



## Utilização de técnicas estatísticas multivariadas para definição de ambiente de produção do peso ao sobreano para o estudo da interação genótipo-ambiente em bovinos Canchim<sup>1</sup>

Maira Mattar<sup>2</sup>, Maurício Mello de Alencar<sup>3</sup>, Fernando Flores Cardoso<sup>4</sup>, Antonio Sergio Ferraudo<sup>5</sup>, Luis Otávio Campos Silva<sup>6</sup>, Ana Carolina Espasandin<sup>7</sup>

<sup>1</sup>Parte da tese de doutorado da primeira autora, financiada pela CAPES

<sup>2</sup>Doutoranda do Programa de Pós-graduação em Zootecnia – UNESP/Jaboticabal. Bolsista da CAPES. e-mail: [mairamattar@hotmail.com](mailto:mairamattar@hotmail.com)

<sup>3</sup>Pesquisador da Embrapa Pecuária Sudeste - São Carlos - SP. email: [mauricio@cnpse.embrapa.br](mailto:mauricio@cnpse.embrapa.br)

<sup>4</sup>Pesquisador da Embrapa Pecuária Sul – Bagé - RS. email: [fcardoso@cppsul.embrapa.br](mailto:fcardoso@cppsul.embrapa.br)

<sup>5</sup>Departamento de Ciências Exatas - FCAV – UNESP/Jaboticabal. email: [fsajago@gmail.com](mailto:fsajago@gmail.com)

<sup>6</sup>Pesquisador da Embrapa Gado de Corte – Campo Grande - MS. email: [locs@cnpqc.embrapa.br](mailto:locs@cnpqc.embrapa.br)

<sup>7</sup>Departamento de Producción Animal y Pasturas, UDELAR, Paysandú, Uruguay. email: [acespasa@fagro.edu.uy](mailto:acespasa@fagro.edu.uy)

**Resumo:** As respostas diferentes de genótipos às variações ambientais são investigadas nos estudos de interação genótipo-ambiente. A definição de ambiente nesses estudos ainda é um desafio, pois muitos fatores não genéticos podem causar efeito sobre a expressão de um conjunto de genes. Neste estudo, com as técnicas de estatística multivariada foram definidos ambientes de produção de bovinos Canchim, por meio de variáveis ambientais, formando grupos homogêneos de municípios do estado de São Paulo com informações de peso ao sobreano desta raça. As técnicas de Agrupamento Hierárquico e não Hierárquico foram eficientes para a simplificação e formação de quatro clusters homogêneos com membros de municípios paulistas, e heterogêneos entre si. Já, a técnica de componentes principais (CP) permitiu discriminar para cada cluster os fatores ambientais mais relevantes em sua formação, através de dois CP que preservaram 81,52% da variabilidade contida no conjunto das variáveis ambientais originais. As técnicas de estatística multivariada foram, portanto, ferramentas eficientes para discriminar ambientes de produção em estudos da interação genótipo-ambiente de bovinos Canchim.

**Palavras-chave:** análise de agrupamento, bovinos Canchim, componentes principais, estatística multivariada, interação genótipo-ambiente, peso ao sobreano

**Abstract:** The different answers of genotypes in environment variations are investigated in the studies of genotype-environment interaction. The definition of environment in these studies is still a challenge, because non genetic factors can cause effect on the expression of a set of genes. In this study, using multivariate statistics method were define environment for production of Canchim cattle, by means of environments variables, creating homogeneous groups from São Paulo State cities with information of post-yearling weight in this race. The hierarchic and non hierarchic cluster analyses were efficient in the simplification and homogeneous formation of four clusters, and heterogeneous among themselves. However, the main principal components technique (PC) allowed discriminating, for each cluster, the most relevant environment factors in its formation, through two PC that preserved 81.52% of the variability contained in the set of the original environments variables. The multivariate statistics methods were efficient tools to discriminate production environments in studies of the genotype-environment interaction of Canchim cattle.

**Keywords:** Canchim cattle, cluster analysis, interaction genotype-environment, multivariate statistic, principal components, post-yearling weight

### Introdução

Existem várias metodologias para investigar a presença de interação genótipo-ambiente (IGA) em características de interesse econômico de bovinos. Nesses estudos, um ambiente é definido como um conjunto de elementos que exerce alguma influência sobre a expressão dos genótipos dos animais. No entanto, existe uma dificuldade para definir um ambiente, pois sua formação depende tanto de informações de variáveis conhecidas como as desconhecidas e inexistentes nos bancos de dados. Como no Brasil existe uma grande diversidade de sistemas de produção de gado bovino, podem ocorrer grandes divergências ambientais em locais próximos. Desta forma, uma alternativa para definir ambientes de produção homogêneos é utilizando variáveis de elementos meteorológicos do clima e fatores geográficos do clima de cada ponto de produção. Com o avanço da informática e a possibilidade do uso de metodologias mais sofisticadas, as técnicas de estatística multivariada proporcionam a análise simultânea

