Categoria 3**

alfa1	1,000
alfa2	0,300
alfa4	0,700

A seguir obteremos algumas informações sobre p_I. Vamos medir também qual é o ganho de informação ao considerar a porção *não observável do teste, isto é, o número de itens que o indivíduo realmente sabe.*

Definimos a função de incerteza como:

F.I. = Função de incerteza = Variância do estimador de p_1

E o ganho relativo de informação como:

Ganho relativo de informação = 100 x F.I. x $Var^{-1}(p_1)$

Como vimos, a variância do estimador de p_1, é obtida por:

Variância do estimador de $p_1 = n\, \alpha_1^{2} \alpha_4\, (\alpha_1 + \alpha_2)^{-1} (\alpha + n)^{-1} (\alpha + 1)^{-1}\, \alpha^{-2}$

Entanto que a variância do parâmetro p_1 é obtida por:

Variância de $p_1 = \alpha_1\, (\alpha_2 + \alpha_4)\ \ \alpha^{-2}\ \ (\alpha + 1)^{-1}$

Portanto o ganho pode ser medido por:

Ganho = $100\ .\ n\ \alpha_1\, \alpha_4 (\alpha_1 + \alpha_2)^{-1}\ \ (\alpha_2 + \alpha_4)^{-1}\ (\alpha + n)^{-1}$

Veja os resultados, a partir dos dados acima estabelecidos:

Esperança de p_1 dado o nº de respostas incorretas no Teste, na categoria 1	0,6154
Variância de p_1 dado o nº de respostas incorretas no Teste, na categoria 1	0,0071
Esperança de p_1 dado o nº de respostas incorretas no Teste, na categoria 2	0,4871
Variância de p_1 dado o nº de respostas incorretas no Teste, na categoria 2	0,0454

Esperança de p_1 dado o nº de respostas incorretas no Teste, na categoria 3	0,3308
Variância de p_1 dado o nº de respostas incorretas no Teste, na categoria 3	0,0507

Variância do estimador de p_1 para a categoria 1 (Função de Incerteza 1)	0,0213
Variância do estimador de p_1 para a categoria 2 (Função de Incerteza 2)	0,0406
Variância do estimador de p_1 para a categoria 3 (Função de Incerteza 3)	0,0359

Variância de p_1 para a categoria 1	0,0300
Variância de p_1 para a categoria 2	0,0700
Variância de p_1 para a categoria 3	0,0833

Ganho relativo de Informação para a categoria 1	71,15%
Ganho relativo de Informação para a categoria 2	58,06%
Ganho relativo de Informação para a categoria 3	43,08%

Observaremos agora o que acontece com o ganho, quando a probabilidade de acerto casual é modificada.

Suponhamos que para a categoria 3 diminuímos a probabilidade de acerto casual de 0,3 para 0,2.

Probabilidade	Teste	Habilidade	Ganho	Alfa
0,20	8	0,50	53,33%	2

O ganho aumenta (de 43,08% para 53,33%). Isto é coerente, pois quanto maior for a probabilidade de acerto casual, o número de respostas conhecidas ficará cada vez mais disfarçado e será mais difícil obter informações sobre o parâmetro de interesse. Quando a probabilidade de acerto casual se aproxima de 1, o ganho diminui rapidamente para zero.

Vamos observar agora o que acontece quando, por exemplo, o valor de alfa é aumentado, mantendo a probabilidade de acerto casual, a habilidade e o tamanho do teste constantes. Quando alfa aumenta, as variâncias em geral, diminuem. Mas, a variância do estimador decresce mais rapidamente em relação à variância do parâmetro, o que significa uma diminuição do ganho relativo de informação. Como comentam Basu & Pereira (1981), quanto maior a variância do estimador, maior é o montante de informação no experimento.

Vejamos um exemplo: mantendo a probabilidade de acerto casual igual a 0,2, a habilidade em 0,7 e o tamanho do teste em 30, observemos que acontece quando alfa vale: 2, 5, 10, 20, 50 e 100.

