

ANÁLISE MULTIVARIADA DA DISPONIBILIDADE *PER CAPITA* DE NUTRIENTES EM 44 PAÍSES¹

PAULO ROBERTO CURI², LUCIA TERADA³,
PATRICK JOHANNES BECKERS e ALEXANDRE ALVES⁴

RESUMO - Foi estudada a situação de países avaliados pela disponibilidade de nutrientes de origem animal e vegetal (dados publicados pela FAO, Anuário da Produção agropecuária 1988), usando-se os métodos estatísticos multivariados de Análise de Componentes Principais e Análise de Agrupamento. O estudo foi efetuado com 44 países cujo critério de escolha foi o de incluir países com os mais variados graus de desenvolvimento e de todos os continentes. O primeiro componente principal foi interpretado como exprimido a disponibilidade de nutrientes de origem animal e o segundo componente principal, a disponibilidade de nutrientes de origem vegetal, o que permitiu a ordenação dos países para cada um destes componentes. As variáveis mais discriminatórias foram: proteína, caloria, riboflavina, lipídio e cálcio, todas de origem animal. As menos discriminatórias foram: niacina, beta caroteno, equivalente retinol e caloria, todas de origem vegetal.

Termos para indexação: Análises de Componentes Principais e de Agrupamento, nutrição, proteínas.

MULTIVARIATE ANALYSIS OF *PER CAPITA* AVAILABILITY OF NUTRIENTS IN 44 COUNTRIES

ABSTRACT - The situation of the countries is studied based on the availability of the animal and vegetal nutrients (data published by FAO in 1988). Two multivariate statistical methods were used: Principal Components Analysis and Cluster Analysis. The study was applied to countries of all the continents, with 44 countries of the most different degrees of development. The first principal component was interpreted as a measure of the available animal nutrients, and the second principal component as a vegetal nutrient, which made it possible the ordering of the countries. The most discriminating variables were the protein, calorie, riboflavin, fat and calcium of animal origin and the less ones were niacin, beta carotene, equivalent retinol and calorie of vegetal origin.

Index terms: Principal Components Analysis, Cluster Analysis, nutrition, proteins.

INTRODUÇÃO

A publicação de dados numéricos na forma de Tabelas, muitas vezes não possibilita ao interessado uma melhor interpretação das informações fornecidas, principalmente nos casos em que as tabelas se avolumam por conterem

um grande número de unidades das quais são fornecidas muitas características.

O Anuário de Produção Agropecuária da FAO publicou em 1988 a disponibilidade de 22 nutrientes de origem animal e vegetal de países de todo o mundo. O presente trabalho ilustra a utilização de dois métodos de análise multivariada para interpretação destes dados, visando proporcionar melhor aproveitamento e mais fácil visualização dos mesmos, já que os resultados finais obtidos podem ser apresentados em forma de figuras de fácil entendimento. Os métodos são a Análise de Componentes Principais e a Análise de Agrupamento (Sneath & Sokal 1973), utilizados com a finalidade de discutir a

¹ Aceito para publicação em 14 de janeiro de 1992.

² Méd.-Vet., Prof.-Titular, Dep. de Melhoramento e Nutrição Animal e do Serviço de Estat. e Computação da FMVZ - UNESP. CEP 18610 Botucatu, SP, Rubião Jr.

³ No curso de Zoot. da FMVZ e bolsista do CNPq.

⁴ No curso de Zoot. da FMVZ e bolsista da FAPESP.

importância de cada um dos nutrientes para a discriminação dos países, pela construção de combinações lineares das 22 variáveis originais de forma a reduzir o espaço das variáveis e possibilitar a localização dos países em figuras. Os resultados das análises permitem o agrupamento de países, com a propriedade de que países pertencentes a um mesmo grupo apresentam maior similaridade (ou homogeneidade) que países de grupos diferentes.

Não se pretende entrar em maiores considerações e explicações sobre a natureza dos resultados, mas sim, discutir a adequação das técnicas multivariadas para a análise deste tipo de dados.

MATERIAL E MÉTODOS

A FAO publica anualmente tabelas contendo a quantidade de nutrientes vegetais e animais *per capita* para países de todo o mundo. Foram estudados 22 nutrientes que refletem a disponibilidade de alimentos, isto é, o que o país produz, *per capita*, e não necessariamente o que é consumido pelos seus habitantes (os dados utilizados não foram aqui transcritos).

