Análise Discriminante Baseada no Quociente das Matrizes de Correlaçoes

Maria Ivete de Barros Brugnerotto

Orientador: Prof. Dr. José Antonio Cordeiro

Campinas, dezembro de 1990

Análise Discriminante baseada no quociente das matrizes de correlações

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra. Maria Ivete de Barros Brugnerotto e aprovada pela Comissão Julgadora.

Campinas, 21 de janeiro de 1991

José Antonio Cordeiro
Orientador

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estaística.

À Thaís, Denis e Lia

AGRADECIMENTOS

Ao Professor Dr. José Antonio Cordeiro, meu orientador.

Ao Professor Dr. Ademir Petenate, meu co-orientador.

À Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC).

À Pesquisadora Dra. Edna Dora Martins Newman Luz da Divisão de Fitopatologia da CEPLAC.

À Professora Dra. Eliana Maria Freitas Marques.

À *Dona* Stellamaris Nalesso Lombardi.

A Adriana B. Santos e Newton Pimenta Neves Júnior.

Aos amigos Elena Ganon, Mauricio Chakkour e Orlando Bordoni.

À amiga Joassy P. N. Jorge, *in memoriam*.

Aos amigos do IMECC.

ÍNDICE

1. INTRODUÇÃO

Em diversas áreas de pesquisa como Psicologia, Agronomia e Economia, muitas vezes os resultados de estudos aparecem como um conjunto de dados multivariados e os pesquisadores têm interesse em utilizar técnicas estatísticas que permitam analisar esses dados em conjunto.

A área da Estatística que possibilita trabalharmos com dados multivariados é denominada de Análise Multivariada. Neste trabalho consideraremos a técnica - Análise Discriminante - que é utilizada com os objetivos de discriminar as populações e alocar novas observações em uma das populações pré-determinadas.

Em 1936, a Análise Discriminante foi utilizada por Fisher (1936) em um problema de taxionomia. Fisher utilizou um conjunto de quatro medidas tomadas em flores de 50 plantas obtidas junto a cada uma das espécies *Iris setosa*, *Iris virginica* e *Iris versicolor*.

Nesse trabalho, primeiramente Fisher considerou somente as espécies *Iris setosa* e *Iris versicolor* a fim de encontrar qual a combinação linear das quatro medidas das flores mais apropriada para discriminar as espécies entre si. Isto é, ele procurava a função linear dessas quatro medidas que maximizava o quociente da diferença, ao quadrado, entre as médias das duas espécies e a variância dentro da espécie. Podemos ainda expressar a idéia de Fisher como sendo a de encontrar a combinação linear das medidas que maximizava a razão entre a soma de quadrados entre as espécies e a soma de quadrados dentro das espécies.

Fisher (1936) considerava que as matrizes de covariância das espécies eram iguais. Entretanto, sabemos que isso nem sempre é verdade. Desta forma, Smith (1947) apresentava a função discriminante quadrática para ser utilizada nos casos em que as covariâncias são distintas.

Outros métodos de determinar a função discriminante foram desenvolvidas e entre eles podemos citar: de máxima verossimilhança, de Bayes, etc. Mais detalhes sobre esses métodos podem ser vistos, por exemplo, em Anderson (1984), capítulo 6.

A regra discriminante encontrada através do método da máxima verossimilhança é baseada na idéia de alocar a observação na população que tem a maior verossimilhança para esse dado.

A regra discriminante de Bayes é fundamentada no fato que existem situações nas quais as populações têm probabilidades *a priori*. Assim, a regra discriminante de Bayes, diz que devemos alocar a observação na população onde o produto da probabilidade *a priori* pela função de verossimilhanca é máxima. Neste caso, a regra discriminante minimiza o custo esperado de classificações incorretas.

Vale a pena ressaltar que se os custos esperados de classificações incorretas forem iguais em cada uma das populações e se também forem iguais as probabilidades *a priori* então as regras de discriminação de Bayes e de verossimilhança coincidem.

Em Ticeran (1988) foi proposto um novo método de discriminação, para o caso de duas populações multivariadas, baseado na comparação das matrizes de covariâncias. Quando comparado com o método de classificação baseado na verossimilhança, o novo método mostrou ser equivalente no caso em que as populações tinham distribuição normal bivariada.

A vantagem do método discriminante baseado na comparação das matrizes de covariâncias em relação ao método clássico é que, independentemente do número inicial de variáveis, ele sempre fornece duas combinações lineares dessas variáveis. Essas duas combinações lineares são obtidas através de uma transformação aplicada aos dados originais e podem ser utilizadas para uma representação gráfica.

Um dos problemas que surge na Análise Discriminante clássica e no método proposto em Ticeran (1988) é que estas técnicas são afetadas pela grandeza dos dados univariados.

Por exemplo, se estamos trabalhando com dados referentes a dois tipos de bovinos e as medidas observadas são: peso do bovino vivo na fazenda (em quilogramas), peso das vísceras do bovino morto (em gramas) e algumas outras variáveis também medidas em gramas, então as regras discriminantes obtidas através dos dois métodos: o clássico e o baseado na comparação das matrizes de covariâncias poderão não ser tão eficientes na classificação de novos indivíduos, devido à grande discrepância existente na grandeza dos dados.

Neste trabalho, propomos estudar um método de discriminação, para o caso de duas populações multivariadas, baseado no quociente das matrizes de correlações.

Para podermos entender melhor esta técnica, comecemos por rever o que é um dado multivariado.

Um dado ou observação multivariada é um conjunto de medições realizadas num mesmo indivíduo. Como exemplo citamos um conjunto de dados utilizado por Flury (1988).

Em uma nota de mil francos suíços, tomaram-se as seguintes medições: comprimento da nota, largura do lado esquerdo, largura do lado direito, largura da margem inferior, largura da margem superior e comprimento da diagonal do quadro central, medida do canto inferior esquerdo até o canto superior direito.

Neste caso, observaram-se seis variáveis para cada nota e essas seis observações conjuntas formam uma observação multivariada.

A partir desse exemplo bem simples, verificaremos como a Análise Discriminante pode ser utilizada.

Primeiramente, consideramos que foram coletadas uma amostra ao acaso de cem notas verdadeiras e uma outra amostra de cem notas falsas.

Suponhamos que essas 200 notas foram separadas e colocadas em duas caixas: uma caixa contendo as notas verdadeiras e outra contendo as notas falsas. Após ter sido feita essa separação, 20 novas notas chegam e devem ser classificadas como verdadeiras ou falsas. Como poderemos fazer essas classificações?

Se solicitarmos a uma pessoa, que faça essa classificação somente pelo método visual, provavelmente, se as notas foram bem falsificadas, ela terá dificuldade na separação e acabará cometendo um certo número de erros de classificação.

Entretanto, essa pessoa poderia utilizar a Análise Discriminante para classificar essas novas notas, minimizando segundo certos critérios, o número de classificações incorretas.

Observamos então que a Análise Discriminante é uma das técnicas de Análise Multivariada e tem por objetivo descobrir regras de discriminação e classificação.

Essas regras possibilitarão classificarmos novos indivíduos, de origem indeterminada, na população correta com a máxima probabilidade possível e descobrirmos que variáveis (e de que forma) atuam mais na separação das populações. Lembramos ainda que para podermos fazer uso desta técnica, devemos saber *a priori* que nosso conjunto de dados vem de populações distintas e conhecidas.

O objetivo deste trabalho é estudar o método discriminante baseado nas matrizes de covariâncias quando as variáveis são padronizadas. Isto equivale a dizer que estudaremos a análise discriminante baseada no quociente das matrizes de correlações.

Outro objetivo deste trabalho é verificar de que forma poderemos utilizar o método proposto na Análise Discriminante para reduzir o número de variáveis originais sem ocasionar grandes perdas na classificacão de novos indivíduos.

Ressaltamos que ao fazermos a padronização das variáveis, estaremos alterando a estrutura inicial da covariância. Essas novas variáveis terão matrizes de covariância igual as matrizes de correlação dos dados originais e portanto, não serão afetadas pela grandeza dos dados coletados.

A seguir apresentamos uma breve descrição do conteúdo de cada capítulo.

No capítulo 2, apresentaremos a técnica clássica de Análise Discriminante para os casos em que apenas duas populações são consideradas. Para o caso de três ou mais populações podemos consultar por exemplo Mardia (1979). No capítulo 3, exporemos o método discriminante apresentado em Ticeran (1988) onde a discriminação é realizada com base no quociente das matrizes de dispersões. O capítulo 4 apresentará o método discriminante apresentado por Ticeran (1988) aplicado aos dados padronizados o que é equivalente a realizar a análise discriminante com base nas matrizes de correlações. No capítulo 5, apresentaremos os testes estatísticos para as hipóteses de igualdade de médias, igualdade de matrizes de covariâncias e igualdade das matrizes de correlações. O capítulo 6 conterá algumas aplicações práticas do método discriminante desenvolvido no capítulo 4 como também apresentará uma comparação deste método com os métodos dos capítulos 2 e 3 para os conjuntos de dados analisados. O capítulo 7 apresentará um resumo da comparação dos três métodos discriminantes para os conjuntos de dados analisados. Conterá também alguns comentários adicionais sobre a técnica exposta no capítulo 4.

O anexo A apresenta algumas técnicas estatísticas utilizadas na determinação dos auto-valores e auto-vetores e apresenta também as aproximações utilizadas para a determinação dos percentis das distribuições Normal, F e χ^2. No anexo B são descritos o programa principal e as sub-rotinas desenvolvidas para os métodos de discriminação apresentados nos capítulos 3 e 4 e para os testes do capítulo 5. O anexo C apresenta as listagens do programa principal e suas sub-rotinas. No anexo D listamos uma saída do programa computacional como um exemplo dos resultados fornecidos por esse programa. O anexo E consta dos gráficos obtidos através do método discriminante baseado no quociente das matrizes de correlações e do baseado no quociente das covariâncias para os conjuntos de dados analisados.

2. ANÁLISE DISCRIMINANTE CLÁSSICA PARA O CASO DE DUAS POPULAÇÕES

Neste capítulo, apresentaremos a técnica de Análise Discriminante clássica para os casos onde apenas duas populações são consideradas.