Alfa	2	5	10	20	50	100
Ganho	69,08%	63,16%	55,26%	44,21%	27,63%	17,00%

Se agora decidíssemos fazer a probabilidade variar, mantendo o alfa e o tamanho do teste constantes, observaremos que quanto maior a probabilidade de acerto casual, menor será o ganho relativo de informação. Isto é natural, pois uma maior probabilidade de acerto casual nos fará perder informação sobre o real grau de conhecimento do indivíduo.

Vejamos um exemplo: mantendo alfa constante é igual a 2, a habilidade em 0,7 e o tamanho do teste em 30 perguntas, observemos como o ganho relativo de informação se modifica para os seguintes valores de probabilidade de acerto casual: 0,001; 0,1; 0,2; 0,25; 0,33; 0,5.

Probab.	0,001	0,1	0,2	0,25	0,33	0,5
Ganho	93,62%	80,91%	69,08%	63,51%	55,03%	38,60%

Vejamos que acontece quando o tamanho do teste é modificado. Para valores muito baixos de perguntas no teste o ganho de informação é pequeno. Isto se explica pelo fato de que a probabilidade de acerto casual tem peso maior quando o teste tem poucas perguntas. Isto significa que se eu fizer mais perguntas, terei mais chances de forçar o indivíduo a mostrar conhecimentos. Mas, o ganho de informação se estabiliza em um certo patamar, a partir do qual qualquer aumento no tamanho do teste não representa ganho relativo de informação.

Veja os exemplos: para uma probabilidade de acerto casual igual a 0,2 (5 alternativas), um alfa de 2, a habilidade igual a 0,7 e valores de tamanho de teste diversos: 2, 15, 30, 50, 500, 5000.

Teste	2	15	30	50	500	5000
Ganho	36,84%	65,02%	69,08%	70,85%	73,39%	73,65%

Vamos imaginar agora que você deseja confeccionar um teste com as seguintes características: a habilidade do indivíduo deve ser de pelo menos 75%, probabilidade de acerto casual igual a 0,2, alfa priori igual a 2, tamanho do teste igual a 30 e que "garanta" que o indivíduo sabe pelo menos 70% das perguntas no Domínio. Qual seria a nota de coorte deste teste?

n	Prob.	Habil.	Alfa	$E(p_i)$	Coorte
30	0,2	75%	2	70%	23

Para os dados acima, a nota de coorte é 23. Isto significa que o indivíduo deve acertar mais de 76% (23 em 30) do teste para considerar que, em média, ele conhece 70% do domínio.

Em muitas ocasiões, a nota de coorte é estabelecida de acordo com os resultados do conjunto de pessoas que participaram do teste.

No Exame Nacional Cursos, que é aplicado a alguns cursos das Universidades do Brasil, para avaliar a "saúde" da educação universitária, as notas de coorte foram determinadas, até o ano de 2000, por percentis da distribuição (P12, P30, P70, P88) para definir o "conceito" da Instituição. Para o ano de 2001 utilizar-se-á uma metodologia um pouco diversa: determinar a média e o desvio padrão das notas absolutas e dividir as categorias segundo elas se afastem da média $\pm$ 1 desvio padrão (Notas A e E), entre -0,5 e 0,5 desvios (Nota C) e entre -1 e -0,5 desvios e 0,5 e 1 desvios para as notas D e B, respectivamente.

No caso específico do Exame Nacional de Cursos (ENC) a determinação prévia das notas de coorte seria mais correto, acreditamos, pois este tipo de metodologia pode apresentar distorções muito importantes. Uma Instituição de Ensino Superior (IES) pode ter tirado uma nota absoluta muito ruim (como vem acontecendo atualmente) mas ser considerada A por ser uma das melhores (dentre as piores).

Inclusive pode acontecer a situação curiosa de um curso melhorar em forma absoluta, mas piorar em termos relativos, de um exame para o outro: um certo ano um curso obtém nota B tendo conseguido nota absoluta, digamos, 5 (na escala 0 a 10) e no ano seguinte ser C, com nota absoluta 6, caso as outras IES melhorem mais ainda. Qual seria a conclusão? A instituição melhorou ou piorou? Uma IES não faz o seu projeto de ensino de acordo com as outras Universidades mas de forma autônoma. Portanto, parece-nos falha esta forma de avaliação.

