

# Análise de Correspondência: Uma Aplicação do Método à Avaliação de Serviços de Vacinação

## Correspondence Analysis: An Application of the Method to the Evaluation of Vaccination Services

Marília Sá Carvalho<sup>1</sup>

Cláudio José Struchiner<sup>2</sup>

CARVALHO, M. S. & STRUCHINER, C. J. *Correspondence Analysis: An Application of the Method to the Evaluation of Vaccination Services*. *Cad. Saúde Públ.*, Rio de Janeiro, 8 (3): 287-301, jul/set, 1992.

*Correspondence analysis is a statistical method developed for descriptive exploratory analysis of large matrices. Through this technique, the most important relations between a large pool of categorical and continuous variables may be plotted graphically. In this paper the described technique was applied to the data of an evaluation process of the National Program of Immunizations (PNI).*

*This evaluation was carried out in 660 health clinics, distributed over 98 Brazilian cities. The variables analysed were related to the performance in PNI activities or potential determinants of this performance. The correspondence analysis was followed by an ascending hierarchical classification process of the Health Clinics, using the computer program "Systeme Portable Pour L'Analyse de Données — SPA".*

*The performance pattern was not homogeneous. The clinics were classified in five groups according to the performance in the PNI activities. In 58% of the Health Clinics, even the most elementary norms for the PNI were not followed.*

*The correspondence analysis is a powerful tool for inspecting large data matrices and selecting the most important variables for a more detailed subsequent analysis.*

**Keywords:** *Correspondence Analysis; Assessment; Vaccination*

### INTRODUÇÃO

A necessidade de se estudar **simultaneamente** as relações existentes entre um grande conjunto de variáveis fez com que fossem desenvolvidas as técnicas estatísticas de análise multivariada (Gouvêa, 1990).

A análise de correspondência, conforme desenvolvida por um grupo de estatísticos franceses desde o início dos anos 60, é teorica-

mente equivalente a outras técnicas desenvolvidas em diferentes contextos desde meados da década de 30 (Greenacre, 1981), fazendo parte de um conjunto de métodos utilizados para a análise descritiva exploratória de grandes tabelas. A concepção geral é semelhante à análise de componentes principais e à análise fatorial, diferenciando-se destas, entre outros aspectos, por permitir a inclusão de variáveis categóricas. Embora seja evidente sua utilidade nos estudos transversais em epidemiologia, especialmente na exploração inicial de dados discretos e/ou contínuos, a análise de correspondência é pouco usual na literatura do setor.

Neste trabalho, aplicou-se a análise de correspondência, seguida de técnica de classificação hierárquica ascendente, aos dados obtidos na avaliação do Programa Nacional de Imuni-

<sup>1</sup> Departamento de Epidemiologia e Métodos Quantitativos em Saúde da Escola Nacional de Saúde Pública. Rua Leopoldo Bulhões, 1480, Rio de Janeiro, RJ, 21041-210, Brasil.

<sup>2</sup> Núcleo de Doenças Endêmicas Samuel Pessoa da Escola Nacional de Saúde Pública. Rua Leopoldo Bulhões, 1480, Rio de Janeiro, RJ, 21041-210, Brasil.

zações (PNI). Esta investigação foi realizada nacionalmente em 1985, através de um convênio entre a Secretaria Nacional de Ações Básicas de Saúde, do Ministério da Saúde, e a Escola Nacional de Saúde Pública.

O objetivo deste artigo é divulgar a técnica estatística empregada. Sem pretensão de substituir as centenas de páginas dos livros texto dedicados ao assunto, acreditamos que este artigo ajudará o pesquisador que estiver se propondo a utilizar a análise de correspondência, através da apresentação dos resultados e da forma como foram interpretados, e possivelmente interessará outros no seu estudo. Além disso, os resultados da avaliação do PNI, extensamente discutidos na monografia de mestrado (Carvalho, 1990) que originou o presente trabalho merecem profunda reflexão por parte dos sanitaristas do país.

## DESCRIÇÃO DO MÉTODO ESTATÍSTICO

A análise de correspondência é especialmente indicada para descrever matrizes com grande volume de dados discretos e sem uma estrutura claramente definida *a priori*. Este método permite a visualização das relações mais importantes de um grande conjunto de variáveis entre si. Os resultados são apresentados sob forma de gráficos, onde estão representadas as categorias de cada variável e onde se pode observar as relações entre estas, através da distância entre os pontos desenhados (Greenacre, 1981; Lebart et al., 1977; Lebart et al., 1984).

Em geral, o ponto de partida é uma matriz

onde nas linhas temos os "casos" e nas colunas, as variáveis de interesse. No nosso exemplo, as linhas são as unidades de saúde e as colunas, as variáveis utilizadas na avaliação do programa de imunizações: situação da geladeira, existência de registro de vacinas, acerto nas indicações e contra-indicações à vacinação, entre outras (ver lista completa na Tabela 1).

Assim, cada unidade de saúde (linha da matriz de dados) apresenta um "perfil" definido por suas respostas em cada quesito do instrumento de coleta de informações utilizado. Da mesma forma, cada variável (coluna) terá um "perfil" em função da distribuição das unidades de saúde. Por exemplo, geladeiras em perfeito estado somente foram encontradas em unidades que também responderam adequadamente em relação às contra-indicações à vacinação. Assim, começa a se configurar um perfil da variável "condição da geladeira", que embora intuitivamente um pouco mais difícil de entender do que o "perfil" da unidade, matematicamente é absolutamente equivalente.

