

A COMPARABILIDADE DE ESCALAS PARA A ATRIBUIÇÃO DE PESOS EN EXAMES DE SELEÇÃO

CARMEN LUCIA DE MELO BARROSO

Fundação Carlos Chagas e Pontifícia Universidade Católica de São Paulo

BRAZIL

RESUMO. Quando a seleção de indivíduos é feita em função de uma medida composta, o peso efetivo de cada variável componente desta medida é definido em termos da contribuição da variável para a variância do composto.

O objetivo do presente trabalho é demonstrar formalmente que o peso efetivo de uma variável depende não só do peso nominal que lhe é deliberadamente atribuído, mas também da variância da variável e de suas covariâncias com as demais variáveis. Sendo razoável supor que, em muitas situações, as covariâncias de cada variável com as demais não apresentem grandes diferenças, impõe-se a necessidade do uso de escalas padronizadas, com variâncias iguais, a fim de que a influência de cada variável seja efetivamente determinada pelo peso que lhe é deliberadamente atribuído. Casos específicos de seleção de candidatos a universidades são apresentados a título de exemplo.

ABSTRACT. When composite scores are used to select individuals, the effective weight of each component variable of these scores is defined in terms of the contribution of the variable to the variance of the composite.

The objective of this paper is to demonstrate formally that the effective weight of a variable depends on the nominal weight deliberately attributed to it as well as on the variance of that variable and its covariance with each of the remaining variables. One might reasonably suppose that in many situations the covariance of each variable with each of the remaining variables will not differ greatly. The need, therefore, arises for the use of standardized scales, with equal variances, so that the influence of each variable might be effectively determined by the weight deliberately attributed to it. Specific examples involving the selection of university students are given for illustrative purposes.

Em concursos para preenchimento de vagas no serviço público, em empresas privadas, ou em estabelecimentos de ensino, raras são as vezes em que a seleção é feita com base em uma única variável¹. Geralmente, os candidatos são selecionados em função de uma medida composta, uma combinação linear de diversas variáveis. Sempre que mais de uma variável fôr utilizada coloca-se o problema de atribuição de pesos. A chamada "média não ponderada" constitui, na verdade, um caso particular de ponderação em que se atribuem pesos iguais a um para todas as variáveis. Neste caso, obviamente, está implícito que todas as variáveis são consideradas igualmente importantes.

Quer se trate desse caso, em que o peso relativo é o mesmo, para cada uma das variáveis, quer se trata de obter um composto para o qual o grau da contribuição de cada variável é controlado através de fatores multiplicativos diversos,² é útil fazer-se uma distinção entre os pesos *nominais*, aqueles atribuídos deliberadamente aos componen-

tes, e os pesos *efetivos*, os que componentes realmente têm no compôsto. A adequação dos pesos nominais deve ser objeto de cuidadosas pesquisas que estabeleçam a validade preditiva das variáveis. O presente artigo apenas delinea um procedimento para fazer com que os pesos efetivos sejam determinados pelos pesos nominais.

Na literatura especializada, o *pêso efetivo* de uma variável num compôsto é definido em termos da contribuição da variável para a variância do compôsto. Ora, isto implica em admitir que a medida composta, para um indivíduo, tem pouco significado em si própria, e seu sentido não é intrínseco, mas derivado da distribuição da medida composta entre todos os indivíduos.

Este é um pressuposto razoável, uma vez que somente escalas ordinais ou intervalares são usadas em psicologia e em medidas educacionais, e que a interpretação de um escore individual requer o uso de normas de referência, já que a unidade de um teste pode ser amplamente diferente da unidade de outro. Por outro lado, como bem assinalou Horst (1965), se pressupomos que certas variáveis psicológicas e educacionais são contínuas, as unidades são definidas arbitrariamente. Assim a comparabilidade e a possibilidade de combinação entre as escalas torna-se, em grande parte, uma questão de definição.