Sejam π_1 e π_2 duas populações multivariadas e seja ${}^{i}U$ a matriz de amostras da população i com n_i observações e p variáveis dada por:

$$\underset{n_i \times p}{{}^{i}U} = \begin{bmatrix} {}^{i}X_{11} & \cdots & {}^{i}X_{1p} \\ \vdots & & \vdots \\ {}^{i}X_{n_i 1} & \cdots & {}^{i}X_{n_i p} \end{bmatrix} \quad , \quad i=1,2 \; .$$

Nesse caso, temos que cada coluna de ${}^{i}U$ representa uma variável aleatória univariada ${}^{i}X_j$, $j = 1,\ldots,p$, com n_i observações.

Para o exemplo das notas falsas e verdadeiras temos que ${}^{1}U$ é a matriz de dados das notas verdadeiras enquanto que ${}^{2}U$ é a matriz referente às notas falsas.

Neste exemplo, ainda temos que $p = 6$, $n_1 = n_2 = 100$, sendo:

${}^{i}X_1$=comprimento da nota ${}^{i}X_2$=largura do lado esquerdo
${}^{i}X_3$=largura do lado direito ${}^{i}X_4$=largura da margem inferior
${}^{i}X_5$=largura da margem superior ${}^{i}X_6$=comprimento da diagonal.

Podemos representar uma observação multivariada desta população π_i por um vetor ${}^{i}X$ dado por:

$$\underset{p\times 1}{{}^{i}X} = ({}^{i}X_1 \quad \dots \quad {}^{i}X_p)^t \quad , \quad i=1,2.$$

Nosso objetivo é separar, com base na regra discriminante d, o espaço das observações multivariadas em duas regiões ou grupos disjuntos, digamos R_1 e R_2 tal que:

$$\mathbb{R}^p = R_1 \cup R_2 \ .$$

Esta regra discriminante deverá minimizar o erro de classificação incorreta dos dados.

Para determinarmos tais regiões, primeiramente necessitamos definir como é avaliado o custo de classificação incorreta.

Definição 2.1: O custo esperado de classificação incorreta da regra discriminante d é definido por:

$$CEM = q_1 C(2|1)\ P(2|1,d) + q_2 C(1|2)\ P(1|2,d)$$

onde:

$q_i = P(X \in \pi_i)$ é a probabilidade *a priori* de obtermos uma observação da população π_i

$C(j|i)$ = o custo de classificarmos uma observação na população π_j sendo que ela pertence a π_i

$P(j|i,d)$ = a probabilidade condicional de classificarmos uma observação de π_i na população π_j, segundo a regra discriminante d.

A regra discriminante que minimiza o custo esperado de classificação incorreta é dada pelo teorema a seguir:

Teorema 2.1: Sejam q_1 e q_2 as probabilidades *a priori* de obtermos uma observação das populações π_1 e π_2, respectivamente. Suponhamos que as populações π_1 e π_2 têm funções densidades de probabilidades f_1 e f_2, respectivamente. Se o custo de classificarmos erradamente uma observação na população π_1 é dado por $C(1|2)$ e o custo de classificarmos uma observação de π_1 na população π_2 é $C(2|1)$, então as regiões de classificação R_1 e R_2 que minimizam o custo esperado de classificação incorreta são dadas por:

$$R_1 = \left\{ x \in \mathbb{R}^p \;\middle|\; \frac{f_1(x)}{f_2(x)} \geq \frac{q_2 \; C(1|2)}{q_1 \; C(2|1)} \right\}$$

$$R_2 = \left\{ x \in \mathbb{R}^p \;\middle|\; \frac{f_1(x)}{f_2(x)} < \frac{q_2 \; C(1|2)}{q_1 \; C(2|1)} \right\} .$$

Para vermos a demonstração deste teorema podemos consultar, por exemplo, Johnson e Wichern (1982), p. 476.

Agora que já temos as regiões de classificação, podemos apresentar a regra discriminante d.

Definição 2.2: A regra discriminante d é definida por:

alocamos x a π_1 se $x \in R_1$ e,

alocamos x a π_2 se $x \in R_2$,

com R_1 e R_2 sendo as regiões determinadas através do teorema (2.1).

Supondo que os custos $C(1|2)$ e $C(2|1)$ sejam iguais, como também sejam iguais as probabilidades *a priori* q_1 e q_2 teremos então, que a regra discriminante d é definida por:

$$\text{alocamos } x \text{ a } \pi_1 \text{ se } \quad \frac{f_1(x)}{f_2(x)} \geq 1$$

e,

$$\text{alocamos } x \text{ a } \pi_2 \text{ se } \quad \frac{f_1(x)}{f_2(x)} < 1.$$

Outra maneira de definirmos a regra discriminante é utilizando a função de verossimilhança. Para isso começaremos apresentando essa função.

Definição 2.3: Sejam $X_1, \ldots, X_n$ vetores aleatórios, independentes e identicamente distribuídos com função de densidade dada por $f(x,\theta)$ onde θ é o vetor de parâmetros desconhecidos.

A função de verossimilhança para essa amostra aleatória é dada por:

$$L(\theta, x_1, \ldots, x_n) = L(\theta, x) = \prod_{i=1}^{n} f(x_i, \theta) .$$

A seguir definimos a regra discriminante baseada na função de verossimilhança.

Definição 2.4: Sejam π_1 e π_2 populações multivariadas com funções de densidades f_1 e f_2, respectivamente

A regra discriminante baseada na função de verossimilhança é dada por:

$$\text{alocamos } x \text{ em } \pi_1 \text{ se } \frac{L_1(\theta,x)}{L_2(\theta,x)} \geq 1$$

$$\text{alocamos } x \text{ em } \pi_2 \text{ se } \frac{L_1(\theta,x)}{L_2(\theta,x)} < 1$$

onde $L_i(\theta,x)$ é a função de verossimilhança da população π_i.

Observamos que esta regra discriminante é equivalente à regra baseada na minimização do custo esperado de classificação incorreta quando $C(1|2) = C(2|1)$ e $q_1 = q_2$.

Mostraremos a seguir a regra de discriminação em alguns casos particulares.

Suponhamos que as populações multivariadas π_1 e π_2 tenham distribuição normal p-variada com parâmetros conhecidos. Sejam μ_1 e μ_2 os vetores de médias e sejam Σ_1 e Σ_2 as matrizes de covariâncias, respectivamente.

<u>Caso 1</u> Se as matrizes de covariâncias, Σ_1 e Σ_2, forem iguais a

uma determinada matriz Σ, então a regra discriminante é dada por:

alocamos x a π_1 se:

$$\exp\left[\frac{1}{2}\left(2\,x^t\,\Sigma^{-1}(\mu_1-\mu_2)+(\mu_2-\mu_1)^t\,\Sigma^{-1}(\mu_2+\mu_1)\right)\right]\geq 1$$

ou ainda se:

$$2\,x^t\,\Sigma^{-1}(\mu_1-\mu_2)+(\mu_2-\mu_1)^t\,\Sigma^{-1}(\mu_2+\mu_1)\geq 0 \qquad (2.1)$$

e, alocamos x a π_2 em caso contrário.

Entretanto, o que ocorre em muitas aplicações é que os parâmetros das distribuições são desconhecidos, portanto devemos utilizar amostras destas populações para estimá-los.

Após os parâmetros μ_1, μ_2 e Σ serem estimados serão colocados na fórmula (2.1) e obteremos desta forma a função de discriminação amostral.

Como sabemos, os estimadores não viciados de μ_i e Σ são dados por:

$$\hat{\mu}_i = \bar{x}_i = \frac{1}{n_i} \, {}^iU^t \, 1_{n_i} \quad , i=1,2$$

$$\hat{\Sigma} = S = \frac{(n_1 - 1)S_1 + (n_2 - 1)\,S_2}{n_1 + n_2 - 2} \, ,$$

onde

$$1^t_{n_i} = (\,1,\ 1,\ \ldots,\ 1)$$

$$\hat{\Sigma}_i = S_i = \frac{1}{n_i - 1} \left({}^iU^t \; {}^iU - n_i \, \bar{x}_i \, \bar{x}_i^{\,t} \right), \qquad i=1,2.$$

Lembramos que a matriz S é chamada de matriz de covariâncias amostral combinadas.

Portanto, substituindo os parâmetros na fórmula (2.1) obtemos a seguinte regra de classificação amostral:

alocamos x em π_1 se:

$$2\, x^t\, S^{-1}\, (\bar{x}_1 - \bar{x}_2) + (\bar{x}_2 - \bar{x}_1)\, S^{-1}\, (\bar{x}_2 + \bar{x}_1) \geq 0$$

e, alocamos x em π_2 em caso contrário.

Caso 2: Se as matrizes Σ_1 e Σ_2 são diferentes, mas conhecidas, temos que a regra discriminante é

alocamos x a π_1 se:

$$\left|\frac{\Sigma_2}{\Sigma_1}\right|^{1/2} \exp\left[\frac{1}{2}\ (x-\mu_2)^t\ \Sigma_2^{-1}(x-\mu_2) - (x-\mu_1)^t\ \Sigma_1^{-1}(x-\mu_1)\right] \geq 1$$

ou ainda se:

$$\ln\frac{|\Sigma_2|}{|\Sigma_1|} + (x-\mu_2)^t\ \Sigma_2^{-1}\ (x-\mu_2) - (x-\mu_1)^t\ \Sigma_1^{-1}\ (x-\mu_1) \geq 0 \qquad (2.2)$$

e, alocamos x a π_2 em caso contrário.

Quando os parâmetros μ_i e Σ_i ,i=1,2, da distribuição são desconhecidos, utilizamos amostras para estimá-los como no caso anterior, obtendo desta forma a seguinte regra de classificação amostral:

alocamos x em π_1 se:

$$\ln \frac{|S_2|}{|S_1|} + (x - \bar{x}_2)^t S_2^{-1} (x - \bar{x}_2) - (x - \bar{x}_1)^t S_1^{-1} (x - \bar{x}_1) \geq 0$$

e, alocamos x em π_2 em caso contrário.

Observamos que o primeiro termo da inequação (2.1) é uma função linear dos componentes do vetor de observações e costuma ser denominada de função discriminante de Fisher.

Já a inequação (2.2) não é mais uma função linear, mas sim uma função quadrática dos componentes do vetor de observações.

Lembramos que essas regras de classificação obtidas para os casos 1 e 2 são as regras baseadas na função de verossimilhança e que coincidem com as regras de classificação que minimizam o custo esperado de classificação incorreta quando $q_1 \; C(2|1) = q_2 \; C(1|2)$.