Para maior facilidade de interpretação, o estudo foi realizado para 44 países escolhidos de forma a incluir países dos mais variados graus de desenvolvimento e de todos os continentes. A relação dos países, bem como seus números de identificação podem ser vistos na apresentação dos resultados (Tabela 1).

A seguir, são listados os nutrientes com suas respectivas siglas de identificação, com o número 1 final indicando origem vegetal e o número 2 a origem animal: caloria (CAL 1, CAL 2), proteína (PROT 1, PROT 2), lipídio (LIP 1, LIP 2), cálcio (CA 1, CA 2), ferro (FE 1, FE 2), retinol total (RET), beta caroteno (BCA 1, BCA 2), equivalente retinol (ERE 1, ERE 2), tiamina (TIA 1, TIA 2), riboflavina (RIB 1, RIB 2), niacina (NIA 1, NIA 2), ácido ascórbico total (AASC).

A Tabela inicial de dados corresponde a uma matriz com 44 linhas (países) e 22 colunas (nutrientes). As variáveis foram transformadas de modo a torná-las adimensionais e com variação no intervalo de 0 a 1. Para tanto, o valor de uma certa variável para um determinado país foi dividido pelo maior valor da mesma variável no conjunto de países estudados.

A análise estatística foi efetuada utilizando-se os métodos multivariados de Análise de Componentes Principais e de Análise de Agrupamento (Sneath &

TABELA 1. Valores dos três primeiros componentes principais (Y1, Y2, Y3), para os países estudados. Ordenação crescente dos países para Y1 e para Y2 (números entre parênteses).

Nº	País	Y1	Y2	Y3
1.	Camarões	-0,172 (5)	0,081 (41)	-0,107
2.	Egito	-0,126 (9)	0,096 (42)	0,043
3.	Líbia	0,003 (23)	0,068 (39)	0,023
4.	Nigéria	-0,181 (2)	0,059 (36)	-0,088
5.	África do Sul	-0,049 (18)	-0,056 (9)	0,031
6.	Uganda	-0,165 (6)	0,072 (40)	0,018
7.	Canadá	0,096 (33)	0,006 (27)	-0,012
8.	Costa Rica	-0,040 (20)	0,004 (26)	-0,125
9.	Cuba	0,005 (24)	-0,057 (8)	0,024
10.	Haiti	-0,173 (3)	0,010 (30)	-0,027
11.	Honduras	-0,113 (10)	-0,069 (5)	-0,004
12.	México	-0,042 (19)	0,008 (29)	0,046
13.	Estados Unidos	0,137 (39)	-0,005 (24)	0,013
14.	Argentina	0,117 (34)	-0,044 (11)	-0,009
15.	Bolívia	-0,103 (13)	-0,062 (7)	0,023
16.	Brasil	-0,089 (14)	-0,063 (6)	0,036
17.	Chile	-0,083 (15)	-0,033 (12)	0,031
18.	Colômbia	-0,073 (16)	-0,013 (21)	-0,064
19.	Paraguai	-0,025 (21)	0,050 (33)	0,054
20.	Peru	-0,108 (11)	-0,075 (4)	0,012
21.	Uruguai	0,062 (27)	-0,104 (2)	-0,005
22.	Venezuela	-0,018 (22)	-0,088 (3)	0,022
23.	Bangladesh	-0,184 (1)	-0,120 (1)	-0,019
24.	China	-0,152 (7)	-0,022 (15)	-0,006
25.	Índia	-0,173 (4)	-0,010 (19)	0,006
26.	Israel	0,082 (30)	0,060 (37)	-0,020
27.	Japão	0,009 (26)	-0,017 (18)	-0,020
28.	Paquistão	-0,135 (8)	-0,017 (17)	0,011
29.	Síria	-0,062 (17)	0,096 (43)	0,067
30.	Turquia	-0,104 (12)	0,051 (34)	0,052
31.	Áustria	0,149 (40)	-0,019 (16)	0,012
32.	Bulgária	0,071 (28)	0,024 (31)	0,053
33.	Tchecoslováquia	0,135 (38)	-0,007 (22)	-0,010
34.	França	0,220 (43)	0,006 (28)	-0,044
35.	Grécia	0,118 (35)	0,133 (44)	-0,008
36.	Itália	0,094 (31)	0,060 (38)	0,014
37.	Holanda	0,203 (42)	-0,008 (20)	-0,007
38.	Portugal	0,009 (25)	0,052 (35)	-0,016
39.	Espanha	0,095 (32)	0,041 (32)	0,024
40.	Suécia	0,180 (41)	-0,051 (10)	-0,020
41.	Reino Unido	0,129 (37)	-0,025 (14)	-0,009
42.	União Soviética	0,079 (29)	-0,006 (23)	0,019
43.	Austrália	0,128 (36)	0,000 (25)	0,010
44.	Nova Zelândia	0,226 (44)	-0,026 (13)	-0,045

Sokal 1973), com o uso do programa NTSYS (Rohlf 1988).