No Vestibular da Fuvest os resultados do teste tem importância relativa desde que a nota de coorte é determinada por um ele-

mento extrínseco ao teste: o número de vagas.

No caso do Exame Nacional de Cursos (ENC), cremos que nada justifica a adoção de um sistema de determinação de notas de coorte a posteriori.

Uma Instituição de Ensino Superior (IES) deve ser considerada A se for realmente ótima. Nos Exames Nacionais de Cursos anteriores, as notas de alguns cursos não superaram 30% de acertos em média! Mesmo assim, as instituições que conseguiram pouco acima de 30% de acertos obtiveram nota A.

É como ser um astro de futebol na Ilhas Fidji. Pode não significar muita coisa...

Atualmente, a prova do ENC divide-se em duas partes, a primeira é um teste com 40 perguntas de múltipla escolha com 5 alternativas por pergunta (probabilidade teórica de 0,2 de acerto casual). A segunda parte é discursiva e, portanto, a probabilidade de acerto casual é zero.

Poderíamos pensar em utilizar o nosso trabalho para atribuir conceitos às IES. Para fazer isto, devemos pensar em algumas características dos alunos, que sejam apropriadas ao perfil de formando que o Ministério de Educação -precursor do ENC- estima adequado. Por exemplo, um valor para a habilidade, que representa aqui o quanto o indivíduo aprendeu no seu curso, poderia ser fixado em 75%.

A Esperança de p_1 representa a capacidade que o indivíduo tem para responder as perguntas do domínio com a habilidade que ele possui e é o que vai determinar as diferentes notas de coorte para cada categoria (A, B, C, D ou E). Fixaremos um alfa neutro igual a 2.

Suponhamos, que decidimos que a IES obterá nota A se a *esperança de p_1 for maior ou igual a 85% (saber mais de 85% do domínio), B será quem souber entre 70 e 85% domínio, C quem souber entre 30 e 70% do domínio, D quem possuir conhecimentos entre 15 e 30% domínio e, finalmente, E quem estiver abaixo de 15% de respostas conhecidas no domínio. Para a parte múltipla-escolha usaremos 0,2 como probabilidade de acerto casual e para a parte discursiva esta probabilidade será zero. Utilizando a nossa "máquina de determinação de notas de coorte", teremos:*

	E	D	C	B	A
Múlt-Esc.	**< 6 questões**	**6 a 11**	**12 a 29**	**30 a 36**	**37 ou +**
*Discursiva**	**< 5 questões**	**5 a 11**	**12 a 27**	**28 a 34**	**35 ou +**

*A parte discursiva tem, na realidade, bem menos perguntas que a de Múltipla-Escolha. Consideramos também 40 perguntas para esta parte discursiva, apenas para comparar com as notas de coorte da parte de Múltipla-Escolha

Observe-se que a exigência de acertos para a prova discursiva é menor pois estamos retirando a aleatoriedade das respostas.

Acreditamos que a forma de determinação das notas de coorte para o ENC que propomos seja mais justa que a utilizada até hoje. Com certeza, a fotografia do ensino superior no Brasil ficaria ainda mais nítida.

Conclusões

Este estudo teve o objetivo de jogar uma luz sobre um aspecto que, geralmente, é desconsiderado tanto na confecção de um teste múltipla-escolha quanto na interpretação dos resultados do mesmo: a porção de respostas que o indíviduo realmente sabe dentre aquelas que acertou. É uma maneira de separar o joio do trigo nas respostas certas do indíviduo.

Tanto para o caso hipergeométrico, quanto para o multinomial obtivemos as distribuições de probabilidades do número de respostas que o indíviduo sabe no domínio, a partir de seus resultados no teste. Com estas distribuições a posteriori calculamos os estimadores bayesianos dos parâmetros de interesse.