No capítulo seguinte, apresentaremos o método de discriminação proposto em Ticeran (1988), o qual é baseado no quociente das matrizes de covariâncias e tem a vantagem de sempre permitir uma visualização gráfica das variáveis transformadas.

3. ANÁLISE DISCRIMINANTE BASEADA NO QUOCIENTE DAS MATRIZES DE COVARIÂNCIAS PARA O CASO DE DUAS POPULAÇÕES

Neste capítulo, apresentaremos a metodologia da Análise Discriminante proposta em Ticeran (1988) que é baseada no quociente das matrizes de covariâncias.

Ressaltamos que este método só é adequado quando as matrizes de covariâncias Σ_1 e Σ_2 diferem entre si, uma vez que ele é baseado na comparação dessas matrizes.

Sejam π_1 e π_2 duas populações com distribuição normal p variada e seja ${}^{i}U$ a matriz de dados da população π_i, formada pelas realizacões da variável p-dimensional ${}^{i}X$, como consideramos no capítulo 2, isto é,

$$\underset{p\times 1}{X^i} = (X_1^i, \ \ldots \ , X_p^i)^t \sim N_p(\mu_i , \Sigma_i), \quad i=1,2.$$

Consideremos agora a transformação proposta em Ticeran (1988), que leva nossa matriz de dados ${}^{i}U$ do espaço p dimensional para o espaço bidimensional.

Sejam γ_1 e γ_p os auto-vetores de $\Sigma_1^{-1}\Sigma_2$ correspondentes ao maior e menor auto-valor, α_1 e α_p, respectivamente e seja γ o vetor definido por:

$$\underset{p\times 2}{\gamma} = (\gamma_1 \mid \gamma_p).$$

Consideremos a seguir a seguinte transformação:

$$\underset{2\times 1}{{}^iW} = \gamma^t\, {}^iX = \begin{bmatrix} \gamma_1^t \\ \\ \gamma_p^t \end{bmatrix} {}^iX = \begin{bmatrix} \gamma_1^t\, {}^iX \\ \\ \gamma_p^t\, {}^iX \end{bmatrix}$$

Observamos então que este novo vetor ${}^iW$ está definido em $\mathbb{R}^2$ sendo que cada um de seus elementos é uma combinação linear do vetor das variáveis originais.

A primeira combinação linear é obtida através do auto-vetor de $\Sigma_1^{-1}\Sigma_2$ correspondente a maior raiz característica, enquanto que a segunda é obtida com o auto-vetor referente ao menor auto-valor. Desta forma essas combinações lineares apresentam respectivamente, a maior e a menor razão de variâncias, dentre as combinações lineares das coordenadas de ${}^1X$ e ${}^2X$.

Por ser ${}^iW$ apenas uma combinação linear de ${}^iX$ temos que:

$${}^iW \sim N_2(\xi_i, \Gamma_i)$$

onde

$$\xi_i = \gamma^t \mu_i \quad , \quad i=1,2$$

$$\Gamma_1 = I_2$$

$$\Gamma_2 = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_p \end{bmatrix}$$

Em Ticeran (1988) é apresentada a seguinte regra de discriminação, que minimiza o custo esperado de classificacão incorreta:

alocamos x na populacão π_1 se:

$$-\frac{1}{2} w^t (I_2 - \Gamma_2^{-1}) w + w^t (I_2 \xi_1 - \Gamma_2^{-1} \xi_2) + b \geq \ln \left[\frac{q_2 C(1|2)}{q_1 C(2|1)} \right]$$

com

$$b = \frac{1}{2} \ln |\Gamma_2| - \frac{1}{2} (\xi_1^t \xi_1 - \xi_2^t \Gamma_2^{-1} \xi_2)$$

e, alocamos x em π_2 em caso contrário.

Nos casos onde os parâmetros da distribuição de ^{i}X são desconhecidos temos conseqüentemente que os parâmetros da distribuição de ^{i}W também são desconhecidos e, portanto, devemos utilizar seus estimadores, para obtermos a função discriminante amostral.

Sejam S_1 e S_2 as matrizes de covariâncias amostrais de ^{1}X e ^{2}X, respectivamente.

Considerando g_j como sendo o auto-vetor de $S_1^{-1}S_2$ correspondente ao auto-valor a_j, $j=1,p$, definimos a transformação para essas amostras por:

$$^{i}w = g^{t}\ {}^{i}x \quad , \quad i=1,2,$$

com $g = (g_1 \mid g_p)$.

Neste caso, os parâmetros ξ_i e Γ_i da distribuição de ^{i}W são estimados por:

$$\hat{\xi}_i = \bar{w}_i \quad , \quad i=1,2$$

$$\hat{\Gamma}_1 = T_1 = I_2$$

$$\hat{\Gamma}_2 = T_2 = \begin{bmatrix} a_1 & 0 \\ 0 & a_p \end{bmatrix}$$

Temos então a seguinte proposta de classificação amostral:

alocamos x em π_1 se

$$-\frac{1}{2} w^t (I_2 - T_2^{-1}) w + w^t (I_2 \bar{w}_1 - T_2^{-1} \bar{w}_2) + \hat{b} \geq \ln \frac{q_2 \ C(1|2)}{q_1 \ C(2|1)}$$

sendo,

$$\hat{b} = \frac{1}{2} \ln |T_2| - \frac{1}{2} (\bar{w}_1^{\,t} \bar{w}_1 - \bar{w}_2^{\,t} T_2^{-1} \bar{w}_2),$$

e, alocamos x em π_2 nos outros casos.

Nos casos em que os custos são iguais e que também o são as probabilidades *a priori*, temos:

alocamos x em π_1 se

$$-\frac{1}{2} w^t (I_2 - T_2^{-1}) w + w^t (I_2 \bar{w}_1 - T_2^{-1} \bar{w}_2) + \hat{b} \geq 0$$

e, alocamos x em π_2 em caso contrário.

Ao comparar este método de discriminação com o método clássico, no caso de duas populações bivariadas, Ticeran (1988) obteve as mesmas estimativas das probabilidades de classificações incorretas, ou seja, quando a hipótese de igualdade das covariâncias é rejeitada, os dois métodos discriminam igualmente as populações.

Uma pergunta que surge é como se comporta este método se em vez de trabalharmos com as variâncias e covariâncias o fizermos somente com as correlações?

O interesse neste caso aparece nas aplicações onde as matrizes de dados são compostas por variáveis medidas em escalas muito diferentes.

Podemos colocar esta questão da seguinte forma: como se apresentaria o método proposto em Ticeran (1988) se, ao invés de considerarmos o quociente das matrizes de covariâncias, trabalharmos com o quociente das matrizes de correlações? No próximo capítulo abordaremos este problema.

4. ANÁLISE DISCRIMINANTE BASEADA NO QUOCIENTE DAS MATRIZES DE CORRELAÇÕES PARA O CASO DE DUAS POPULAÇÕES

Neste capítulo, desenvolveremos a metodologia para a análise discriminante baseada na comparação das matrizes de correlações, como também, apresentaremos uma forma de eliminação de variáveis com base nessa análise.

4.1 Análise Discriminante baseada no quociente das matrizes de correlações

Novamente, consideraremos as duas populações normais multivariadas π_1 e π_2, como vimos no capítulo 2, isto é:

$$\underset{p \times 1}{{}^{i}X} \sim N_p(\mu_i, \Sigma_i), \ i=1,2.$$

Nosso interesse nesta seção é analisar o método de discriminação proposto em Ticeran (1988) nos casos em que eliminamos o efeito da magnitude das covariáveis. Isto equivale a dizer que desejamos estudar a discriminação de duas populações, baseando-nos no quociente das matrizes de correlações

Para realizarmos essa análise, é necessário que as matrizes de correlações das populações sejam diferentes, pois o método será baseado na comparação dessas correlações; logo, não fará sentido aplicar essa metodologia quando as matrizes de correlações forem iguais.

Iniciamos o processo padronizando os dados originais ou seja, começamos transformando os dados originais através de uma matriz diagonal cujos elementos são os desvios-padrões das coordenadas da matriz de dados.

Seja essa matriz diagonal definida por:

$$\underset{p\times p}{\Delta_i} = \text{diag}({}^{i}\sigma_{11} , \ldots , {}^{i}\sigma_{pp}) , \quad i=1,2,$$

onde ${}^{i}\sigma_{jj}$, $j=1,\ldots,p$, é o desvio padrão da j-ésima coordenada de ${}^{i}X$.

Definimos o vetor aleatório padronizado por:

$$\underset{p\times 1}{{}^{i}Z} = \Delta_i^{-1} \; {}^{i}X, \quad i=1,2.$$

Este novo vetor não tem a mesma estrutura de covariância dos dados originais e sua distribuição de probabilidade é:

$$\underset{p\times 1}{{}^{i}Z} \sim N_p(\Delta_i^{-1} \mu_i , \Delta_i^{-1} \Sigma_i \Delta_i^{-1}) = N_p(\Delta_i^{-1} \mu_i , \Psi_i) ,$$

onde Ψ_i é a matriz de correlação de ${}^{i}X$, i=1,2.

Observamos, portanto, que esse novo vetor ${}^{i}Z$ tem matriz de covariância igual a matriz de correlação de ${}^{i}X$.

Seguindo o processo, nosso problema agora é definir as combinações lineares de ${}^{i}Z$ que apresentam a maior e a menor razão de variâncias, referentes às duas populações padronizadas.

Neste sentido, estamos procurando as combinações lineares que maximizam e minimizam a função:

$$h(c) = \frac{c^t \, \Psi_2 \, c}{c^t \, \Psi_1 \, c} \quad , \quad \forall \, c \in \mathbb{R}^p .$$

O teorema a seguir ajudará solucionarmos esse problema.

Teorema 4.1.1: Seja A uma matriz simétrica n x n e seja D uma matriz n x n positiva definida. Sejam $\gamma_1 \geq \ldots \geq \gamma_n$ os auto-valores de $D^{-1}A$ correspondentes aos auto-vetores v_1, ..., v_n. Então

$$\sup_x \frac{x^t \, A \, x}{x^t \, D \, x} = \gamma_1$$

e

$$\inf_x \frac{x^t A x}{x^t D x} = \gamma_n$$

com os limites sendo atingidos quando $x = v_1$ e $x = v_n$, respectivamente.