A Análise de Componentes Principais (ACP) foi empregada para reduzir o conjunto das 22 variáveis originais a novas variáveis não correlacionadas que são os componentes principais. Cada componente (Y) é uma combinação linear das variáveis originais (X), com a capacidade de reter a maior quantidade da informação fornecida pelas variáveis originais. Cada país, que era definido pelo conjunto de 22 variáveis passou, então, a ser definido pelos componentes principais. No caso estudado foram utilizados os 3 primeiros componentes (Y_1, Y_2, Y_3). Os componentes principais padronizados de cada país foram calculados com base na matriz original de dados de forma a serem centrados na origem. Em notação matricial, a matriz Y (44 x 3) de componentes principais, é dada por:

$Y = (I - (1/n)E) X A$, onde I é a matriz identidade, X (44 x 22) é a matriz de dados originais, E (44 x 44) é uma matriz com números 1 em cada posição, A (22 x 3) é uma matriz cujas colunas são os três primeiros vetores característicos da matriz de covariância dos dados e n = 44 é o número de unidades (países).

A ACP possibilitou verificar a capacidade discriminatória das variáveis originais no processo de formação dos agrupamentos (Morrison 1967). A contribuição de cada variável para certo componente principal foi avaliada pelo coeficiente de correlação linear ($r_{X_j Y_k}$) entre a variável (X_j ; j = 1, ..., 22) e o componente principal (Y_k ; k = 1, 2, 3), sendo $r_{X_j Y_k} = a_{jk} * l_k^{1/2} / s_j$, onde a_{jk} é o coeficiente da variável X_j no componente Y_k , l_k é a k-ésima maior raiz característica da matriz de covariância e s_j é o desvio-padrão da variável X_j . Portanto, quanto maior for o valor absoluto do coeficiente de correlação ($r_{X_j Y_k}$) maior será a contribuição da variável X_j para a formação do componente principal (Y_k).

Por outro lado, o estudo do significado prático dos componentes principais permite considerações de interesse na interpretação dos resultados (Sneath & Sokal 1973, Jolicoeur & Mosimann 1960). Este significado depende da intensidade da contribuição de cada variável ao componente principal, traduzido pelo valor de $r_{X_j Y_k}$.

A adequação da ACP foi verificada pela quantidade da informação das variáveis originais retida pelos três componentes principais (percentagem da variância total das 22 variáveis acumulada pelos três primeiros componentes).

A Análise de Agrupamentos (AA) foi realizada com os dados transformados, calculando-se a Distância Euclidiana Média entre os países para o conjunto

das 22 variáveis e utilizando o algoritmo UPGMA (unweighted pair group method with arithmetic average), conforme recomendado por Rohlf (1970). O resultado final da análise foi representado na forma de um dendrograma. A utilização conjunta do dendrograma da AA e da ordenação dos países obtido na ACP permite inferir, com maior segurança, os agrupamentos formados (Curi 1983). Os agrupamentos devem ter como características a maior homogeneidade (ou similaridade) entre países do mesmo grupo e a heterogeneidade entre países de grupos diferentes.

RESULTADOS E CONCLUSÕES

Como os coeficientes de correlação entre os nutrientes de origem animal e o primeiro componente foram sempre próximos de 1, enquanto os mesmos coeficientes envolvendo os nutrientes vegetais resultaram pequenos, o primeiro componente principal pode ser entendido como exprimindo principalmente a contribuição dos nutrientes de origem animal. Países com valores grandes de Y_1 possuem maior disponibilidade *per capita* de nutrientes de origem animal, o contrário ocorrendo com países com os menores valores de Y_1 .

A Tabela 1 apresenta a relação dos países estudados e os respectivos valores dos três primeiros componentes principais.

A Tabela 2 apresenta os coeficientes de correlação entre as variáveis originais e os três primeiros componentes principais, responsáveis por cerca de 83% da informação contida no conjunto das 22 variáveis originais.