Nossa matriz "unidades X variáveis" matematicamente pode ser representada como uma nuvem de "n" (número de Unidades), pontos em um espaço de "p" (número de variáveis), dimensões. Ou seja, ao invés da representação usual em um espaço de três dimensões (x,y,z), cada variável é uma dimensão de um espaço, e as unidades podem ser localizadas neste espaço através de coordenadas, que são os valores que cada variável adquire. Por exemplo, as Unidades de São João e de Buraco Quente assumem o seguinte perfil na nossa matriz, neste exemplo com quatro dimensões apenas:

Variáveis	Categorias	Unidades	
		São João	Buraco Quente
Geladeira	Ótima	S	N
	Regular	N	N
	Errada	N	N
Taxa de Cobertura Vacinal	Calcula	S	N
	Recebe	N	N
	Desconhece	N	S
Controle de Estoque	Sim	S	N
	Não	N	S
Contra-Indicações	Corretas	S	N
	Erradas	N	S

N — não apresenta  
S — apresenta

TABELA 1. Número, Contribuição Absoluta, Cosseno Quadrado e Valor Teste nos Dois Eixos de Maior Auto-Valor Para as Variáveis Ativas

Variável	Símbolo = Categoria	Número	Contribuição Absoluta		Cosseno Quadrado		Valor Teste	
			Eixo 1	Eixo 2	Eixo 1	Eixo 2	Eixo 1	Eixo 2
Geladeira	GELE = Inadequada	199	69	5	20	1	- 11.37	- 2.84
	GELM = Parcialmente Correta	145	1	20	0	3	- 1.33	- 5.21
	GELO = Ótima	306	56	1	21	0	11.79	1.48
Registro de Vacinas	*GELI = Ignorado	4	-	-	-	-	- 0.91	2.25
	GELF = Ausente	6	-	-	-	-	- 0.41	- 0.63
	SRUC = Correto	617	2	0	8	0	7.29	- 0.90
Dose p/ Cálculo Cob. Vac.	NRUC = Inexistente	42	37	0	8	0	- 7.45	0.77
	*IRUC = Ignorado	1	-	-	-	-	0.50	0.90
	SDOS = Correta (3 <sup>ª</sup> )	275	59	14	20	3	11.63	- 4.88
Taxa de Cobertura Vacinal	NDOS = Incorreta/Não Sabe	385	42	10	20	3	- 11.63	4.88
	CBRN = Ausente	457	70	6	46	3	- 17.38	- 4.67
	CBRT = Recebe Taxa Calculada	133	113	213	28	40	13.78	15.97
Taxa de Abandono do Programa	CBRS = Realiza o Cálculo	61	45	202	10	33	8.23	- 14.86
	*CBRI = Ignorado	9	-	-	-	-	0.94	0.45
	ABNN = Ausente	564	26	10	38	10	- 15.63	- 8.19
Controle de Estoque	ABNT = Recebe Taxa Calculada	74	128	212	29	35	13.92	14.94
	ABNS = Realiza o Cálculo	12	35	181	7	27	6.30	- 11.58
	*ABNI = Ignorado	10	-	-	-	-	2.27	- 2.29
Contra-Indicações	SSTQ = Correto	377	23	13	11	4	8.18	- 5.39
	NSTQ = Inexistente	273	32	19	11	4	- 8.74	5.81
	*ISTQ = Ignorado	10	-	-	-	-	2.08	- 1.59
Multivacinação	SCID = Corretas	79	84	1	19	0	11.29	- 1.37
	NCID = Erradas	564	11	0	16	0	- 10.34	- 11.60
	ICID = Ignorado	17	0	40	0	6	- 0.12	6.38
Estratégias de Vacinação	SVMU = Correta	538	10	0	12	0	9.06	- 1.48
	NVMU = Errada	113	49	0	12	0	- 8.77	0.56
	*VMU = Ignorado	9	-	-	-	-	- 11.82	3.16
Estratégias de Vacinação	ETTO = Só demanda espontânea	424	27	15	15	6	- 10.13	- 6.40
	ETTI = Uma outra estratégia	143	68	12	16	2	10.45	- 3.87
	ETT2 = Duas outras ou mais	89	5	14	1	2	3.15	- 4.32
	ETTI = Ignorado	4	-	-	-	-	- 0.16	0.43

\* Categorias com total inferior a 11 unidades tratadas como suplementares

A Unidade de São João apresenta todas as variáveis da avaliação do PNI exatamente conforme o recomendado nas normas. Já a de Buraco Quente... É evidente que, no espaço multidimensional definido pelas variáveis, as Unidades de São João e Buraco Quente ficam em situações opostas. Da mesma forma, pode-se inverter as posições, imaginando a representação das "p" categorias de cada variável em um espaço de "n" dimensões, cada uma sendo uma unidade. Ou seja, categorias de variáveis que tendem a acontecer nas mesmas unidades tenderiam a ficar em locais próximos neste espaço, e vice-versa. No exemplo acima, ficariam próximas: geladeiras corretas, cálculo de cobertura vacinal, controle de estoque e contra-indicações corretas.

A representação gráfica usual separa as informações, apresentando, por exemplo, um gráfico de barras da distribuição das categorias de cada variável. A representação gráfica obtida através da análise de correspondência possibilita visualizar a distribuição das variáveis na sua relação com todas as outras. Cada categoria de cada variável é representada por um ponto, e as distâncias entre os pontos representam as relações entre as variáveis que se deseja analisar. Por conseguinte, é essencial a compreensão do modelo utilizado para a determinação destas distâncias (Greenacre, 1981).

A distância entre duas categorias dispostas em uma coluna da matriz é calculada a partir de seu perfil de coluna e levando em consideração o "peso" de cada linha, ou seja, o total desta linha no total de elementos da matriz. Esta operação é repetida para as linhas. Assim, estas categorias têm uma localização no espaço que é função, simultaneamente, de seu perfil de linha e do "peso" de cada coluna no total da matriz. Ou seja, há uma perfeita correspondência entre linhas e colunas, sendo a frequência relativa, ou perfil de linha, o "peso" das colunas, e vice-versa. Este aspecto permite a visualização de pontos de linhas e colunas em um mesmo espaço. A distância calculada desta forma é chamada "distância qui-quadrado".

Assim, cada categoria de cada variável tem calculada sua distância para todas as demais, configurando uma nuvem de pontos em um espaço multidimensional. Idealmente, observaríamos a distribuição desta "nuvem" de variáveis

no espaço, verificando a relação entre elas. Entretanto, não é possível inspecionar visualmente um espaço multidimensional. Porém, esta nuvem pode ser projetada em planos. Estes planos são selecionados pela sua capacidade de preservar, ao máximo, a distância entre os pontos, refletindo, o melhor possível, as relações entre as categorias. Matematicamente, a solução deste problema é dada pelos auto-vetores da matriz de dados. Infelizmente, não nos foi possível explicar como é feita esta passagem, desenvolvida em capítulo da álgebra.

Cada auto-vetor é um dos eixos que define nosso espaço multidimensional, respondendo por uma percentagem da variância total da matriz. A importância deste percentual é grande, pois permite avaliar a capacidade de cada eixo representar a "nuvem" de pontos. Quanto maior o percentual, melhor a representação da distribuição dos pontos no espaço e, conseqüentemente, das relações existentes entre as variáveis.