Podemos ver a demonstração deste teorema em Seber (1984), p. 527.

Para aplicarmos este teorema no nosso problema faremos as seguintes considerações: Seja $\Psi_1 = A$ e $\Psi_2 = D$. Sejam $\lambda_1, \ldots, \lambda_p$ os auto-valores de $\Psi_1^{-1}\Psi_2$, em ordem decrescente e sejam $\beta_1, \ldots, \beta_p$ os auto-vetores correspondentes.

Temos então que:

$$\max_c \ h(c) = \frac{\beta_1^t \ \Psi_2 \ \beta_1}{\beta_1^t \ \Psi_1 \ \beta_1} = \lambda_1$$

e

$$\min_c \ h(c) = \frac{\beta_p^t \ \Psi_2 \ \beta_p}{\beta_p^t \ \Psi_1 \ \beta_p} = \lambda_p$$

Agora que já temos solucionado esse problema, começamos definindo uma transformação baseada nos auto-vetores de $\Psi_1^{-1}\Psi_2$ que será aplicada na matriz de dados padronizados

Seja:

$$\underset{2\times 1}{{}^{i}Y} = \beta^{t}\ {}^{i}Z \quad , \quad i=1,2$$

onde

$$\underset{p\times 2}{\beta} = (\beta_1 \ \vdots\ \beta_p).$$

O vetor ${}^{i}Y$ assim definido é uma combinação linear das coordenadas da matriz de dados ${}^{i}Z$, sendo o primeiro elemento formado através da combinação linear com o auto-vetor de $\Psi_1^{-1}\Psi_2$ correspondente ao maior auto-valor, enquanto que o segundo elemento é obtido com o auto-vetor correspondente ao menor auto-valor.

Esta transformação β aplicada a ${}^{i}Z$ permite-nos deixarmos o espaço p-dimensional e começarmos a trabalhar em apenas duas dimensões.

Nossa preocupação, então, será encontrar a distribuição de ${}^{i}Y$. Como ${}^{i}Y$ é apenas uma combinação linear de ${}^{i}Z$ temos que:

$$\underset{2\times 1}{{}^{i}Y} \sim N_2(\nu_i, \Lambda_i), \qquad i=1,2.$$

onde

$$\nu_i = E(\,{}^iY\,) = \beta^t\, E(\,{}^iZ\,) = \beta^t\, \Delta_i^{-1}\, \mu_i$$

$$\Lambda_i = Var(\beta^t\ {}^iZ\,) = \beta^t\, Var({}^iZ\,)\,\beta =$$

$$= \beta^t\, \Delta_i^{-1}\, \Sigma_i\, \Delta_i^{-1}\, \beta = \beta^t\, \Psi_i\, \beta\,, \qquad i=1,2.$$

Analisando mais detalhadamente $\Lambda_i = \beta^t\, \Psi_i\, \beta$ vemos que

$$\Lambda_i = \beta^t\, \Psi_i\, \beta = \begin{bmatrix} \beta_1^t \\ \beta_p^t \end{bmatrix} \Psi_i \begin{bmatrix} \beta_1 & \beta_p \end{bmatrix} =$$

$$= \begin{bmatrix} \beta_1^t\, \Psi_i\, \beta_1 & \beta_1^t\, \Psi_i\, \beta_p \\ \beta_p^t\, \Psi_i\, \beta_1 & \beta_p^t\, \Psi_i\, \beta_p \end{bmatrix} \qquad (4.1.1)$$

O teorema a seguir será utilizado para analisarmos mais detalhadamente $\beta_j^t\, \Psi_i\, \beta_k$.

Teorema 4.1.2: Dadas as matrizes A e B simétricas e positivas definidas, então existe uma matriz não singular F, tal que

a) $F^t\, B\, F = L = diag\,(\varphi_1, \ldots, \varphi_p)$

com $\varphi_1 \geq \ldots \geq \varphi_p > 0$ sendo as raízes características de $A^{-1}B$ e,

$$b) \quad F^t A F = I.$$

Se as raízes são distintas então $F = (r_1 \vdots \ldots \vdots r_p)$, onde r_j, $j=1,\ldots,p$, é o auto-vetor de $A^{-1}B$ correspondente ao auto-valor φ_j.

Se uma raiz tem multiplicidade m, então um conjunto de m desses auto-vetores linearmente independentes, pode ser trocado por m combinações lineares desses vetores característicos.

No nosso caso, as matrizes de correlações Ψ_1 e Ψ_2 são positivas definidas. Portanto, aplicando o teorema 4.1.2 obtemos:

$$a) \quad F^t \Psi_1 F = I$$

$$b) \quad F^t \Psi_2 F = L$$

onde F é a matriz formada pelos auto-vetores de $\Psi_1^{-1}\Psi_2$ e L é a matriz diagonal cujos elementos são os auto-valores correspondentes.

Suponhamos que as raízes características de $\Psi_1^{-1}\Psi_2$ sejam todas distintas. Nesse caso temos que:

$$F = (\beta_1 \mid \ldots \mid \beta_p).$$

Logo,

$$\beta_j^t \, \Psi_1 \, \beta_k = \begin{cases} 0 \; , & j \neq k \\ 1 \; , & j = k \end{cases} \quad , \quad j,k = 1,\ldots,p \qquad (4.1.2)$$

$$\beta_j^t \, \Psi_2 \, \beta_k = \begin{cases} 0 \; , & j \neq k \\ \lambda_j \; , & j = k \end{cases} \quad , \quad j,k = 1,\ldots,p. \qquad (4.1.3)$$

Portanto, substituindo os resultados (4.1.2) e (4.1.3) em (4.1.1) obtemos:

$$\Lambda_1 = \beta^t \, \Psi_1 \, \beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

$$\Lambda_2 = \beta^t \, \Psi_2 \, \beta = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_p \end{bmatrix}.$$

Agora que conhecemos a distribuição de ${}^{i}Y$, podemos encontrar a regra de classificação.

Em geral, quando estamos trabalhando com Análise Discriminante, desejamos encontrar uma regra que sirva não só para discriminar as populações, como também, que possa ser utilizada na classificação de novas observações.

Nos capítulos 2 e 3 tal regra era encontrada através da aplicação de uma observação no quociente das funções de verossimilhança. Porém, neste capítulo, não será possível determinar a regra de discriminação e classificação desta forma.

Para o método de discriminação proposto, observamos que quando uma nova observação chega para ser classificada, não sabemos *a priori*, a qual das populações ela pertence; logo, não sabemos fazer a padronização desta observação. Conseqüentemente não conseguimos definir a variável transformada Y que seria utilizada na função discriminante.

Para solucionarmos esse problema e encontrarmos uma regra de discriminação que possa não só ser utilizada na separação das populações, como também ser empregada na classificação, propomos o método descrito a seguir.

Supomos que a observacão x vem da população π_1 e a padronizamos usando Δ_1. Desta forma criamos:

$$^{1}Z_{*} = \Delta_1^{-1} X .$$

Em seguida, supomos que essa mesma observacão x pertence a π_2 e determinamos:

$$^{2}Z_{*} = \Delta_2^{-1} X .$$

De posse desses novos valores $^{1}Z_{*}$ e $^{2}Z_{*}$, definimos as transformações.

$$^{1}Y_{*} = \beta^{t}\ {}^{1}Z_{*}$$

$$^{2}Y_{*} = \beta^{t}\ {}^{2}Z_{*}$$

Calculamos então o quociente entre a densidade de $^{1}Y_{*}$ no ponto $^{1}y_{*}$ e a densidade de ^{2}Y em $^{2}y_{*}$ e definimos a seguinte regra de classificação:

$$\text{alocamos } x \text{ em } \pi_1 \text{ se } \frac{f(^{1}y_{*})}{f(^{2}y_{*})} \geq 1$$

ou, equivalentemente

$$\text{alocamos } x \text{ em } \pi_1 \text{ se } \ln \frac{f(^{1}y_{*})}{f(^{2}y_{*})} \geq 0.$$

e, alocamos x em π_2 em caso contrário.

Isso significa que estamos classificando x em π_1 quando a densidade da observação padronizada pela matriz de covariância de π_1 for maior que a densidade dessa mesma observação padronizada pela matriz de covariância da população π_2.

Como vimos anteriormente, ^{i}Y tem distribuição normal bivariada e conseqüentemente $^{i}Y_{*}$ também, logo substituindo as densidades temos que:

alocamos x em π_1 se

$$\ln \left[\frac{|\Lambda_1|^{-1/2} \exp \left[-\frac{1}{2} ({}^1y_* - \nu_1)^t \Lambda_1^{-1} ({}^1y_* - \nu_1) \right]}{|\Lambda_2|^{-1/2} \exp \left[-\frac{1}{2} ({}^2y_* - \nu_2)^t \Lambda_2^{-1} ({}^2y_* - \nu_2) \right]} \right] \geq 0$$

e, alocamos x em π_2 em caso contrário.

Lembrando que $\Lambda_1 = I$ e trabalhando algebricamente a expressão anterior obtemos a seguinte regra de classificação:

alocamos x em π_1 se

$$\begin{aligned} &{}^2y_*^t\, \Lambda_2^{-1}\, {}^2y_* - 2\, {}^2y_*^t\, \Lambda_2^{-1}\, \nu_2 - {}^1y_*^t\, {}^1y_* + 2\, {}^1y_*^t\, \nu_1 - \nu_1^t\, \nu_1 + \\ &+ \nu_2^t\, \Lambda_2^{-1}\, \nu_2 + \ln |\Lambda_2| \geq 0 \qquad (4.1.4) \end{aligned}$$

e, alocamos x em π_2 em caso contrário.

Para determinarmos a função discriminante amostral, iniciamos supondo que foram coletadas amostras aleatórias em cada uma das populações π_i e consideramos R_i como sendo a matriz de correlação amostral da população π_i, i=1,2

Sejam $b_1, \ldots, b_p$ os auto-vetores de $R_1^{-1} R_2$ correspondentes aos valores característicos $r_1, \ldots, r_p$, respectivamente. Supomos também que esses auto-valores estão em ordem decrescente. Seja $b = (b_1 \vdots b_p)$.