A definição de unidade mais comum tem sido o item, na prova objetiva, ou a questão, na prova tradicional. Ora, o item, ou a questão, como unidade de medida, apresenta grandes limitações.³ É óbvio que nada se esclarece com a informação de que o candidato *i* respondeu, acertadamente, um número *x* de questões da prova *j*. Se a este dado se acrescentar a informação do número de perguntas que constam da prova, o significado deste resultado ainda permanece obscuro, pois nada se sabe sobre o grau de dificuldade da prova utilizada. É evidente que esta dificuldade deve ser calculada para o grupo específico com o qual o candidato está competindo, pois trata-se de um índice que varia em função do nível de habilidade do grupo considerado.

Por simples e evidente que seja esta noção, ela não é amplamente difundida. É comum, por exemplo, que o candidato procure saber quantos pontos obtêve no vestibular, sem considerar que a obtenção de um certo número de pontos pode ter significados completamente diversos, dependendo da distribuição dos pontos obtidos pelos demais candidatos.⁴

Sem pretender analisar a teoria implícita nestes procedimentos, registremos apenas que, atualmente, a prática generalizada em psicometria é a definição da unidade de medida em termos da variação entre indivíduos.⁵

Uma vez que, em exames de seleção, não se trata de tomar de-

A COMPARABILIDADE DE ESCALAS PARA A ATRIBUIÇÃO DE PESOS

cisões absolutas em escalas arbitrárias (come poderia ser o caso se se procurasse verificar a habilitação), mas sim de escolher os “melhores”, o julgamento do resultado de um indivíduo depende da sua posição relativa no grupo. Se esta posição é determinada em função de uma medida composta, que resulta da combinação de diversas variáveis, o peso efetivo de cada variável no composto será definido em termos de sua contribuição para a variância do composto.

Stanley e Wang (1968) realizaram uma pesquisa sobre os métodos de atribuição diferencial de pesos, e sobre os estudos empíricos a eles relacionados. Este estudo inclui sistemas fixos e variáveis, estabelecidos empiricamente ou “a priori”, enfim, todos os principais procedimentos usados nos últimos 50 anos. Entre as várias conclusões a que chegaram, interessa-nos a de que a definição de peso efetivo de uma variável num composto em termos de sua contribuição para a variância do composto só não seria adequada no caso de decisões absolutas em escalas arbitrárias.

Como podemos verificar no Apêndice, a contribuição de uma variável para a variância do composto depende, não só do peso nominal que lhe é atribuído, mas, também, da variância desta variável e de suas covariâncias com as demais variáveis.⁶

Ora, se queremos que os pesos efetivos sejam iguais aos pesos nominais, isto é, se queremos que a nossa atribuição de importância a cada variável seja respeitada, precisamos controlar as variâncias e as covariâncias.

Em muitas situações, é razoável supor que, em média, as covariâncias de cada variável com as demais não apresentem grandes diferenças. Este pode ser facilmente obtido pela transformação de cada uma das variáveis em uma escala padronizada, com variância igual a um. Esta escala é o resultado da transformação do escore bruto de cada candidato em escore padrão, de acordo com a seguinte fórmula:

$$Z_{ij} = \frac{X_{ij} - M_j}{S_j}$$

onde

Z_{ij} = escore padrão do candidato.

X_{ij} = n° de questões certas (escore bruto) do candidato i , na prova j .

M_j = média dos escores brutos dos candidatos, na prova j .

S_j = raiz quadrada da variância (desvio padrão) dos brutos da prova j .

Assim, a classificação dos candidatos pode ser feita em função de um composto em que a influência de cada um dos componentes é

efetivamente determinada pelo pêso que lhe é deliberadamente atribuído, já que todos os componentes terão variâncias iguais.

Uma transformação conveniente é adotada pelo College Entrance Examination Board, em que a média é fixada em 500 e o desvio-padrão em 100. Dêste modo, evitam-se três problemas: o uso de escores negativos, o uso de decimais (características do escore padrão) e a ocorrência freqüente de empates, que se verifica quando o número de candidatos é grande e a escala utilizada não possibilita muita discriminação (como é o caso de estanino, por exemplo.)⁷

Em exames de seleção, a adoção do sistema de transformação dos escores para a uniformização das variâncias deve levar em consideração que a opinião pública, pouco esclarecida, talvez encontre dificuldade em compreender e, portanto, aceitar, os fundamentos que justificam o sistema. A noção de que um escore deriva seu significado da distribuição dos escores, apesar de simples, não é de compreensão imediata.