Ordenando-se os valores de Y_1 , do menor para o maior, resultará a correspondente ordenação crescente dos países para a disponibilidade *per capita* de nutrientes de origem animal (valores entre parênteses da Tabela 2). Assim procedendo, destacaram-se Nova Zelândia, França, Holanda, Suécia, Áustria, Estados Unidos, Tchecoslováquia, Reino Unido e Austrália como os países com maior disponibilidade, e Bangladesh, Nigéria, Índia, Camarões, Haiti, Uganda, China, Paquistão e Egito como os de menor disponibilidade. A situação do Brasil resultou bastante incômoda, colocando-se junto com Bolívia, Honduras e Peru, mais próximo ao

TABELA 2. Coeficiente de correlação ($r_{X_jY_k}$) entre as variáveis originais (X) e os três primeiros componentes principais (Y1, Y2, Y3). Percentagem da informação (% Var. e % Var. acum.) retida pelos componentes.

Componente principal (Y)	Y1	Y2	Y3
Var. originais (X)			
CAL1	0,06	0,63	0,35
CAL2	0,98	-0,05	0,01
PROT1	-0,27	0,64	0,52
PROT2	0,99	-0,01	-0,01
LIP1	0,45	0,64	0,13
LIP2	0,97	-0,04	0,02
CA1	-0,18	0,83	0,11
CA2	0,94	-0,05	-0,02
FE1	-0,44	0,79	0,11
FE1	0,88	0,00	0,00
RET	0,85	0,09	-0,03
BCA1	0,00	0,56	-0,82
BCA2	0,91	0,05	0,00
ERE1	0,00	0,57	-0,82
ERE2	0,87	0,09	-0,03
TIA1	-0,42	0,66	0,40
TIA2	0,94	0,04	0,03
RIB1	-0,03	0,89	0,26
RIB2	0,98	-0,03	0,02
NIA1	0,06	0,45	0,24
NIA2	0,81	-0,08	0,00
AASC	0,36	0,72	0,25
% Var.	63,27	13,38	6,76
% Var. acum.	63,27	76,65	83,41

bloco de países com menor disponibilidade de nutrientes, e superior, somente, a 13 dos 44 países estudados.

Para o segundo componente (Y2) as maiores contribuições resultaram das variáveis de origem vegetal, em especial, da riboflavina, do ferro e do cálcio (com coeficientes de correlação variando de 0,79 a 0,89). As variáveis de origem animal pouco contribuíram para Y2. Portanto, países com grandes valores de Y2 apresentam boa disponibilidade de nutrientes de origem vegetal, entre os quais destacaram-se: Grécia, Sí-

ria, Egito, Camarões, Uganda e Líbia. Os menores valores de Y2 foram apresentados por Bangladexe, Uruguai e Venezuela. Deve ser ressaltado que o Uruguai apresentou-se bem situado para alimentos de origem animal.

O terceiro componente (Y3), responsável por 6,76% da informação contida nas variáveis originais, foi de mais difícil interpretação. Para este componente contribuíram de forma mais acentuada quatro variáveis de origem vegetal: beta caroteno e tiamina (com sinal negativo) e proteína e caloria com sinal contrário.

Considerando-se os valores dos coeficientes de correlação das variáveis originais com os três primeiros componentes principais, as variáveis com maior poder discriminatório foram proteína, caloria, riboflavina e lipídio, todas de origem animal. As menos discriminatórias foram niacina, beta caroteno, equivalente retinol e caloria, todas de origem vegetal.

Baseando-se nos valores dos 3 componentes principais (Tabela 2) seria possível construir uma figura tridimensional com a localização especial dos 44 países estudados. Esta figura não foi apresentada pois optou-se por colocar o dendrograma (Fig. 1) obtido na AA.