Como os eixos são escolhidos pela sua capacidade de preservar as relações entre as categorias das variáveis, cada uma delas contribui para a definição dos eixos de forma diferenciada. Esta contribuição é chamada "contribuição absoluta". Assim, algumas categorias têm papel preponderante na construção de um ou outro eixo dos gráficos, indicando tratar-se de categorias que apresentam-se fortemente correlacionadas com este eixo e entre si. Ao final, a soma das contribuições absolutas de todas as categorias em cada eixo é igual a 1 (um).

Um aspecto importante é a seleção das variáveis que participarão da análise. Na definição do espaço de representação gráfica, algumas variáveis têm papel dito "ativo", ou seja, são as variáveis cuja distribuição é a base para o cálculo dos eixos. Outras serão colocadas no gráfico após a definição dos eixos, assumindo papel "ilustrativo", também chamado "suplementar". Este papel pode ser invertido, e o grupo inicialmente "ativo" passar a "ilustrativo", e vice-versa. O objetivo desta separação é melhorar a análise, agrupando as variáveis segundo o que se deseja observar.

O essencial para apoiar a compreensão da técnica utilizada foi apresentado até aqui. Entretanto, alguns outros elementos subsidiam a análise e interpretação dos gráficos, sendo

também apresentados nos resultados. Destes, os mais importantes são:

- Cosseno Quadrado — o cosseno do ângulo definido pelo vetor de cada ponto com cada eixo, elevado ao quadrado, multiplicado por 100, que indica a proximidade do ponto ao eixo;
- Valor-Teste — distância de cada ponto à origem do eixo, em número de desvios padrões, que orienta a análise das variáveis suplementares. Utilizando um marco de 5% de probabilidade de ocorrência casual da localização de determinado ponto no eixo, considera-se significativos apenas os pontos com valores-teste acima de 2 (dois), em números absolutos (Benzecri, 1979; Cesia, 1982; Cesia, 1985).

A etapa seguinte do trabalho consiste em agrupar as unidades por classificação hierárquica ascendente. Retornando à nossa "nùvem" de unidades, onde as dimensões do espaço são dadas pelas variáveis, pode-se agrupar estes pontos segundo os grupamentos naturalmente formados, em função dos diferentes desempenhos das unidades. É exatamente por terem perfil semelhante que as unidades se localizam em pontos relativamente próximos no espaço.

Na classificação, foram utilizados apenas os eixos com auto-valores maiores do que o inverso do número de variáveis ativas, considerados os eixos com menor probabilidade de construção meramente casual (Morineau, 1984). Existem diferentes critérios de agrupamento dos pontos. O critério utilizado aqui foi o de menor redução da variância total da matriz (método de Ward) (Lebart et al., 1977). As unidades são agrupadas de duas em duas, de forma que os grupos resultantes representem a menor redução da variância total da matriz. Este processo se repete, juntando agora os pequenos grupos iniciais formados, e assim sucessivamente, construindo-se uma árvore de classificação que ascende até a junção dos dois últimos grupos e reconstituição do universo original. A junção de cada dois elementos ou dois conglomerados é denominada "nó". Cada nó apresenta, portanto, um determinado "peso", ou número de elementos que o compõem, e um índice de agregação. Este corresponde à distância qui-quadrado ponderada entre as classes.

Estes índices são apresentados através de um gráfico de barras, que permite a visualização de grandes saltos entre dois "nós", que representariam classes naturais. A última barra do gráfico corresponde à junção das duas últimas classes. A definição do número de classes em que se repartiu o conjunto de unidades é feita tendo como elemento essencial este gráfico.

A divisão em classes assim obtida é então submetida a outro procedimento, cujo objetivo é melhorar o posicionamento das unidades nas classes, "realocando" algumas que se distanciam mais do seu grupo, otimizando, portanto, a classificação obtida. Esta "rearrumação" é feita através de partições sucessivas ao redor de centros móveis. As coordenadas do centro de gravidade de um grupo de unidades são calculadas como a média das coordenadas em cada eixo para cada unidade. A técnica de "rearrumação" das unidades consiste basicamente em:

- Cálculo do centro de gravidade de cada classe obtida por classificação hierárquica ascendente;
- Se algum ponto classificado em um grupo encontrar-se mais próximo de outro centro de classe, ele é "realocado";
- Os centros de classe são recalculados em função das modificações e novamente os pontos são "rearrumados" em função de sua proximidade ao centro de gravidade das classes.

Este processo se interrompe quando as distâncias entre os centros de gravidade das classes não mais sofrem alterações ao serem "realocadas" as unidades. Ou seja, quando a distância entre os centros de classe é máxima, o que indica ter sido atingida a melhor distribuição possível com a técnica utilizada.

A descrição qualitativa das classes é feita através das categorias com maior significado na construção de cada classe.

Para avaliar esta questão, foi criado um indicador, denominado "Critério", que é construído calculando-se a distância entre a percentagem da categoria na classe e na população em geral, em número de desvios-padrão. Este é o principal elemento que permite a descrição das classes conforme seus elementos mais característicos. A análise baseia-se, essencialmente, na comparação da distribuição da categoria na

população e na classe.

Assim, está criada uma nova variável: a classe onde cada unidade foi alocada. Desta forma, a variável "classe" pode ser situada no mesmo espaço gráfico definido pela análise de correspondência. Assim, os centros de cada classe de unidades avaliadas por seu desempenho no PNI foram também colocados no gráfico da análise de correspondência para permitir a visualização das relações entre as classes e as variáveis.

## MATERIAL E MÉTODOS

Neste trabalho, são avaliadas a estrutura e o processo de ações normatizadas do PNI, executadas em todo o território nacional. Foram visitados 98 municípios, sendo sempre incluídas as capitais e, em seguida, os municípios de maior população em cada estado.

Em cada município, foram avaliadas as unidades de atendimento ambulatorial que contavam com médico no seu quadro de pessoal e que pertenciam a alguma das seguintes instituições: Secretaria Estadual de Saúde (SES), Secretaria Municipal de Saúde (SMS) ou Instituto Nacional de Assistência Médica e Previdência Social (Inamps). Além disso, foram visitados também os hospitais com maior número de leitos de pediatria em cada município selecionado, independente da instituição mantenedora. Deve-se observar que não é possível a generalização estatística das conclusões obtidas para os municípios não visitados, pois a seleção destes não foi aleatória. Entretanto, a equipe que trabalhou no projeto, e que discutiu os resultados descritivos apresentados no relatório da pesquisa (ENSP, 1986), não encontrou razão para acreditar que os demais municípios do mesmo porte tivessem, em geral, práticas muito diferentes.