de múltiplas variáveis sobre qualquer objeto estudado (Hair, et al. 2005), e permite a redução do espaço dimensional das variáveis para uma maior compreensão da análise. Para definir ambiente, é possível utilizar técnicas de agrupamento e componentes principais, onde a primeira permite a classificação de objetos em grupos diferentes, buscando alcançar homogeneidade dentro e heterogeneidade entre os grupos, em relação às variáveis estudadas e, a segunda usada para medir o poder de cada variável nos agrupamentos formados. O objetivo desse trabalho foi definir grupos homogêneos de municípios do estado de São Paulo com informações de peso ao sobreano de bovinos Canchim para o estudo da interação genótipo-ambiente desta raça.

### Material e Métodos

Os ambientes estudados foram 51 municípios do estado de São Paulo com rebanhos de bovinos da raça Canchim e com informações de peso ao sobreano. Para simplificar a diversidade ambiental foi realizada uma análise de agrupamento, para formar grupos de municípios definidos por variáveis ambientais de características meteorológicas e geográficas do clima (altitude (m), chuva (mm), classificação climática de Koeppen (Cck), Latitude (°), Longitude (°), Temperatura média mínima anual (°C) e Temperatura média máxima anual (°C)), obtidos pelo CEPAGRI (Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura). Por não possuir hipótese a “priori” sobre a estrutura e comportamento das variáveis ambientais nos 51 municípios do estado de São Paulo, o método de análise de agrupamento hierárquico foi realizado para efetuar uma organização hierárquica de grupos de municípios, por meio da similaridade que os torna membros ou não de um mesmo grupo com características ambientais comuns. O coeficiente de semelhança usado para quantificar a similaridade entre os municípios em relação às variáveis ambientais estudadas foi a distância euclidiana. O método de agrupamento usado foi de Ward, que busca o mínimo desvio padrão entre os dados de cada grupo. Com as informações de agrupamentos a “priori” obtidas pelo método hierárquico foi realizada uma análise através do método não hierárquico de agrupamento k-médias que é um método que busca dividir os dados de acordo com sua proximidade a um número arbitrário de centróides (média das variáveis agrupadas), sendo a distância euclidiana utilizada para medir a distância entre cada município e um centróide. Após a simplificação dos ambientes pelo método de agrupamento k-médias foi avaliado a relevância da quantidade de informações contidas em cada variável ambiental por meio da análise de componentes principais, onde se pôde observar o poder discriminatório das variáveis ambientais para cada cluster, cada um representante de um ambiente distinto. As técnicas de análises multivariadas foram realizadas através do programa Statsoft (2004).

### Resultados e Discussão

O dendrograma obtido pelo método hierárquico de agrupamento mostrou a formação de quatro grandes grupos de municípios do estado de São Paulo por meio de uma seqüência crescente de partições, estes quatro grupos foram avaliados posteriormente pelo método de agrupamento não-hierárquico k-médias como mostra a figura 1.

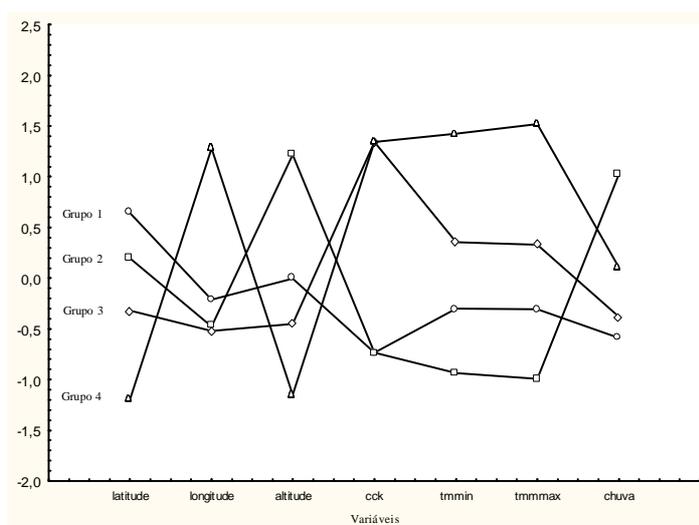


Figura 1. Médias das variáveis ambientais de cada cluster obtidas via agrupamento não hierárquico k-médias.