Consideramos a matriz

$$D_i = \text{diag}\ ({}^i s_{11}, \ldots, {}^i s_{pp}) \quad i=1,2,$$

como sendo a matriz diagonal cujo elemento ${}^i s_{jj}$, $j=1,\ldots,p$, é o desvio padrão amostral da j-ésima componente de ${}^i X$.

Definimos então a matriz de observações transformadas ${}^i y$ por:

$${}^i y = b^t \ D_i^{-1} \ {}^i x \quad , \ i=1,2.$$

Os estimadores de ν_i e Λ_i são:

$$\hat{\nu}_i = \bar{y}_i$$

$$\hat{\Lambda}_i = b^t \ R_i \ b = \begin{cases} I_2 & i = 1 \\ \text{diag}\ (r_1, \ r_p) & i = 2 \end{cases}$$

Para determinarmos a regra de classificação amostral, basta agora substituirmos os parâmetros ν_i e Λ_i na fórmula (4.1.4) por seus estimadores. Fazendo estas substituições, obtemos a seguinte regra:

alocamos x em π_1 se

$$
{}^{2}y_*^t \; \hat{\Lambda}_2^{-1} \; {}^{2}y_* - 2 \; {}^{2}y_*^t \; \hat{\Lambda}_2^{-1} \; \bar{y}_2 - {}^{1}y_*^t \; {}^{1}y_* - \bar{y}_1^t \; \bar{y}_1 +
$$

$$
+ \; 2 \; {}^{1}y_*^t \; \bar{y}_1 + \bar{y}_2^t \; \hat{\Lambda}_2^{-1} \; \bar{y}_2 + \ln |\hat{\Lambda}_2| \geq 0
$$

e, alocamos x em π_2 em caso contrário.

Para aplicarmos este método discriminante não é necessário nos atermos na escolha de qual população será considerada π_1 e qual será a π_2. Isso se deve ao fato de que os mesmos resultados serão obtidos se trabalharmos com a transformação baseada nos auto-vetores de $R_2^{-1}R_1$ pois os vetores característicos de uma matriz são idênticos aos da inversa dessa matriz, só que em ordem inversa. E ainda, os auto-valores da matriz inversa são os inversos dos auto-valores.

4.1 Eliminação de variáveis

Quando trabalhamos com dados experimentais, além de desejarmos classificar novas observações em um dos grupos pré-determinados, temos interesse em reduzir se possível, o número de medidas (variáveis) observadas. Isto significa que desejamos reduzir a dimensionalidade do problema.

Isto pode ser resolvido através da técnica multivariada de Componentes Principais que tem por objetivo determinar algumas combinações lineares das variáveis que seriam utilizadas para reduzir os dados.

O método discriminante proposto, pode ser utilizado com esse objetivo da seguinte maneira:

Como o vetor ${}^{i}y$ é bidimensional podemos fazer uma representação gráfica de seus componentes. Mais ainda, podemos colocar nesse mesmo gráfico as duas populações observadas, representadas por símbolos diferentes.

Suponhamos que no eixo vertical seja colocada a primeira componente de ${}^{i}y$, correspondendo a transformação feita a partir do auto-vetor referente ao maior auto-valor de $R_1^{-1}R_2$. Denominaremos este eixo de y_{max}. No eixo horizontal, colocamos a segunda componente e o denominaremos de y_{min}.

Se ao analisarmos este gráfico, virmos que as populações estão mais separadas no eixo y_{max} do que no eixo y_{mim}, por exemplo, temos então que o auto-vetor correspondente ao maior auto-valor é que está proporcionando essa discriminação.

Como estamos interessados em reduzir o número de variáveis, devemos, por conseguinte analisar os componentes do auto-vetor correspondente ao menor auto-valor.

Iniciamos, portanto, o processo de redução de dimensionalidade, eliminando a variável correspondente ao menor componente, em valor absoluto, do auto-vetor referente ao menor auto-valor.

Em seguida, aplicamos o método de discriminação baseado no quociente das matrizes de correlações. Se agora, as percentagens de classificações corretas aumentarem ou permanecerem constantes podemos continuar com o processo de eliminação de variáveis.

Se desejarmos continuar eliminando variáveis, devemos atentar que a segunda variável a ser eliminada será aquela referente ao segundo menor componente, sempre em valor absoluto, do auto-valor utilizado na eliminação da primeira variável.

Esse processo será continuado até que as percentagens de classificações corretas em cada uma das populações sejam inferior as obtidas originalmente.

Na descrição dos métodos de Análise Discriminante apresentados vimos que se existe a igualdade das matrizes de covariâncias então a regra discriminante obtida pelo método clássico é uma função linear.

Quando as matrizes de covariâncias são diferentes, a regra discriminante obtida pelo método clássico é uma funcão quadrática e vimos também que neste caso, podemos aplicar o método discriminante baseado no quociente das matrizes de covariâncias

Para o método de discriminação baseado no quociente das matrizes de correlações ser adequado à aplicação, é necessário que as matrizes de correlações sejam diferentes

No próximo capítulo apresentaremos os testes de hipóteses que permitem avaliar cada uma dessas situações.

5. Alguns testes de hipóteses utilizados com a Análise Discriminante

Quando vamos aplicar a Análise Discriminante em um conjunto de dados, várias pressuposições devem ser testadas antes de realizarmos a análise propriamente dita.

Nesta seção, apresentaremos os testes para verificar as hipóteses de igualdade das matrizes de covariâncias, igualdade das matrizes de correlações e finalmente o teste para a igualdade de médias.

5.1 Teste de igualdade de duas matrizes de covariâncias

Para utilizarmos a técnica clássica de Análise Discriminante, para o caso de duas populações multinormais, necessitamos primeiramente saber se as duas matrizes de covariâncias podem ser consideradas estatisticamente iguais ou não.

Se não existem evidências para rejeitarmos a hipótese de igualdade das matrizes de covariâncias, então, a função discriminante obtida é uma combinação linear das variáveis univariadas analisadas. No caso da hipótese ser rejeitada, a função discriminante será uma função quadrática dessas variáveis.

O teste de igualdade das matrizes de covariâncias também deve ser realizado quando pretendemos utilizar o método de discriminação baseado no quociente das matrizes de covariância, uma vez que este método só deve ser empregado nos casos onde rejeitamos a hipótese de homogeneidade das matrizes de covariâncias.

Seja ${}^{i}U$ a matriz de dados da população π_i e seja ${}^{i}X$ o vetor aleatório p-dimensional. Suponhamos que ${}^{i}X$ tem distribuição normal p-variada, isto é,

$$ {}^{i}X \sim N_p(\mu_i, \Sigma_i), \quad i=1,2. $$

Desejamos testar a hipótese de igualdade das matrizes de covariâncias que pode ser escrita como:

$$ H_0: \Sigma_1 = \Sigma_2 $$

$$ H_1: \Sigma_1 \neq \Sigma_2 . $$

Seja n_i o número de observações independentes obtidas da população π_i e seja S_i o estimador não viciado de Σ_i, $i=1,2$.

A estatística do teste modificado da razão de verossimilhança é:

$$V = (n_1 + n_2 - 2) \ln |S| - \sum_{i=1}^{2} (n_i - 1) \ln |S_i| ,$$

onde S é a matriz de covariância amostral combinada, isto é,

$$S = \frac{(n_1 - 1)\, S_1 + (n_2 - 1)\, S_2}{n_1 + n_2 - 2}$$

Box (1949) mostrou que se introduzirmos um fator de escala em V, obteremos uma quantidade com distribuição qui-quadrado quando n_1 e n_2 tendem ao infinito.

Esse fator de escala é dado por:

$$C = 1 - \frac{2p^2 + 3p - 1}{6\,(p + 1)} \left[\sum_{i=1}^{2} \frac{1}{(n_i - 1)} - \frac{1}{n_1 + n_2 - 2} \right]$$

e a quantidade $VC = V * C$, tem distribuição χ^2 com $0.5p(p+1)$ graus de liberdade quando n_1 e n_2 tendem ao infinito.

5.2 Teste de igualdade de duas matrizes de correlações

Suponhamos que R_1 e R_2 sejam as matrizes de correlação de amostras independentes, com n_i observações, das populações normais p-variadas.

Nosso objetivo é testar a hipótese de igualdade dessas matrizes, isto é, desejamos testar

$$H_0 : R_1 = R_2$$

$$H_1 : R_1 \neq R_2 .$$

Observamos que se a hipótese nula não é rejeitada não temos interesse em aplicar a Análise Discriminante baseada no quociente das correlações.

Larntz e Perlman (1985) propusseram um teste que tem por base a aplicação da transformação z de Fisher em cada um dos coeficientes de correlação r_{ij}.

Quando aplicaram a transformação z de Fisher em r_{ij}, obtiveram as quantidades ${}^{i}z_{jk}$ definidas por:

$${}^{i}z_{jk} = \frac{1}{2} \ln \left[\frac{1 + {}^{i}r_{jk}}{1 - {}^{i}r_{jk}} \right]$$

Seja

$$d = \frac{(n_1 - 3)\,(n_2 - 3)}{(n_1 + n_2 - 6)}$$

Portanto, definindo a estatística do teste T_3 por:

$$T_3 = d^{1/2} \max_{1 \leq j < k \leq p} \left| {}^{1}z_{jk} - {}^{2}z_{jk} \right| ,$$

tem-se que para qualquer $b > 0$

$$Pr \left[T_3 \leq b \right] \geq \left[\Phi(b) - \Phi(-b) \right]^{p(p-1)/2}$$

assintoticamente sob H_0, com $\Phi(c)$ representando o valor da distribuição Normal padrão calculada no ponto c.

Assim, o teste de nível α é dado por:

rejeitamos H_0 se $T_3 > b_\alpha$,

onde $b_\alpha > 0$ é escolhido tal que

$$\left[\Phi(b_\alpha) - \Phi(-b_\alpha) \right]^{p(p-1)/2} = 1 - \alpha$$

5.3 Teste de igualdade dos vetores de médias

Quando as evidências não indicam a rejeição da hipótese de igualdade das matrizes de covariâncias, é interessante testarmos também a igualdade dos vetores de médias. Se essa hipótese também não for rejeitada, então as duas populações normais podem ser consideradas iguais e neste caso, não temos interesse em utilizar a Análise Discriminante.