Sem ignorar êsses pontos, podemos concluir que, sempre que se quiser selecionar os indivíduos que melhor se classificam dentro de um grupo, em função de um escore global, resultado da combinação de escores de diversas provas, impõe-se o uso de escores padronizados, para assegurar que a influência relativa de cada prova sobre o escore global seja igual ao pêso que lhe é atribuído deliberadamente.

REFERENCIAS

- APA, AERA, NCME. *Standards for Educational and Psychological Tests and Manuals*. Washington, D. C.: APA, 1966.
- Ghiselli, E. *Theory of Psychological Measurement*. New York: McGraw-Hill, 1964.
- Horst, P. *Psychological Measurement and Prediction*. Belmont, Calif: Wadsworth, 1965.
- Lord, F. M. e Novick, M. D. *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1968.
- Stanley, C. e Wang, M. D. *Differential Weighting: A Survey of Methods and Empirical Studies*. Princeton, N. J.: College Entrance Examination Board, 1968.
- Vaughn, K. W. Planning the Objective Test. In E. F. Lindquist (Ed.), *Educational Measurement*. Washington, D. C.: ACE, 1950.

APENDICE

A inspeção da fórmula da variância de uma soma de variáveis (ver, por exemplo, Lord e Novick, 1968, p. 96)—desconsiderada a influência dos pesos nominais—mostra que o pêso efectivo natural, ou seja, a contribuição de cada variável para a variância do composto, resulta, não somente da sua própria variância, mas, também, de suas covariâncias com as demais variáveis.

Nos exames vestibulares, a variância dos escores globais, antes da atribuição dos pesos nominais, é igual a:

$$\text{Var} (y) = \text{Var} (x_1) + \text{Var} (x_2) + \dots + \text{Var} (x_n) + 2 \text{Cov}(x_1x_2) + \dots + 2 \text{Cov}(x_{i-1} x_i)$$

onde

A COMPARABILIDADE DE ESCALAS PARA A ATRIBUIÇÃO DE PESOS

$$y = x_1 + x_2 + \dots + x_n = \sum_{j=1}^n x_j$$

x_1 = escore de Química

x_2 = escore de Português

.....

x_n = escore do Exame Prático

Atribuindo um conjunto de pesos nominais p_j ($j = 1, \dots, n$) ao conjunto das variáveis x_j ($j = 1, \dots, n$), o escore global de cada candidato i é dado por:

$$y_i = p_1 x_{1i} + p_2 x_{2i} + \dots + p_n x_{ni}$$

sendo x_{1i} = escore de Química do candidato, $p_1 = 100$, peso da prova de Química, etc.

A variância do composto é dada por:

$$\begin{aligned} \text{Var}(p_1 x_1 + p_2 x_2 + \dots + p_n x_n) &= \\ &= p_1^2 \text{Var}(x_1) + p_2^2 \text{Var}(x_2) + \dots + p_n^2 \text{Var}(x_n) + \\ &+ 2 p_1 p_2 \text{Cov}(x_1 x_2) + 2 p_1 p_3 \text{Cov}(x_1 x_3) + \dots + \\ &+ 2 p_{n-1} p_n \text{Cov}(x_{n-1} x_n) \end{aligned}$$

Naturalmente, esta fórmula pode ser reduzida para:

$$\text{Var}(y) = \sum_{j, j'=1}^n p_j p_{j'} \text{Cov}(x_j, x_{j'})$$

onde

$$\text{Cov}(x_j, x_{j'}) = \text{Var}(x_j) \text{ se } j = j'$$

A contribuição da variável j para a variância do composto é dada por:

$$C_j = P_1 P_j \text{Cov}(x_1 x_j) + P_2 P_j \text{Cov}(x_2 x_j) + \dots + \\ + P_{j-1} P_j \text{Cov}(x_{j-1} x_j) + P_j^2 \text{Var}(x_j) + P_{j+1} P_j \\ \text{Cov}(x_{j+1} x_j) + \dots + P_n P_j \text{Cov}(x_n x_j)$$

Portanto, a contribuição da variável i para a variância total depende:

- a) dos pesos nominais p_i
- b) da variância de x_i
- c) das covariâncias de x_i com as outras variáveis.