Como sugestão de aplicação dos resultados deste estudo propusemos uma alternativa à forma de implementação e interpretação atual do Exame Nacional de Cursos do Ensino Superior.

Apêndice

A demonstração do lema da página 34 é baseada em resultados obtidos utilizando os Teoremas de Basu, especialmente o Teorema 1.

O teorema 1 de Basu diz que dada uma estatística suficiente e completa T, a estatística Y é ancilar se ela for (condicionalmente) independente de T para cada valor do parâmetro Θ.

Suponha uma seqüência $X_1, X_2, ...$de variáveis gama mutuamente independentes com parâmetros $\alpha_1, \alpha_2, ...$. Isto é, a função densidade de probabilidade de X_n é $f(x_n) = C\, x^{\alpha_n - 1} e^{-x}$, $x > 0$, $n = 1, 2, ...$

Seja $T_n = X_1 + X_2 + ... + X_n$ e $Y_n = T_n / T_{n+1}$, $n = 1, 2, ...$ T_n possui distribuição gama com parâmetro $\alpha_1 + \alpha_2 + ... + \alpha_n$ e Y_n tem distribuição beta com parâmetros $\alpha_1 + \alpha_2 + ... + \alpha_n$ e α_{n+1}, como pode ser provado facilmente.

Agora, o fato de $Y_1, Y_2, ..., Y_{n-1}, T_n$ terem de ser mutuamente independentes não é trivial. Vamos mostrar que isto procede.

Introduziremos um parâmetro de escala Θ na distribuição conjunta de $X_1, X_2, ...$ Assim, as X_n's são mutuamente independentes com função densidade de probabilidade igual a: $f(x_n) = C(\Theta)\ x^{\alpha_n - 1} e^{-x/\Theta}$, $x > 0$, $n = 1, 2, ...$ Considere os α's como constantes positivas e Θ como o parâmetro desconhecido.

Seja $\boldsymbol{X}^{(n)} = (X_1, X_2, ...X_n)$ a amostra. Então $T_n = X_1 + X_2 + ... + X_n$ é uma estatística suficiente completa. A estatística-vetor $\boldsymbol{Y}^{(n-1)} = (Y_1, Y_2, ..., Y_{n-1})$ é invariante em escala. Como Θ é o parâmetro de escala, concluímos que $\boldsymbol{Y}^{(n-1)}$ é ancilar. Do teorema 1 de Basu segue que, para cada n, $\boldsymbol{Y}^{(n-1)}$ é independente de T_n. Como $(\boldsymbol{Y}^{(n-1)}, T_n)$ é função dos primeiros n X_i's, o par $(\boldsymbol{Y}^{(n-1)}, T_n)$ é independente de X_{n+1}. Desta forma $\boldsymbol{Y}^{(n-1)}$, T_n e X_{n+1} são mutuamente independentes.

Em decorrência do anterior temos que $\boldsymbol{Y}^{(n-1)}$ é independente de $Y_n = T_n / (T_n + X_{n+1})$. Portanto, para cada $n \geq 2$, o vetor $Y_1, Y_2, ..., Y_{n-1}$ é independente de Y_n e isto significa que os Y_i's são mutuamente independentes.

Basu (1982) caracterizou as distribuições de Dirichlet por transformações de variáveis aleatórias gama independentes.

Para a adaptação do anterior ao nosso lema, considere p_1, p_2, p_4 gamas independentes, com parâmetros α_1, α_2, α_4, respectivamente. Então a distribuição conjunta de (p_1, p_2, p_4) é $D(\alpha_1, \alpha_2, \alpha_4)$.

Considere $T_2 = y = p_1 + p_2$ e $Y_1 = p_1 / (p_1 + p_2)$. Assim definidos, temos que $Y_1 \sim B(a_1, a_2 + a_4)$, o que representa a afirmação (ii).

Pela conclusão acima observada temos que y é independente de p_1 / y. Isto prova a afirmação (iii) do lema.