Seja ${}^{i}X$ o vetor aleatório com distribuição normal p-variada, isto é,

$$ {}^{i}X \sim N_p(\mu_i \ , \Sigma_i) \quad , \quad i=1,2. $$

Consideramos n_i como sendo o número de observações independente obtidas da população π_i , i=1,2.

Nesta seção, desejamos testar a hipótese:

$$ H_0: \ \mu_1 = \mu_2 $$

$$ H_0: \ \mu_1 \neq \mu_2 $$

supondo que $\Sigma_1 = \Sigma_2 = \Sigma$, porém desconhecida.

Seja a estatística T^2 de Hotteling definida por:

$$ T^2 = \frac{n_1 \ n_2}{n_1 + n_2} \ (\bar{x}_1 - \bar{x}_2)^t \ S^{-1} \ (\bar{x}_1 - \bar{x}_2) \ , $$

onde $\bar{x}_i$ é o vetor de médias da população π_i, i=1,2, e S é a matriz de covariância amostral combinada.

Temos então que a estatística F^* tem distribuição F com p e (n_1+n_2-p-1) graus de liberdade, onde:

$$F^* = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)\,p} \; T^2 .$$

Neste caso, temos a seguinte regra de decisão:

não rejeitamos H_0 se $F^* \leq F_{\alpha;p,(n_1+n_2-p-1)}$

onde $F_{\alpha;p,(n_1+n_2-p-1)}$ é o quantil de ordem $(1-\alpha)$ de F^*, sob H_0.

A metodologia da Análise Discriminante foi descrita nos capítulos 2, 3 e 4. Neste capítulo apresentamos os testes necessários para checarmos as hipóteses de igualdade das matrizes de covariâncias, de correlações e a igualdade dos vetores de médias.

Assim sendo, no próximo capitulo apresentaremos os resultados da aplicação dessa técnica em alguns conjuntos de dados.

6. Aplicações Práticas

Neste capítulo, apresentaremos a descrição dos conjuntos de dados analisados e para cada um desses conjuntos, os resultados dos testes das hipóteses de igualdade das matrizes de covariâncias, igualdade dos vetores de médias e igualdade das matrizes de correlações.

Aplicaremos a Análise Discriminante usando os métodos: clássico, baseado no quociente das matrizes de covariância e baseado no quociente das matrizes de correlações. Assim, para os dados analisados será possível compararmos esses três métodos.

6.1 Conjuntos de dados analisados

1) Flury (1988), descreve o conjunto de dados composto pelas medições realizadas em uma nota de mil francos suiços. Coletou-se uma amostra de cem notas falsas e uma amostra de cem notas verdadeiras.

Para cada uma dessas notas, as seguintes variáveis foram medidas:

$^{i}X_1$ = comprimento da nota
$^{i}X_2$ = largura do lado esquerdo
$^{i}X_3$ = largura do lado direito
$^{i}X_4$ = largura da margem inferior
$^{i}X_5$ = largura da margem superior

${}^{i}X_6$ = comprimento da diagonal.

As duas populações que formam o conjunto de dados são:

π_1 = notas verdadeiras

π_2 = notas falsas.

Por simplicidade, denotamos esse conjunto de dados por NOTAS.

O próximo conjunto de dados encontra-se em Roveratti (1979). Trata-se de um estudo sobre o efeito da castração de bovinos da raça Nelore, realizado em Onda Verde, no município de São José do Rio Preto. Foram amostrados 10 animais ao acaso em cada uma das seguintes populações:

π_1 = bovinos inteiros (tourinhos)

π_2 = bovinos castrados (novilhos)

Para estudar o efeito da castração de bovinos muitas variáveis foram coletadas. Porém, para aplicarmos a Análise Discriminante selecionamos apenas três subconjuntos dessas variáveis que são descritos a seguir.

2) O primeiro subconjunto é formado pelas seguintes variáveis:

${}^{i}X_1$ = peso vivo na fazenda (Kg)

${}^{i}X_2$ = peso dos ossos do crâneo (Kg)

${}^{i}X_3$ = peso dos ossos da mandíbula (Kg)
${}^{i}X_4$ = peso do cérebro (Kg)

Chamaremos a esse conjunto de dados de BOVINO.

3) O segundo subconjunto refere-se à gordura industrial e é formado pelas variáveis:

${}^{i}X_1$ = gordura peri-renal e pélvica (Kg)
${}^{i}X_2$ = gordura inguinal (Kg)
${}^{i}X_3$ = peso do epiloo (Kg)
${}^{i}X_4$ = gordura mesentérica (Kg)

A este conjunto de dados denominaremos de GORDI.

4) Finalmente, o terceiro subconjunto, referente a carnes aproveitáveis é composto por:

${}^{i}X_1$ = carnes aproveitáveis (Kg)
${}^{i}X_2$ = retalhos magros (Kg)
${}^{i}X_3$ = aparas de gorduras (Kg)
${}^{i}X_4$ = ossos (Kg)

Este conjunto será denominado de CARNEA.

5) O próximo conjunto de dados é formado por uma amostra de vinte isolados de *Phytophthora*, em cada uma das populações e foi fornecido pela Pesquisadora Dra. Edna Dora Martins Newman Luz da Divisão de Fitopatologia da CEPLAC.

Em cada uma dessas amostras, as seguintes variáveis foram observadas:

${}^{i}X_1$ = comprimento dos esporângios (μm)
${}^{i}X_2$ = largura dos esporângios (μm)
${}^{i}X_3$ = papila dos esporângios (μm)
${}^{i}X_4$ = pedicelo dos esporângios (μm)
${}^{i}X_5$ = diâmetro médio de lesões em cacau (mm).

As populações são:

π_1 = *Phytophthora capsici*
π_2 = *Phytophthora palmivora*.

A este conjunto de dados chamaremos de ISOLADOS.

Os próximos conjuntos de dados foram fornecidos pelo Professor Dr. Ademir Petenate, IMECC, UNICAMP e referem-se a um estudo sobre as medidas morfométricas de moscas. Uma amostra de quarenta observações foi coletada em cada uma das quatro regiões distintas do estado de São Paulo que serão denominadas de região 1, região 2, região 3 e região 4.

O primeiro conjunto de dados será denominado de ASA e é formado pelas medidas de distância de nervuras na asa sendo:

${}^{i}X_1$ = distância da primeira nervura
${}^{i}X_2$ = distância da segunda nervura
${}^{i}X_3$ = distância da terceira nervura
${}^{i}X_4$ = distância da quarta nervura
${}^{i}X_5$ = distância da quinta nervura
${}^{i}X_6$ = distância da sexta nervura

${}^{i}X_7$ = distância da sétima nervura

Como os métodos discriminantes: baseado no quociente das matrizes de covariâncias e baseado no quociente das matrizes de correlações só foram estudados para o caso de duas populações, esse conjunto de dados foi analisado considerando os seguintes subconjuntos:

6) ASA12 - formado pelas regiões 1 e 2
7) ASA13 - formado pelas regiões 1 e 3
8) ASA14 - formado pelas regiões 1 e 4
9) ASA23 - formado pelas regiões 2 e 3
10) ASA24 - formado pelas regiões 2 e 4
11) ASA34 - formado pelas regiões 3 e 4

No segundo conjunto de dados, denominado de CONCOR, foram tomadas as seguintes medidas:

${}^{i}X_1$ = medida da antena
${}^{i}X_2$ = comprimento do escapo
${}^{i}X_3$ = distância entre os olhos
${}^{i}X_4$ = altura dos olhos
${}^{i}X_5$ = comprimento da abertura bucal
${}^{i}X_6$ = comprimento do primeiro tarsômero
${}^{i}X_7$ = comprimento do segundo tarsômero
${}^{i}X_8$ = comprimento do terceiro tarsômero
${}^{i}X_9$ = comprimento do quinto tarsômero

Pelo mesmo motivo exposto para o conjunto de dados anterior, dividimos esse conjunto de dados nos seguintes subconjuntos:

12) CONCOR 1 e 2 = formado pelas regiões 1 e 2
13) CONCOR 1 e 3 = formado pelas regiões 1 e 3
14) CONCOR 1 e 4 = formado pelas regiões 1 e 4
15) CONCOR 2 e 3 = formado pelas regiões 2 e 3
16) CONCOR 2 e 4 = formado pelas regiões 2 e 4
17) CONCOR 3 e 4 = formado pelas regiões 3 e 4

6.2 Resultados obtidos

A tabela 6.2.1, dada a seguir, mostra os resultados dos testes de igualdade das matrizes de covariâncias, igualdade dos vetores de médias e igualdade das matrizes de correlações para cada conjunto de dados descrito em (6.1).

Os testes foram realizados com os níveis de significância de 5% e 1%, sendo seus resultados apresentados nas colunas denominadas de 5% e 1%.

A seguinte notação será utilizada:

SIG - indicando que o teste foi significativo, isto é, a hipótese nula foi rejeitada

NS - indicando que o teste foi não significativo, isto é, que a hipótese testada não foi rejeitada.

Tabela 6.2.1: Testes de igualdade de médias, de matrizes de covariâncias e de matrizes de correlações.

Conjuntos de dados	teste de covariâncias		teste de médias		teste de correlações	
	5%	1%	5%	1%	5%	1%
nota	sig	sig	-	-	sig	sig
bovino	ns	ns	sig	sig	sig	sig
carnea	ns	ns	sig	ns	ns	ns
gordi	ns	ns	sig	ns	sig	ns
isolados	sig	sig	-	-	sig	sig
asa 1 e 2	sig	ns	sig	sig	sig	sig
asa 1 e 3	sig	sig	-	-	sig	sig
asa 1 e 4	sig	sig	-	-	sig	sig
asa 2 e 3	sig	sig	-	-	sig	sig
asa 2 e 4	ns	ns	sig	sig	sig	sig
asa 3 e 4	sig	sig	-	-	sig	sig
concor 1 e 2	sig	sig	-	-	sig	sig
concor 1 e 3	sig	sig	-	-	sig	sig
concor 1 e 4	sig	sig	-	-	sig	sig
concor 2 e 3	sig	sig	-	-	sig	sig
concor 2 e 4	sig	sig	-	-	sig	sig
concor 3 e 4	sig	sig	-	-	sig	sig

Para podermos comparar os métodos de discriminação, separamos os conjuntos de dados apresentados na tabela 6.2.1, em quatro grupos.