As variáveis utilizadas para analisar o desempenho das unidades no PNI foram selecionadas com base em experiências anteriores (Benzecri, 1979; SESH-RJ, 1983; SES-BA, 1983). As variáveis "determinantes" foram selecionadas entre os fatores geralmente aceitos como tendo influência sobre o processo, alguns ligados estritamente ao setor, outros mais amplos, relacionados ao funcionamento global dos serviços.

Foram incluídas na análise as seguintes variáveis: adequação das geladeiras de vacinas às normas do PNI; registro de vacinas aplicadas; cálculo de cobertura vacinal e de taxa de abandono; estratégias de vacinação utilizadas pela unidade; controle de estoque de vacinas; conhecimento das normas relativas a contra-indicações à vacinação; indicação de vacinação múltipla; existência de normas e de boletins epidemiológicos na unidade; local onde é feita a programação das atividades; supervisões recebidas; instituição mantenedora e posição hierárquica da unidade; disponibilidade de dados demográficos; relação com a comunidade; presença de profissional capacitado; escolaridade do entrevistado.

Os dados foram coletados diretamente nas unidades de saúde, com um questionário fechado, em final de 1985. Cada tipo de instituição incluída no universo foi tratada como um extrato independente. O número de unidades locais amostradas foi definido em função do total de unidades de cada instituição em cada município. Foram visitadas 660 unidades, sendo 314 das Secretarias Estaduais de Saúde, 226 das Secretarias Municipais de Saúde, 77 do Inamps e 43 hospitais.

Os dados obtidos foram submetidos às técnicas estatísticas descritas, com o auxílio do "SPAD — *Système Portable pour l'Analyse des Données*", conjunto articulado de programas desenvolvido especialmente para a análise estatística descritiva de grandes tabelas. Foi utilizada a versão de 1985, cedida pelo Instituto de Matemática Pura e Aplicada — Impa, para utilização no computador de grande porte do Laboratório Nacional de Computação Científica.

Neste trabalho, as variáveis de avaliação das atividades estudadas foram designadas "ativas" e aquelas ligadas a fatores potencialmente determinantes deste desempenho, como a presença de supervisão, a existência de dados demográficos, entre outros, "ilustrativas". Em seguida, estes papéis foram invertidos como forma de verificar a consistência da análise realizada. Algumas variáveis sempre desempenham papel suplementar, como é o caso da "Escolaridade do Entrevistado", que funciona como controle da qualidade da entrevista, não tendo papel na conformação do padrão de desempenho das unidades. As categorias presentes em menos de

11 unidades — 1,67% do total — não foram consideradas ativas para evitar as distorções causadas pela concordância de respostas "ignorado", sempre nas mesmas unidades, que prejudicariam a análise. Para efeito dos cálculos dos auto-valores, estas unidades foram "realocadas" aleatoriamente pelas demais respostas.

Serão apresentados nos resultados alguns gráficos e listagens resumidas provenientes do SPAD. Informações mais detalhadas poderão ser obtidas com a autora.

## RESULTADOS

O gráfico de barras do percentual da variância total representado por cada auto-valor (Figura 1) sugere que sejam considerados, para análise dos gráficos, os dois primeiros planos, uma vez que os três primeiros eixos respondem por 34% da variância, com saltos importantes após o 1º e o 3º eixo. Não há regra estabelecida definindo até que plano analisar os gráficos, devendo ser lembrado aqui, novamente o caráter **exploratório** da técnica descrita. Neste trabalho, embora tenhamos analisado diversos planos, apresentaremos apenas o gráfico composto pelos dois primeiros eixos (Figura 2). As categorias de cada variável são representadas no gráfico por siglas de 4 (quatro) letras, apresentadas na Tabela 1.

O 1º eixo é formado com a contribuição de praticamente todas as variáveis, exceto "**Registro de Vacinação**", algumas categorias de pequeno efetivo e para duas categorias que caracterizam situações de desempenho regular (ver Tabela 1 — "contribuição absoluta"). Ou seja, este eixo é construído pelo conjunto de categorias escolhidas, refletindo, portanto, uma forte correlação entre quase todas as variáveis indicativas do desempenho das unidades na vacinação. Ainda nesta mesma listagem, observa-se que o 2º eixo é definido, principalmente, pelas variáveis "**Cálculo de Cobertura**" e "**Taxa de Abandono**" e, dentro destas, pelas categorias que indicam se é a própria unidade que calcula o indicador ou se ela o recebe calculado por outra instância.

Pela Figura 1, composto pelos eixos 1 e 2, alguns aspectos ficam muito evidentes. As categorias positivas em relação ao cumprimento

das atividades do PNI concentram-se à direita do eixo 1 (vertical), e o não cumprimento das atividades, à esquerda. Próximo à origem dos eixos situam-se as que sugerem desempenho regular, como, por exemplo, geladeira com algo errado (GELM) ou apenas uma estratégia de vacinação além da demanda espontânea (ETT1).

Também as variáveis ilustrativas acompanham este padrão, concentrando-se à direita no gráfico as categorias indicativas de situação boa do serviço, como a existência de supervisão, de boletim epidemiológico, de egresso de curso especial de treinamento em atividade de vigilância epidemiológica, o CBVE. Nas variáveis que apresentam um gradiente entre os extremos, como é o caso de "**Normas do PNI**", "**Relação Comunitária**" e "**Dados Demográficos**", entre os fatores determinantes, e "**Consolidado Informante**" e "**Envio de Notificação**", todas ilustrativas, as categorias intermediárias situam-se próximo à origem, desenhando, portanto, uma curva de desempenho que vai da situação mais precária, à esquerda do gráfico, para a melhor, à direita.

A variável determinante "**Instituição**" apresenta correlação significativa somente com o eixo 1 (ver Tabela 2 — "valores teste"). As unidades das Secretarias Estaduais de Saúde concentram-se à direita e as demais, à esquerda, sugerindo que o desempenho das unidades das Secretarias Municipais de Saúde, do Inamps e hospitalares é bastante precário nas atividades do PNI. Também a variável ilustrativa "**Escolaridade do Entrevistado**" mostra correlação significativa com este eixo, embora pequena.