NOTAS

¹Nos vestibulares do CESCEM (Centro de Seleção de Candidatos às Escolas Médicas e de Currículo Biológico), por exemplo, os candidatos dos cursos das carreiras biológicas são selecionados em função da combinação dos escores obtidos em 7 provas de aferição de conhecimentos e habilidades em campos específicos.

²Este é o caso mais comum. Utilizando o exemplo do CESCEM novamente: os escores de Português são multiplicados por 10, os de Inglês por 50. Assim, se um indivíduo obtiver o mesmo escore nas duas provas, o seu escore de Português (o seu conhecimentos de Português, na medida em que seu escore nesta prova for capaz de medi-lo) contribuirá duas vezes mais para o composto do que o escore de Inglês. (Outras provas recebem pesos que variam de carreira para carreira).

³Gertrude Stein definiu: "uma rosa é uma rosa, é uma rosa".. Mas nem sempre um item é um item, é um item. Nos exemplos abaixo, a natureza elástica dos itens os torna inadequados como unidade de medida de conhecimento matemático, porque é evidente que um indivíduo que responde acertadamente o item 2 demonstra um grau de conhecimento muito maior que o do indivíduo que é capaz de responder ao item 1.

item 1 2 + 2 = ?

item 2 A potência enésima (n inteiro) de um número imaginário puro é sempre um número imaginário puro?

É claro que estes 2 itens não seriam incluídos numa mesma prova. No entanto eles servem de exemplo exagerado do que ocorre freqüentemente, quando não temos informação sobre o grau exato de dificuldade do item. O problema é maior quando se trata de comparar itens de provas diferentes.

⁴Se, por exemplo, 90% dos candidatos obtiveram escores inferiores a 380, quem obteve este resultado (380) demonstrou um nível de habilidade superior ao da maioria e, sem qualquer dúvida, será admitido na Universidade, se o número de vagas for suficiente para atender a pelo menos 10% dos candidatos. Se, ao contrário, 90% dos candidatos obtiveram escores superiores a 380, a obtenção de 380 indica um nível de habilidade inferior ao da grande parte do grupo, e um candidato nestas condições só será aceito na Universidade se o número de candidatos pouco exceder o número de vagas.

⁵Veja-se, por exemplo: Magnusson (1967), APA, NCME, AERA (1966), e Ghiselli (1964).

⁶Alguns exemplos podem contribuir para esclarecer este ponto. Argumentando pelo absurdo: um test de Matemática é construído e seu nível de dificuldade é excessivamente alto: todos os candidatos obtêm zero. Por maior que seja o peso nominal atribuído a esse teste, seu peso efetivo será zero. E isto ocorre, não por que o escore obtido pelos candidatos seja zero, mas porque todos os candidatos obtiveram o mesmo escore. O peso efetivo também seria zero se todos tivessem obtido o escore máximo. O que importa, no caso, é a ausência da variância, o que podemos demonstrar com outro exemplo. Aplicamos uma prova de

A COMPARABILIDADE DE ESCALAS PARA A ATRIBUIÇÃO DE PESOS

Português e outra de Matemática a 3 candidatos: A, B e C. à prova de Português atribuímos o peso nominal 4 e à de Matemática, o peso 6. Os resultados obtidos são os seguintes:

<i>Candid.</i>	<i>Port.</i>	<i>Matem.</i>	<i>Escore Global</i>	<i>Classificação</i>
A	10	4	$10 \times 4 + 4 \times 6 = 64$	1°
B	5	4	$5 \times 4 + 4 \times 6 = 44$	2°
C	0	4	$0 \times 4 + 4 \times 6 = 24$	3°

A classificação de cada um dos candidatos permaneceria exatamente a mesma com ou sem a inclusão da prova de Matemática. Como essa prova em nada altera a posição relativa dos candidatos (e é isso o que nos interessa) seu peso nominal seja maior que o de Português.

Esse é o caso em que a variância é nula, mas pode-se mostrar que, controlando-se os pesos nominais e as covariâncias, quanto maior a variância de um teste, maior será sua influência na classificação final dos candidatos segundo um escore global, para o qual este teste contribui.

Magnumsson (1967) descreve detalhadamente as transformações lineares.