O primeiro grupo é formado pelos conjuntos de dados onde o teste de igualdade das matrizes de covariâncias foi não significativo ao nível de 5%. O segundo grupo, apresenta os conjuntos de dados onde o teste foi significativo ao nível de 5%.

O terceiro grupo é formado pelos conjuntos de dados onde o teste de igualdade das matrizes de correlações foi não significativo ao nível de 5% e, finalmente, o quarto grupo é composto pelos conjuntos onde o teste foi significativo a 5%

Como vimos nos capítulos anteriores, o método de discriminação baseado no quociente das matrizes de covariâncias deve ser aplicado nos casos onde não existe a igualdade das matrizes de covariâncias. Já o método baseado no quociente das matrizes de correlações deve ser utilizado quando a igualdade das matrizes de correlações é rejeitada. Apesar dessas restrições aplicamos os métodos mesmo quando eles não eram recomendados a fim de verificar como se comportam nessas outras situações.

As tabelas 6.2.2, 6.2.3, 6.2.4, 6.2.5 e 6.2.6 apresentadas a seguir mostram as percentagens de classificacões corretas, calculadas por re-substituicão, para cada um dos métodos de discriminação.

A partir deste momento utilizaremos a seguinte notação nas tabelas:

Método (1) = método de discriminação baseado no quociente das matrizes de correlações

Método (2) = método de discriminação baseado no quociente das matrizes de covariâncias

Método (3) = método de discriminação clássico

Tabela 6.2.2: Conjuntos de dados onde o teste de igualdade das matrizes de covariâncias foi não significativo ao nível de 5%.

Conjuntos de dados	Percentagens de Classificação Correta pelo					
	Método (1)		Método (2)		Método (3)	
	π_1	π_2	π_1	π_2	π_1	π_2
bovino	90.0	90.0	90.0	70.0	100.0	100.0
carnea	90.0	80.0	60.0	60.0	90.0	80.0
gordi	70.0	100.0	90.0	60.0	80.0	90.0
asa 2 e 4	87.5	50.0	60.0	67.5	77.5	72.5

Tabela 6.2.3: Conjuntos de dados onde o teste de igualdade das matrizes de covariâncias foi significativo ao nível de 5%.

Conjuntos de dados	Percentagens de Classificação Correta pelo					
	Método (1)		Método (2)		Método (3)	
	π_1	π_2	π_1	π_2	π_1	π_2
nota	100.0	100.0	99.0	100.0	99.0	100.0
isolados	100.0	100.0	100.0	100.0	100.0	100.0
asa 1 e 2	60.0	72.5	62.5	72.5	82.5	87.5
asa 1 e 3	80.0	62.5	77.5	72.5	67.5	85.0
asa 1 e 4	85.0	65.0	72.5	87.5	85.0	87.5
asa 2 e 3	67.5	75.0	65.0	90.0	85.0	87.5
asa 3 e 4	75.0	70.0	85.0	77.5	90.0	95.0
concor 1 e 2	87.5	80.0	80.0	85.0	97.5	95.0
concor 1 e 3	100.0	77.5	82.5	92.5	100.0	97.5
concor 1 e 4	65.0	90.0	97.5	65.0	97.5	95.0
concor 2 e 3	85.0	77.5	80.0	75.0	100.0	95.0
concor 2 e 4	62.5	90.0	90.0	55.0	95.0	85.0
concor 3 e 4	82.5	72.5	95.0	70.0	95.0	82.5

Tabela 6.2.4: Conjuntos de dados onde o teste de igualdade das matrizes de correlações foi não significativo ao nível de 5%.

Conjuntos de dados	Percentagens de Classificação Correta pelo					
	Método (1)		Método (2)		Método (3)	
	π_1	π_2	π_1	π_2	π_1	π_2
carnea	90.0	80.0	60.0	60.0	90.0	80.0

Tabela 6.2.5: Conjuntos de dados onde o teste de igualdade das matrizes de correlações foi significativo ao nível de 5%.

Conjuntos de dados	Percentagens de Classificação Correta pelo					
	Método (1)		Método (2)		Método (3)	
	π_1	π_2	π_1	π_2	π_1	π_2
nota	100.0	100.0	99.0	100.0	99.0	100.0
bovino	90.0	90.0	90.0	70.0	100.0	100.0
gordi	70.0	100.0	90.0	60.0	80.0	90.0
isolados	100.0	100.0	100.0	100.0	100.0	100.0
asa 1 e 2	60.0	72.5	62.5	72.5	82.5	87.5
asa 1 e 3	80.0	62.5	77.5	72.5	67.5	85.0
asa 1 e 4	85.0	65.0	72.5	87.5	85.0	87.5
asa 2 e 3	67.5	75.0	65.0	90.0	85.0	87.5
asa 2 e 4	87.5	50.0	60.0	67.5	77.5	72.5
asa 3 e 4	75.0	70.0	85.0	77.5	90.0	95.0
concor 1 e 2	87.5	80.0	80.0	85.0	97.5	95.0
concor 1 e 3	100.0	77.5	82.5	92.5	100.0	97.5
concor 1 e 4	65.0	90.0	97.5	65.0	97.5	95.0
concor 2 e 3	85.0	77.5	80.0	75.0	100.0	95.0
concor 2 e 4	62.5	90.0	90.0	55.0	95.0	85.0
concor 3 e 4	82.5	72.5	95.0	70.0	95.0	82.5

Podemos resumir os resultados na seguinte tabela:

Tabela 6.2.6: Percentuais de classificações incorretas quando as hipóteses de igualdade de matrizes de covariâncias e correlações são rejeitadas ou não

	Percentagem de classificações incorretas usando		
	Método (1)	Método (2)	Método (3)
$\Sigma_1 = \Sigma_2$	23.57	32.86	18.57
$\Sigma_1 \neq \Sigma_2$	18.48	16.87	7.68
$\Psi_1 = \Psi_2$	15.00	40.00	15.00
$\Psi_1 \neq \Psi_2$	19.11	18.31	8.79

Antes de apresentarmos os resultados referentes a eliminação de variáveis, gostaríamos de lembrar que uma das vantagens do método baseado no quociente das matrizes de correlações e do baseado no quociente das covariâncias é o fato de que, independentemente, do número inicial de variáveis, sempre obtemos duas combinações lineares dessas variáveis para cada uma das populações. Se representarmos em um gráfico essas combinações teremos uma visão geométrica de como as populações foram discriminadas.

Esses mesmos gráficos permitem compararmos os dois métodos de discriminação para os conjuntos de dados analisados

O anexo E apresenta os gráficos obtidos para os conjuntos de dados em análise.

No eixo vertical foi colocado o componente y_{max} e no eixo horizontal o y_{min}.

Na tabela 6.2.7, dada a seguir, mostraremos os resultados da eliminação de algumas variáveis nos conjuntos de dados em análise.

Na coluna, Eliminação da variável, estão indicadas as variáveis retiradas da análise. Nas três colunas seguintes, apresentamos os resultados dos testes de igualdade de covariâncias, médias e de correlações, respectivamente.

Finalmente, na última coluna, mostramos as percentagens de classificações corretas, obtidas pelo método discriminante baseado no quociente das matrizes de correlações, após a eliminação das variáveis indicadas.

Tabela 6.2.7: Resultados dos testes e da análise discriminante baseada nas matrizes de correlações, após a eliminação de variáveis

Conjunto de dados	Eliminação da variável	$\Sigma_1 = \Sigma_2$		$\mu_1 = \mu_2$		$R_1 = R_2$		Percentagem de classificação correta	
		5%	1%	5%	1%	5%	1%	π_1	π_2
nota	X_3	sig	sig	-	-	sig	sig	97.0	97.0
bovino	X_1	sig	ns	-	-	sig	sig	60.0	90.0
carnea	X_4	ns	ns	sig	sig	ns	ns	90.0	70.0
gordi	X_2	ns	ns	sig	sig	sig	ns	70.0	100.0
	X_2 e X_3	ns	ns	ns	ns	ns	ns	60.0	80.0
isolados	X_3	sig	sig	-	-	sig	sig	100.0	100.0
	X_3 e X_2	sig	sig	-	-	sig	sig	100.0	100.0
	X_3, X_2, X_1	sig	sig	-	-	sig	sig	100.0	100.0
asa 1 e 2	X_6	sig	sig	-	-	sig	sig	50.0	80.0
asa 1 e 3	X_4	sig	sig	-	-	sig	sig	82.5	67.5
	X_2	sig	sig	-	-	sig	sig	72.5	57.5
	X_4 e X_5	sig	sig	-	-	sig	sig	77.5	57.5
asa 1 e 4	X_6	sig	sig	-	-	sig	sig	77.5	47.5
	X_3	sig	sig	-	-	sig	sig	85.0	65.0
	X_2 e X_3	sig	ns	sig	sig	sig	ns	77.5	50.0
asa 2 e 3	X_5	sig	ns	sig	sig	sig	sig	77.5	60.0
asa 2 e 4	X_6	sig	ns	sig	sig	sig	sig	92.5	37.5
asa 3 e 4	X_6	sig	ns	sig	sig	sig	sig	70.0	70.0
concor 1 e 2	X_6	sig	sig	-	-	sig	sig	90.0	82.5
	X_6 e X_3	sig	sig	-	-	sig	-	90.0	72.5

Tabela 6.2.7: continuação

Conjunto de dados	Eliminação da variável	$\Sigma_1 = \Sigma_2$		$\mu_1 = \mu_2$		$R_1 = R_2$		Percentagem de classificação correta	
		5%	1%	5%	1%	5%	1%	π_1	π_2
concor 1 e 3	X_1	sig	sig	-	-	sig	sig	100.0	72.5
concor 1 e 4	X_9	sig	sig	-	-	sig	sig	57.5	90.0
concor 2 e 3	X_1	sig	sig	-	-	sig	sig	85.0	82.5
	X_1 e X_7	sig	sig	-	-	sig	sig	90.0	80.0
	X_1, X_7, X_2	sig	sig	-	-	sig	sig	85.0	77.5
concor 2 e 4	X_1	sig	sig	-	-	sig	sig	62.5	87.5
concor 3 e 4	X_2	sig	sig	-	-	sig	sig	80.0	60.0

6.3 Análise dos resultados

Nesta seção, apresentaremos as conclusões obtidas para os conjuntos de dados analisados e também teceremos alguns comentários sobre o método de discriminação proposto.