O 2º eixo é composto, principalmente, pelas categorias que indicam a capacidade das unidades realizarem, de forma autônoma, o cálculo de alguns indicadores de avaliação das atividades de imunização. Ficam graficamente bem separadas as unidades que calculam a cobertura vacinal e a taxa de abandono do programa das que recebem estes indicadores já calculados.

Também as variáveis ilustrativas relativas à autonomia, como o cálculo de doses de vacina necessárias às atividades, a disponibilidade de dados demográficos na unidade, a relação estabelecida com a comunidade, são discriminadas por este eixo. Deve-se notar que a presença de egresso do CBVE também se destaca neste eixo.

**FIGURA 1. Gráfico de Barras da Distribuição Percentual da Variância Representada por Cada Auto-Valor**

Auto-Valor	Porcentagem	Percentual Acumulado	
1	0.22448260	14.43	*****
2	0.16511536	10.51	*****
3	0.14249939	9.16	*****
4	0.12097335	7.78	*****
5	0.11668879	7.50	*****
6	0.11305845	7.27	*****
7	0.10767037	6.92	*****
8	0.09963369	6.41	*****
9	0.09570169	6.15	*****
10	0.09165871	5.89	*****
11	0.08600187	5.53	*****
12	0.07539493	4.35	*****
13	0.06964457	4.48	*****
14	0.04701521	3.02	*****

**FIGURA 2. Categorias das Variáveis de Avaliação do PNI — Análise de Correspondência**

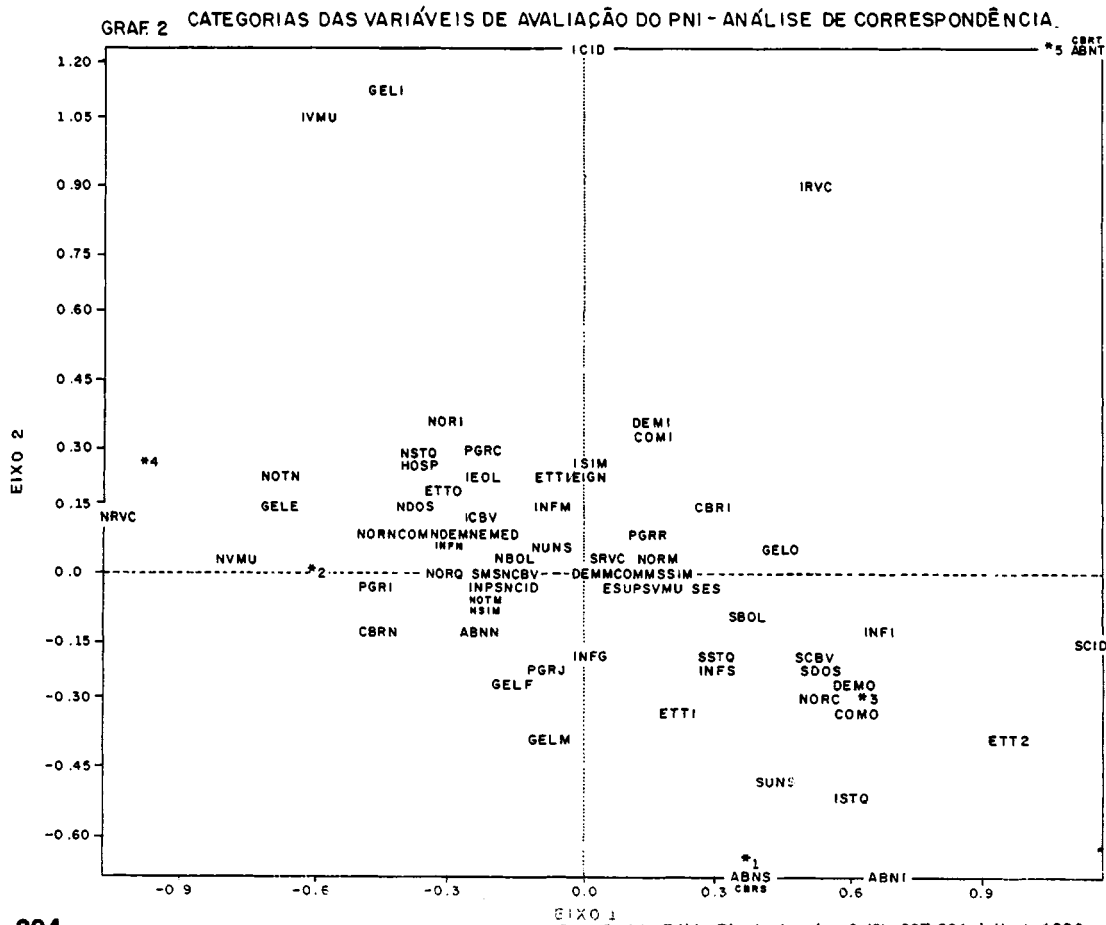




TABELA 2. Número e Valor-Teste nos Dois Eixos de Maior Auto-Valor Para as Variáveis Ilustrativas