A figura 1 apresenta os conjuntos de médias (centróides) das variáveis ambientais de cada agrupamento, onde cada uma destas médias representa um ponto no espaço de casos (municípios) deste conjunto de

dados. Uma boa divisão de grupos pôde ser interpretada, pois os centróides de cada cluster, referentes às variáveis ambientais, em sua maioria, foram distintas entre si. A análise de variância apresentou resultados estatisticamente significativos a 95% de probabilidade para cada uma das variáveis ambientais dos clusters definidos pelo método k-médias. Assim, a representação de todas as variáveis ambientais como critérios de classificação foram importantes para a formação classificatória dos agrupamentos contendo municípios. A menor média de latitude encontrada foi para o cluster 4, proporcionando um agrupamento com maiores médias Tmmín e Tmmáx, devido a proximidade da linha do equador dos municípios pertencentes a este cluster. Já o inverso ocorreu para o cluster 1, com municípios de inverno mais rigoroso. O cluster 2 apresentou média mais alta de altitude e chuva, conferindo um agrupamento com municípios com as menores Tmmín e Tmmáx. O cluster 3 possuiu médias intermediárias de altitude, Tmmín, Tmmáx e chuva. As médias dos clusters evidenciaram que os agrupamentos formaram clusters climáticos distintos de municípios, alguns com climas mais quentes, intermediários e mais frios. Na análise de componentes principais, a dispersão dos municípios é mostrada na Figura 2 utilizando-se os dois primeiros componentes principais (CP1 e CP2) cujos autovalores da matriz de covariância foram superiores a unidade (Kaiser, 1958) e conseguiram preservar 81,52% da variabilidade contida no conjunto das variáveis ambientais originais. Assim, cada município ficou definido nesse espaço bidimensional pelas coordenadas de CP1 e CP2. Os coeficientes de correlação mais altos entre as variáveis ambientais e CP1 foram -0,97 e -0,93 (Tmmáx e Tmmín), respectivamente. No CP2, a chuva foi a variável com mais informação retida, com correlação igual a -0,91. Correlações negativas indicam que a posição dos municípios são discordantes com o eixo do CP1 ou CP2, como mostra a figura 2.

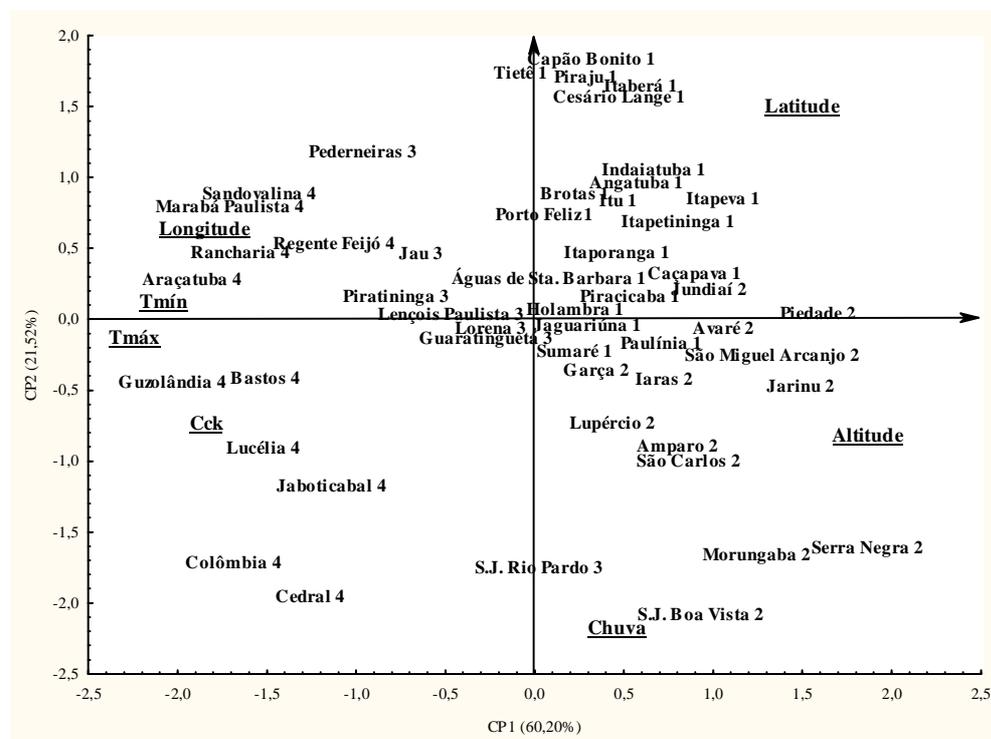


Figura 2. Gráfico biplot dos clusters e seus municípios e as variáveis ambientais.

### Conclusões

As análises estatísticas multivariadas de agrupamento hierárquico e não hierárquico, e componentes principais mostraram ser eficientes ferramentas utilizadas para discriminar ambientes de produção em estudos da interação genótipo-ambiente para a característica peso ao sobreano, formados segundo seus padrões ambientais, em bovinos da raça Canchim.

### Literatura citada

HAIR, J.F, ANDERSON, R.E., TATHAM, R.L. et al. **Análise Multivariada de dados**. Porto Alegre. RS, 5ª ed., 2005. Editora: Bookman Editora.

KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, 23, p. 187-200, 1958

STATSOFT, Inc. (2004). **STATISTICA** (data analysis software system), version 7. [www.statsoft.com](http://www.statsoft.com).