A interpretação da Fig. 1, apoiada pelos dados apresentados na Tabela 1, possibilitou definir os agrupamentos de países descritos a seguir:

G1: Camarões (1) e Nigéria (4), caracterizados por pequena disponibilidade *per capita* de nutrientes de origem animal (Y1 pequenos);

G2: Haiti (10), Índia (25), Paquistão (28), China (24), com Bangladexe (23) unindo-se a este grupo em nível inferior de similaridade;

G3: Egito (2), Turquia (30), Síria (29), Uganda (6) pode ser incluído neste grupo, embora de forma menos nítida;

G4: Honduras (11), Bolívia (15), Peru (20), Brasil (16) e Chile (17) constituíram um subgrupo de G4 (grupo mais amplo) que incluiu outro subgrupo formado por África do Sul (5), Cuba (9) e Venezuela (22). Menos caracterizado neste grupo, situou-se ainda, o México (12);

G5: Costa Rica (8) e Colômbia (18) formaram um grupo que, pela Fig. 1, parece estar

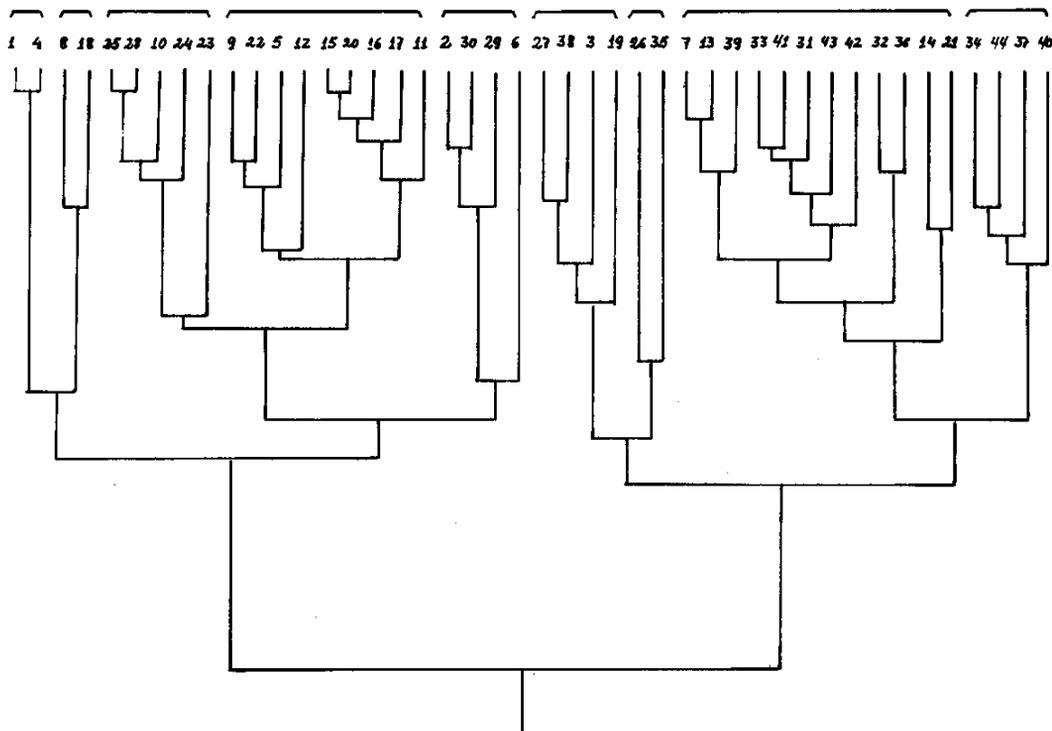


FIG. 1. Dendrograma resultante da Análise de Agrupamento, utilizando-se Distância Euclidiana Média entre os países para o conjunto das 22 variáveis.

mais próximo do grupo formado por Camarões (1) e Nigéria (4) o que não se confirmou com a ordenação tridimensional dos países, onde Costa Rica e Colômbia apareceram próximos a G4;

G6: Líbia (3), Japão (27), Portugal (38) e Paraguai (19) constituíram um grupo menos homogêneo, o mesmo ocorrendo com o grupo G7 formado por Israel (26) e Grécia (35);

Os dois grupos seguintes foram constituídos por países que apresentaram as maiores disponibilidades de nutrientes de origem animal:

G8: Canadá (7), Estados Unidos (13), Espanha (39), Áustria (31), Tchecoslováquia (33), Reino Unido (41), Austrália (43), União Soviética (42), Bulgária (32), Itália (36), Argentina (14) e Uruguai (21);

G9: França (34), Nova Zelândia (44), Holanda (37) e Suécia (40).

Pela Fig. 1, ainda, foi possível constatar a separação dos 44 países em dois grandes conglomerados: aquele formado pelos grupos G1, G2, G3, G4 e G5, caracterizado por países com menor disponibilidade *per capita* de nutrientes de origem animal, e aquele formado pelos demais grupos, englobando os países com maior disponibilidade de nutrientes de origem animal.