Variáveis Ilustrativas	Símbolo - Categoria	Número	Valor teste	
			Eixo 1	Eixo 2
Instituição	SES - Estado	314	6.32	- 0.86
	SMS - Município	226	- 3.61	0.23
	INPS - Inamps	77	- 2.48	- 0.28
	HOSP - Hospital	43	- 2.62	1.67
Escolaridade Entrevistado	ESUP - Nível Superior	474	2.60	- 1.76
	EMED - Nível Médio	154	- 2.72	1.19
	EIGN - Ignorado	32	- 0.09	1.34
Dados Demográficos	DEMO - População e Área	132	7.63	- 3.09
	DEMM - População ou Área	236	- 0.03	0.33
	DEMN - Nenhum	291	- 6.13	2.15
	DEMI - Ignorado	1	0.15	0.32
Relação Comunitária	COMO - Variada	88	6.35	- 3.13
	COMM - Alguma Atividade Comunitária	314	2.67	0.01
	COMN - Nenhuma Atividade Comunitária	257	- 7.17	2.15
	COMI - Ignorado	1	0.15	0.32
Notificantes	INFM - Somente Funcionários	282	- 2.34	3.68
	INFS - Outras Unidades	163	4.75	- 3.20
	INFG - Escolas, Creches	113	- 0.41	- 2.07
	INFN - Ninguém	92	- 2.94	1.12
	INFI - Ignorado	10	2.26	- 0.41
Envio de Notificação	NOTS - Envia Notif. Corretamente	410	5.86	- 0.22
	NOTM - Envia Notif. Irregularmente	194	- 3.77	- 0.69
	NOTN - Não Envia Notificação	47	- 4.65	1.57
	NOTI - Ignorado	9	0.58	0.14
Programação	PGRU - Unidade Programa	244	- 1.12	- 4.34
	PGRC - Centro de Saúde	67	- 2.26	2.42
	PGRR - Regional ou Central	316	3.72	2.77
	PGRI - Não Sabe Informar	33	- 2.89	- 0.09
Unidade Subordinada	SUNS - Tem Unidade Subordinada	68	3.93	- 4.12
	NJNS - Não Tem Subordinada	587	- 3.80	3.85
	IJNS - Ignorado	5	- 0.06	0.51
CBVE	SCBV - Algum CBVE	182	8.62	- 3.09
	NCBV - Nenhum CBVE	200	- 2.31	2.76
	ICBV - Sem Informação	278	- 5.65	- 0.06
Supervisão	SSIM - Sup. em Imunizações	356	4.25	- 0.46
	NSIM - Não Sup. Imun.	278	- 4.22	1.31
	ISIM - Ignorado	26	- 0.18	0.14
Presença de Normas	NORQ - Apenas Quadro Vacinação	139	- 4.05	1.09
	NORM - Esquema e Conservação	297	3.02	- 3.06
	NORC - Norma Vac. Completa	104	5.89	1.04
	NORN - Não Norma Vacinação	111	- 5.06	1.11
	NORI - Ignorado	9	- 0.92	- 1.46
Recebe Boletim	SBOL - Tem Boletim	212	6.66	1.34
	NBOL - Não Tem Boletim	442	- 6.51	0.57
	IBOL - Ignorado	6	- 0.49	- 13.30
Classes Criadas	*1 - Grupo 1 PNI	44	7.51	- 0.15
	*2 - Grupo 2 PNI	333	- 15.45	- 4.07
	*3 - Grupo 3 PNI	166	9.10	2.02
	*4 - Grupo 4 PNI	53	- 7.33	15.57
	*5 - Grupo 5 PNI	64	13.17	

**FIGURA 3. Classificação Hierárquica Ascendente — Descrição dos 45 Nós da Hierarquia de Índices Mais Elevados**

Nó	Maior	Menor	Nº	Peso	Índice	Barra dos Índices de Nível
1275	1228	1222	31	31	0.0018215	**
1276	1229	1200	17	17	0.0018598	**
1277	1220	1254	15	15	0.0019780	**
1278	1240	1267	61	61	0.0020125	**
1279	1278	1195	79	79	0.0020786	**
1280	1077	979	58	58	0.0023079	**
1281	1250	1180	10	10	0.0023653	**
1282	1258	1217	9	9	0.0024734	**
1283	206	1265	5	5	0.0027013	**
1284	1270	1238	34	34	0.0027703	**
1285	1269	1207	13	13	0.0027775	**
1286	1285	1242	17	17	0.0028281	**
1287	1263	1066	70	70	0.0028363	**
1288	1231	1276	27	27	0.0028577	**
1289	1280	1212	80	80	0.0028825	**
1290	1273	1247	89	89	0.0029157	**
1291	1241	1266	27	27	0.0030160	**
1292	1275	1230	52	52	0.0031512	***
1293	1257	1227	30	30	0.0031947	***
1294	1251	1264	19	19	0.0035601	***
1295	1272	1261	31	31	0.0036866	***
1296	1249	1210	14	14	0.0038236	***
1297	1262	1259	7	7	0.0039950	***
1298	1284	1296	48	48	0.0047174	***
1299	1237	1288	52	52	0.0048358	****
1300	1277	1271	36	36	0.0049748	****
1301	1288	1274	24	24	0.0050657	****
1302	1295	1255	36	36	0.0057462	****
1303	1297	1282	16	16	0.0067026	*****
1304	1292	1287	122	122	0.0076152	*****
1305	1260	1301	49	49	0.0081330	*****
1306	1279	1291	106	106	0.0104983	*****
1307	1299	1300	88	88	0.0105351	*****
1308	1290	1294	58	58	0.0116568	*****
1309	1298	1286	65	65	0.0132647	*****
1310	1289	1293	110	110	0.0140815	*****
1311	1307	1306	194	194	0.0219916	*****
1312	1304	1310	232	232	0.0243564	*****
1313	1281	1305	59	59	0.0316214	*****
1314	1311	1309	259	259	0.0403949	*****
1315	1302	1303	52	52	0.0530046	*****
1316	1314	1313	318	318	0.0683513	*****
1317	1312	1316	550	550	0.0742019	*****
1318	1315	1317	602	602	0.0889665	*****
1319	1308	1318	660	660	0.1258790	*****

Soma dos Índices do Níveis = 0.76935

A análise dos determinantes do desempenho do PNI, obtida invertendo-se os papéis "ativo" e "suplementar" das variáveis, corroborou, em linhas gerais, as observações anteriores.

Os resultados da classificação hierárquica ascendente são apresentados a seguir. O gráfico de barras dos índices dos níveis da classificação hierárquica ascendente não apresenta pontos de corte muito definidos indicando a inexistência de classes bem demarcadas (Figura 3). A distância entre duas partições é ligeiramente maior entre 5 e 6 classes, sendo então escolhida a partição em 5 (cinco).

Nas Tabelas 3 a 7 são apresentadas as principais características de cada classe ou grupo de unidades: variáveis e categorias que permitem a diferenciação dos grupos, percentual em que estas categorias estão presentes no universo e percentual em que estão presentes no grupo. Os centros destas classes podem ser visualizados no gráfico da análise de correspondência identificadas por um asterisco seguido de seu número de identificação (\*1, \*2, \*3, etc.). A numeração das classes é casual, não tendo significado hierárquico.