Para a afirmação (i) temos que se $(p_1, p_2, p_4) \sim D(\alpha_1, \alpha_2, \alpha_4)$, então $(p_1 + p_2, p_4) \sim D(\alpha_1 + \alpha_2, \alpha_4)$ ou, escrito de outra forma, $(p_1 + p_2) \sim B(\alpha_1 + \alpha_2, \alpha_4)$, como é convencionado.

BIBLIOGRAFIA

Andrade, D. F. e Valle, R. C. Introdução à Teoria da Resposta ao Item. Estudos em Avaliação Educacional, 18, 13-32, 1998.

Andrade, D. F., Tavares, H. R. e Valle, R. C. Teoria da Resposta ao Item: Conceitos e Aplicações, Associação Brasileira de Estatística, 2000.

Baker, F.B. Item Response Theory – Parameter Estimation Techniques. New York: Marcel Deker, Inc., 1992

Barlow, R. E. e Pereira C. A. B., The Bayesian Operation and Probabilistic Influence Diagrams. Topics in Statistical Dependence, Block & Savits Editors, IMS: 19-44, 1990

Basu, D. Basu Theorems, in The Encyclopedia of Statistical Sciences, eds. N. L. Johnson & S. Kotz, 1, pp. 193-196, 1982.

Basu, D. e Pereira C. A. B., On the Bayesian Analysis of Categorical Data: the Problem of Nonresponse, Journal of Statistical and Inference 6, 345-362, 1982.

Goldstein, H. e Wood, R. Five decades of item response modelling. British Journal of Mathematical and Statistical Psychology, 42, 139-167, 1989.

Hambleton, R., Swaminathan, H. and Algina, J. A Bayesian Decision-Theoretic procedure for use with criterion-referenced tests. Journal of Educational Measurement, 1975, 12, 2.

Hively, W., Maxwell, G., Rabhel, G., Sension, D. e Lundin, S. Domain referenced curriculum evaluation: A technical handbook and case study. CSE Monograph Series in Evaluation, Number 1. Los Angeles: Center for the study of evaluation. University of California, 1973.

Linden, W. J. van der e Hambleton, R. K. Handbook of modern Item Response Theory. New York: Springer-Verlag, 1997

Lord, F. M. Applications of Item Response Theory to Practical Testing Problems. Hillsdale: Lawrence Erlbaum Associates, Inc., 1980

Lord F.M. e Novick M. R., Statistical Theories of Mental Test Scores. Addison-Wesley Publishing Company, Inc., 1968

Millman, J. Sampling plans for domain referenced tests. Educational Technology, 1972, 14, 17.

Millman, J. Determining Test Length and Passing Scores for Objectives-based Tests. Los Angeles: Instructional Objectives Exchanges. 1972

Millman, J. Criterion referenced measurement. Evaluation in Educational Current Applications, H. Popham, ed. American Educational Research Association, McCutchan Pub. Co., 1974.

Mislevy, R.J. Bayes modal estimation in item response models. Psychometrika, 51, 177-195, 1986.

Nitko, A. Problems in the development of criterion-referenced tests: The IPI experience. CSE Monograph Series in Evaluation, Number 3. Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Novick, M., e Lewis, C. Prescribing test length for criterion referenced measurement: CSE Monograph Series in Evaluation, Number3. Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Paulino, C. D. M., Pereira, C. A. B., Bayesian Analysis of Categorical Data Informatively Censored. Commun. Statist. A 21, 2689-705, 1992

Paulino, C. D. M., Pereira, C. A. B., Bayesian Methods for Categorical Data under Informative General Censoring, Biometrika, 82, 2, pp. 439-46, 1995.

Popham, W., e Husek, T. Implications of criterion referenced measurement. Journal of Educational Measurement , 1969, 6, 1.

Popham, W. Selecting objectives and generating test items for objectives-based tests. CSE Monograph Series in Evaluation, Number 3. Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Visco, R. The Classification of Students with respect to Achievement, with Applications for Statewide Assessment. Doctoral dissertation , Florida State University, 1977.

Zúñiga, L. An Inferential Method for Domain Referenced Tests. Doctoral dissertation, Florida State university, 1977.