Observando-se os dados da Tabela 1 foi possível constatar que:

1. Austrália, Estados Unidos e países europeus, em geral, apresentaram valores grandes para Y1 e intermediários para Y2. Constituem casos a parte Portugal, com valor baixo para Y1, e Grécia, Itália e Espanha que apresentaram valores elevados também para Y2;

2. Com relação aos países da América do Sul, Brasil, Bolívia, Peru e Venezuela se carac-

terizaram por pequena disponibilidade de nutrientes tanto de origem animal como de vegetal (Y1 e Y2 pequeno). Uruguai e Argentina apresentaram disponibilidade acentuadamente de origem animal (Y1 grande e Y2 pequeno), o contrário ocorrendo com o Paraguai;

3. Camarões, Egito, Nigéria, Uganda, Síria e Turquia mostraram predominância de disponibilidade de nutrientes de origem vegetal (Y1 pequenos e Y2 grandes).

4. Bangladexe e Paquistão destacaram-se como países com baixa disponibilidade de nutrientes de ambas as origens;

5. O Japão, apesar de seu extraordinário nível de desenvolvimento econômico e tecnológico, não se destacou, apresentando disponibilidade *per capita* intermediária para os nutrientes de origem animal e vegetal.

A Tabela 3 apresenta valores *per capita* por dia de disponibilidade total (origens animal e

TABELA 3. Valores *per capita* por dia da disponibilidade total (animal e vegetal de calorias, proteínas (em grammas) e lipídios (em grammas) para alguns países, continentes e para o mundo. Percentagem em relação aos valores dos Estados Unidos, usados como referência.

Local	CAL	PROT	LIP
Mundo	2694 (74.0)	70.3 (66.0)	65.3 (39.7)
Europa	3397 (93.0)	101.5 (95.3)	135.4 (82.4)
Ásia	2485 (68.2)	60.7 (57.0)	42.9 (26.1)
América do Sul	2622 (72.0)	66.2 (62.2)	63.1 (38.4)
África	2299 (63.1)	57.4 (53.9)	46.9 (28.5)
Estados Unidos	3642 (100.0)	106.5 (100.0)	164.4 (100.0)
Canadá	3425 (94.0)	96.4 (90.5)	154.1 (93.7)
Brasil	2643 (72.6)	61.1 (57.4)	57.5 (35.0)
Austrália	3326 (91.3)	100.7 (94.6)	137.0 (83.3)
Nova Zelândia	3407 (93.5)	110.4 (103.7)	142.8 (86.9)
União Soviética	3394 (93.2)	105.6 (99.2)	101.6 (61.8)
China	2628 (72.2)	62.0 (58.2)	41.4 (25.2)
Índia	2204 (60.5)	53.9 (50.6)	35.7 (21.7)
Bangladexe	1922 (52.8)	40.6 (38.1)	18.8 (11.4)

vegetal) de calorias, proteínas e lipídios para alguns países, para continentes e para o Mundo. Usando-se os Estados Unidos como referência (100%) foram calculadas as percentagens dos outros locais.

Para estes nutrientes o Brasil mostrou-se inferior ao Mundo, similar à América do Sul e à China e superior à Ásia e à África. Comparado com os Estados Unidos, Canadá, União Soviética, Austrália, Nova Zelândia e Europa, a inferioridade se evidenciou de forma marcante, principalmente para lipídios e proteínas.

REFERÊNCIAS

- ANUÁRIO DA PRODUÇÃO AGROPECUÁRIA, Brasília: FAO, v.42, p.1-350, 1988.
- CURI, P.P. Resultados comparativos de alguns métodos de análise multivariada aplicados a um conjunto de dados. *Revista de Matemática Estatística*, v.1, p.55-67, 1983.
- JOLICOEUR, P.; MOSIMANN, J.E. Size and shape variation in the painted turtle. A principal component analysis. *Growth*, v.24, p.339-354, 1960.
- MORRISON, D.F. *Multivariate statistical methods*. New York: Mc Graw Hill, 1967. 338p.
- ROHLF, F.J. Adaptive hierarchical clustering schemes. *Systematic Zoology*, v.19, p.58-82, 1970.
- ROHLF, F.J. *Numerical Taxonomy and multivariate analysis system*. New York: Exeter Publishing, 1988. 100p.
- SNEATH, P.H.A.; SOKAL, R.R. *Numerical taxonomy*. San Francisco: W. H. Freeman, 1973. 573p.