**TABELA 3.** Nº e Percentual de Unidades na Classe \*1 de Desempenho no PNI, Categorias que as Caracterizam e Percentual Destas Categorias no Universo Total e em Cada Classe, Brasil, 1985

Classe *1 — 44 Unidades — 6,7% do Total				
Variável	Categorias Características	% no Total	% na Classe	Critério
Cálculo Cobertura	Calcula	9,2	81,8	10,43
Normas	Completa	15,8	52,3	5,62
Dados Demográficos	Completo	20,0	54,5	5,17
Informante	Outras unidades	24,7	54,5	4,41
Estratégias	Mais de uma	21,7	50,0	4,30
Unidade Subordinada	Presente	10,3	31,8	4,02
CBVE	Presente	27,6	54,5	3,97
Geladeira	Parcialmente adequada	22,0	47,7	3,95
Dose p/ Cobertura	Correta	41,7	65,9	3,49
Registro Vacinas	Presente	93,5	00,0	3,47
Atividade Comunitária	Variada	13,3	31,8	3,36
Controle Estoque	Presente	57,1	77,3	3,07
Boletim Epidemiológico	Presente	32,1	52,3	2,99
Instituição	Estado	47,6	65,9	2,68
Escolarid. Entrev.	Superior	71,8	86,4	2,56

**FIGURA 4.** Nº e Percentual de Unidades na Classe \*2 de Desempenho no PNI, Categorias que as Caracterizam e Percentual Destas Categorias no Universo Total e em Cada Classe, Brasil, 1985

Classe *2 — 333 Unidades — 50,5% do Total				
Variável	Categorias Características	% no Total	% na Classe	Critério
Contra-indicações	Incorretas	85,5	99,7	12,74
Taxa de Abandono	Inexistente	85,5	99,1	11,68
Cálculo de Cobertura	Inexistente	69,2	88,6	11,16
Dose Cobertura	Incorreta	58,3	77,5	10,17
Registro de Vacinas	Presente	93,5	100,0	9,44
Estratégias	Só demanda	64,2	81,1	9,24
Geladeira	Inadequada	30,2	41,7	6,69
Normas	Apenas quadro das vacinas	21,1	28,2	4,69
Multivacinação	Incorreta	17,1	23,4	4,49
Dados Demográficos	Ausentes	44,1	52,0	4,19
Atividade Comunitária	Ausentes	38,9	46,2	3,97
Instituição	Município	34,2	41,1	3,86
Escolaridade Entrev.	Médio/Elementar	23,3	29,4	3,85
Boletim Epidemiológico	Ausente	67,0	73,6	3,73
CBVE	Sem informação	42,1	48,9	3,67

**FIGURA 5.** Nº e Percentual de Unidades na Classe \*3 de Desempenho no PNI, Categorias que as Caracterizam e Percentual Destas Categorias no Universo Total e em Cada Classe, Brasil, 1985

Classe *3 — 166 Unidades — 25,2% do Total				
Variável	Categorias Características	% no Total	% na Classe	Critério
Geladeira	Adequada	46,4	80,7	10,49
Dose Cobertura	Correta	41,7	73,5	9,59
Multivacinação	Correta	81,5	98,2	8,32
Contra-indicações	Corretas	12,0	31,9	8,18
Estratégias	Mais de duas	13,5	34,3	8,17
Taxa de Abandono	Inexistente	85,5	95,8	5,20
Registro Vacinas	Presente	93,5	99,4	4,93
CBVE	Presente	27,6	41,6	4,53
Controle Estoque	Presente	57,1	67,5	3,24
Notificação	Correto	62,1	71,7	3,08
Instituição	Estado	47,6	56,6	2,79
Supervisão	Presente	53,9	62,0	2,52
Atividade Comunitária	Ausente	47,6	55,4	2,43
Normas	Esquema e conservação	45,0	51,8	2,12
Escolaridade Entrev.	Superior	71,8	77,7	2,08

Começando pela classe "5", esta pode ser caracterizada, em linhas gerais, como um grupo onde o desempenho das unidades é o preconizado nas normas. A principal característica da classe, entretanto, conforme pode ser visto pelo alto valor do indicador "Critério", é a dependência em relação aos níveis regional e/ou central para o cálculo dos indicadores do programa e para a programação de doses de vacina necessárias.

O centro da classe encontra-se no mesmo ponto que a categoria indicativa do cálculo de cobertura realizado fora da unidade CBRT (Figura 2). No lado oposto em relação ao eixo 2 encontra-se o grupo "1", que, mantendo, no geral, o correto desempenho, caracteriza-se, fundamentalmente por calcular sua cobertura vacinal. Coerente com esta característica é a disponibilidade de dados de população e área, obviamente necessários para permitir o cálculo de cobertura, o que também aparece como característica importante do grupo. Gráficamente, há uma superposição das categorias indicativas da realização do cálculo de cobertura vaci-

nal e taxa de abandono do programa com o centro da classe "1".

Ainda com respostas corretas em relação ao desempenho nas atividades do PNI encontra-se o grupo "3", também com predomínio relativo de unidades das Secretarias Estaduais de Saúde. Neste grupo, os indicadores de cobertura vacinal e taxa de abandono do programa não estão disponíveis na unidade.

Estas três classes, "1", "3" e "5", localizam-se do lado direito do gráfico, junto com as categorias indicativas de desempenho adequado.

O grupo "4" reúne as unidades com desempenho inteiramente fora das normas do programa, inclusive em relação a questões básicas, como o controle de estoque e o registro adequado de crianças vacinadas, sem o qual a orientação adotada no programa é não repassar o imunizante.

Também o grupo "2" agrupa unidades com desempenho ruim, especialmente em relação às atividades que exigem um grau maior de conhecimento das normas e recursos humanos capacitados. Esta classe congrega 333 unidades, metade dos 660 serviços avaliados.

**TABELA 6.** Nº e Percentual de Unidades na Classe 4 de Desempenho no PNI, Categorias que as Caracterizam e Percentual Destas Categorias no Universo Total e em Cada Classe, Brasil, 1985

Classe 4 — 53 Unidades — 8,0% do Total				
Variável	Categorias Características	% no Total	% na Classe	Critério
Registro Vacinas	Inexistente	6,4	75,5	11,88
Contra-indicações	Ignorado	2,6	24,5	6,13
Respons. Programação	Não sabe informar	5,0	20,8	4,16
Controle Estoque	Inexistente	41,4	66,0	3,89
Normas	Inexistente	16,8	35,8	3,56
Geladeira	Inadequada	30,2	50,9	3,40
Dose Cobertura	Incorreta	58,3	77,4	3,20
Multivacinação	Incorreta	17,1	32,1	2,90
Instituição	Hospital	6,5	17,0	2,89
Atividade Comunitária	Nenhuma	38,9	56,6	2,84
Supervisão	Inexistente	42,1	56,6	2,35
Taxa de Abandono	Inexistente	85,5	94,3	2,33
Cálculo Cobertura	Inexistente	69,2	81,1	2,19
Notificação	Inexistente	7,1	13,2	1,91
Respons. Programação	Centro de Saúde	10,2	17,0	1,84

**TABELA 7.** Nº e Percentual de Unidades na Classe \*5 de Desempenho no PNI, Categorias que as Caracterizam e Percentual Destas Categorias no Universo Total e em Cada Classe, Brasil, 1985

Classe *5 — 64 Unidades — 9,7% do Total				
Variável	Categorias Características	% no Total	% na Classe	Critério
Taxa de Abandono	Recebe calculado	11,2	96,9	14,58
Cálculo Cobertura	Recebe calculado	20,2	100,0	14,43
Respons. Programação	Nível regional ou central	47,9	68,8	3,68
Geladeira	Adequada	46,4	65,6	3,39
Dose Cobertura	Correta	41,7	57,8	2,85
Contra-indicações	Corretas	12,0	23,4	2,83
Normas	Esquema e conservação	45,0	60,9	2,82
Boletim Epidemiológico	Presente	32,1	46,9	2,71
Notificação	Correta	62,1	75,0	2,43
Dados Demográficos	Completo	20,0	31,3	2,39
Multivacinação	Correta	81,5	90,6	2,29
Registro Vacinas	Presente	93,5	98,4	2,25
Supervisão	Presente	53,9	65,6	2,12
Estratégias	Mais de uma	21,7	31,3	2,04
Instituição	Estado	47,6	57,3	1,85

## CONCLUSÕES

A utilização da análise de correspondência permitiu que os dados da avaliação do PNI fossem abordados de forma a extrair o máximo possível de informações, a partir dos elementos disponíveis. Mais ainda, evitou algumas interpretações equivocadas a que seríamos induzidos pela análise de cada variável independente, separada das demais. O processo de classificação acrescenta ainda um aspecto interessante: a possibilidade de definir estratégias de atuação baseadas nos agrupamentos observados.

O exemplo apresentado neste trabalho limitou a análise a variáveis categóricas, embora seja possível analisar variáveis contínuas simultaneamente. Além disso, as técnicas descritas estão disponíveis para utilização em microcomputadores.

Os procedimentos empregados poderiam ser seguidos de outros métodos estatísticos de análise onde fossem testadas as associações encontradas. A análise de correspondência em epidemiologia pode servir como um auxiliar

poderoso na primeira inspeção de grandes volumes de dados, que após seleção cuidadosa seriam então submetidos a outras técnicas de maior precisão.

## RESUMO

CARVALHO, M. S. & STRUCHINER, C. J. **Análise de Correspondência: Uma Aplicação do Método à Avaliação de Serviços de Vacinação.** Cad. Saúde Públ., Rio de Janeiro, 8 (3): 287-301, jul/set, 1992.

A análise de correspondência é um método estatístico desenvolvido para a análise descritiva exploratória de grandes tabelas. Através desta técnica, pode-se visualizar, graficamente, as relações mais importantes de um grande conjunto de variáveis categóricas e contínuas. Neste trabalho, aplicou-se a técnica descrita, seguida de um processo de classificação hierárquica ascendente, disponíveis no programa "Système Portable Pour L'Analyse de Données — SPAD", aos

dados obtidos a partir de uma avaliação de desempenho do Programa Nacional de Imunizações realizada em 660 unidades públicas de saúde, distribuídas em 98 municípios brasileiros. Foram incluídas na análise diversas variáveis relacionadas ao cumprimento das normas do PNI e outras "determinantes" potenciais do desempenho das unidades na atividade.

Os serviços avaliados não apresentaram padrão uniforme de desempenho, sendo classificados em cinco grupos, segundo sua atuação no PNI. Desrespeito às normas mais elementares foi observado em 58% das unidades.

A análise de correspondência em epidemiologia pode servir como um auxiliar poderoso na primeira inspeção de grandes volumes de dados, que após seleção cuidadosa seriam então submetidos a outras técnicas de maior precisão.

**Palavras-Chave:** Análise de Correspondência; Avaliação; Vacinação

## REFERÊNCIAS BIBLIOGRÁFICAS

- BENZECRI, J. P., 1979. Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers de L'Analyse des Données*, 3: 377-384.
- CARVALHO, M. S., 1990. *Avaliação da Prática de Vigilância Epidemiológica nos Serviços Públicos de Saúde no Brasil*. Tese de Mestrado, Rio de Janeiro: Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz.
- CESIA (Centre de Statistique et D'Informatique Appliquées), 1982. *Systeme portable pour l'analyse des données (SPAD); notices d'utilization*. Paris. (Mimeo)
- \_\_\_\_\_, 1985. *Systeme portable pour l'analyse des données (SPAD)*. Paris. (Mimeo)
- ENSP (Escola Nacional de Saúde Pública), 1986. *Avaliação do impacto da estratégia de treinamento da SNABS/MS em controle de doenças transmissíveis; Relatório Final*. Rio de Janeiro. (Mimeo)
- GOUVÊA, V., 1990. *Análise de correspondências (Textos Didáticos)*, Rio de Janeiro: ENCE/IBGE. (Mimeo)
- GREENACRE, M. J., 1981. Practical correspondence analysis. In: *Looking at Multivariate Data*, Cap. III, New York: J. Wiley & Sons.
- LEBART, L.; MORINEAU, A. & WARWICK, K. M., 1984. *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: J. Wiley & Sons.
- LEBART, L.; MORINEAU, A. & FÉNELON, J. P., 1977. *Traitement des Données Statistiques; Méthodes et Programmes*. 2ª ed. Paris: Dunod.
- MORINEAU, A., 1984. Note sur la caracterisation statistique d'une classe et les valeurs-tests. *Bulletin Technique CESIA*, 2: 20-27.
- SES-BA (Secretaria de Saúde da Bahia), 1983. *Informes sobre a Avaliação do Programa de Imunizações do Estado da Bahia*. (Mimeo)
- SESH-RJ (Secretaria de Estado de Saúde e Higiene do Rio de Janeiro), 1983. *Primeira Avaliação Operacional Multidisciplinar do Programa de Imunizações do Estado do Rio de Janeiro*. Rio de Janeiro. (Mimeo)