Ao analisarmos a tabela 6.2.6 apresentada na seção anterior, observamos que quando não existem evidências para a hipótese de igualdade das matrizes de covariâncias ser rejeitada, o método clássico é o que tem o menor número de classificações incorretas. Em segundo lugar aparece o método baseado no quociente das matrizes de correlações e finalmente o método discriminante com base no quociente das matrizes de covariâncias.

Este resultado já era esperado uma vez que nesse caso o método clássico tem função discriminante linear e o método baseado nas covariâncias não deveria ser empregado.

Convém ressaltar que para esse grupo de dados, nada podemos afirmar sobre a igualdade ou não; estatisticamente falando; das matrizes de correlações pois a igualdade de covariâncias não implica necessariamente em igualdade de correlações.

Assim, ao considerarmos o grupo de dados onde a igualdade das covariâncias não foi rejeitada, estamos considerando os casos onde as matrizes de correlações são iguais ou diferentes indistintamente. Logo, estão aí incluídos os casos onde este método não deveria ser empregado.

Considerando agora o grupo de dados onde a igualdade das matrizes de covariâncias foi rejeitada, temos que o número de classificações incorretas foram 86, 189 e 207 para o método clássico, o baseado nas covariâncias e o baseado nas correlações, respectivamente.

Observamos que embora esse grupo seja aquele onde o método discriminante baseado no quociente das matrizes de covariâncias deveria apresentar o melhor resultado, as percentagens de classificações incorretas utilizando este método não diferem muito do método baseado nas matrizes de correlações. Os percentuais obtidos foram 16.87% e 18.48% de classificações erradas pelos métodos baseado nas covariâncias e baseado nas correlações, respectivamente.

Apesar de termos tido somente um conjunto de dados (CARNEA) onde a igualdade das correlações foi aceita apresentamos a conclusão da comparação dos métodos para esse conjunto.

Tanto o método clássico como o método baseado nas correlações apresentaram o mesmo número de classificações incorretas enquanto que o método baseado no quociente das covariâncias apresentou um número 2.6 vezes maior que os anteriores.

Finalmente, para o grupo onde a igualdade das matrizes de correlações foi rejeitada, teoricamente o melhor grupo para aplicarmos o método proposto, vemos que esse método foi o que apresentou o maior número de erros de classificação. O método clássico foi o que obteve o menor número de erros de classificação sendo seguido pelo método das covariâncias.

O método proposto não apresentou melhores resultados de reclassificação que o método clássico e o baseado no quociente das covariâncias, no entanto, foi o método que melhor discriminou as populações, como podemos ver nos gráficos apresentados no anexo E.

Esses gráficos mostram que para o conjunto de dados denominado NOTA as duas populações estão bem separadas e a reclassificação correta dos dados atingiu 100%.

Já para os dados de ASA 3 e 4 embora tenhamos as duas populações bem separadas, as percentagens de reclassificação incorreta foram de 25% e 30% para as regiões 3 e 4, respectivamente.

Para o conjunto CONCOR 2 e 4 temos que através do gráfico obtido pelo método baseado no quociente das correlações podemos observar que a matriz de dispersão da região 2 é circular enquanto que a dispersão da região 4 é elíptica.

A pior discriminação de populações ocorreu com as regiões 3 e 4 de CONCOR e os gráficos obtidos pelos dois métodos discriminantes (baseado no quociente das correlações e baseado nas covariâncias) são muito semelhantes.

Quanto as eliminações de variáveis temos as seguintes conclusões: para o conjunto de dados apresentados em Flury (1988) referente as notas de mil francos suiços, talvez pudéssemos eliminar a largura da margem inferior, uma vez que após essa eliminação obtivemos 98% e 100% de classificações corretas, nas populações de notas verdadeiras e falsas, respectivamente.

Para o conjunto de dados denominado BOVINO, podemos eliminar as variáveis peso dos ossos do crâneo e peso dos ossos da mandíbula e ainda continuar com as mesmas percentagens de classificação (90% e 90%) obtidas originalmente.

Após eliminarmos a variável gordura inguinal no conjunto de dados referentes à gordura industrial, obtemos as mesmas percentagens de classificação.

No conjunto de dados ISOLADOS, as medições de pedicelo de esporângios e diâmetro médio de lesões de cacau são suficientes para obtermos 100% de reclassificação corretas nas populações de *Phytophtora capsici* e *Phytophthora palmivora*.

Para ASA 1 e 3 podemos eliminar a quarta distância de nervuras melhorando desta forma os percentuais de classificações corretas de 70.0 e 57.5% para 82.5 e 67.5%, respectivamente.

As regiões 1 e 4 dos dados referentes a ASA mantiveram as mesmas percentagens ao eliminarmos a terceira distância entre nervuras.

Para os dados de CONCOR 1 e 2 melhoramos as percentagens de classificacão ao eliminarmos a medida do comprimento do primeiro tarsômero.

CONCOR 2 e 3 é o conjunto de dados que permitiu o maior número de eliminacões (três) mantendo os mesmos percentuais de classificação. Nesse caso as variáveis eliminadas foram medida da antena, comprimento do escapo e comprimento do segundo tarsômero. Se eliminarmos apenas as variáveis medida da antena e comprimento do segundo tarsômero melhoramos esses percentuais.

Os conjuntos ASA 1 e 2, ASA 2 e 3, ASA 2 e 4, ASA 3 e 4, CONCOR 1 e 3, CONCOR 1 e 4, CONCOR 2 e 4 e CONCOR 3 e 4 não permitem a redução de dimensionalidade sem que ocorra um aumento nos erros de classificação.

7. Resumo e alguns comentários adicionais

Neste capítulo, apresentaremos um resumo das técnicas apresentadas neste trabalho com breves comentários.

Três técnicas de discriminação foram apresentadas: método clássico (devido a Fisher, 1936), método baseado no quociente das matrizes de covariâncias (apresentado em Ticeran, 1988) e finalmente o método baseado nas matrizes de correlações.

O método de discriminação entre duas populações multivariadas efetuado pela comparação das dispersões (padronizadas ou não) pode ser útil quando as suas médias forem iguais ou quando houver interesse em investigar como se diferenciam do ponto de vista das dispersões.

Os resultados obtidos com os conjuntos de dados não são conclusivos para os métodos, mas fornecem uma indicação do comportamento desses métodos.

Porém, valeria a pena continuar investigando um pouco mais o método proposto, pois apesar de seus resultados não serem melhores que os outros métodos, para os conjuntos de dados analisados, a discriminação das populações foi melhor.

Isso nos leva a crer que talvez o problema esteja na escolha da forma de encontrar a função discriminante, ou nos cálculos dos auto-valores e auto-vetores uma vez que não apresentaram os valores esperados. Talvez ainda o problema esteja no poder do teste de igualdade das matrizes de correlações.

Estas dúvidas indicam um caminho de investigação que poderá ser seguindo com a continuação do trabalho aqui apresentado.

ANEXOS

A. ALGUMAS TÉCNICAS DE ESTATÍSTICA COMPUTACIONAL

Neste anexo apresentaremos alguns teoremas referentes a matrizes que quando aplicados minimizam o problema da precisão computacional.

Esses teoremas foram utilizados para determinar os auto-valores e auto-vetores das matrizes $R_1^{-1}R_2$ e $S_1^{-1}S_2$.

Teorema A.1: (Decomposição de Cholesky) Seja A uma matriz positiva definida. Então existe uma única matriz triangular inferior, T, cujos elementos da diagonal principal são positivos, tal que:

$$A = T\ T^t.$$

A demonstração deste teorema pode ser vista, por exemplo, em Anderson (1984), p.586.

Teorema A.2: (Decomposição do valor singular) Seja $A{:}n \times p$, uma matriz de posto n. Então A pode ser escrita como

$$A = U\ L\ V^t$$

onde $U{:}n \times r$ e $V{:}p \times r$ são matrizes ortogonais por colunas e $L{:}r \times r$ é uma matriz diagonal com elementos positivos.

Neste caso, temos que as colunas de U são os auto-vetores de $A\,A^t$ enquanto que as colunas de V são os auto-vetores de $A^t\,A$.

Encontramos a demonstração em Mardia (1979), p.473.

Teorema A.3: (Decomposição espectral) Qualquer matriz simétrica A:pxp, pode ser escrita na forma:

$$A = \Gamma\,\Lambda\,\Gamma^t$$

com Λ sendo a matriz diagonal cujos elementos são os auto-valores de A e Γ uma matriz ortogonal onde as colunas são os auto-vetores padronizados.

Podemos ver a demonstração deste teorema em Mardia (1979), p.469.

Mostraremos a seguir como foram determinados os auto-valores e auto-vetores de $R_1^{-1}R_2$.

A.1. Determinacão dos auto-valores e auto-vetores de $R_1^{-1}R_2$

Seja λ um auto-valor de $R_1^{-1}R_2$ e ν seu respectivo auto-vetor, isto é,

$$(R_2 - \lambda\ R_1)\ \nu = 0. \qquad (A.1.1)$$

Aplicando o teorema (A.1), decomposição de Cholesky, em R_1 temos que:

$$R_1 = G_1\ G_1^t \qquad (A.1.2)$$

Substituindo (A.1.2) em (A.1.1) obtemos

$$(R_2 - \lambda\ G_1 G_1^t\ )\ \nu = 0.$$

Temos ainda que:

$$(R_2 - \lambda\ G_1 G_1^t\ )\ \nu = 0 \ \Rightarrow\ (G_1^{-1} R_2 - \lambda\ G_1^t\ )\ \nu = 0. \qquad (A.1.3)$$

Portanto, chamando de ν^* o vetor $G_1^t\ \nu$ temos que a equação (A.1.3) pode ser escrita como:

$$(G_1^{-1} R_2 (G_1^t)^{-1} - \lambda\ I\ )\ \nu^* = 0. \qquad (A.1.4)$$

Desta forma, obtemos que λ e ν^* são os respectivos auto-valor e auto-vetor de $G_1^{-1}R_2(G_1^{-1})^t$

Para analisarmos um pouco mais a matriz $G_1R_2(G_1^{-1})^t$ aplicamos agora o teorema (A.1), decomposição de Cholesky, na matriz R_2, obtendo:

$$R_2 = G_2G_2^t \tag{A.1.5}$$

e

$$G_1^{-1}R_2(G_1^{-1})^t = G_1^{-1}G_2G_2^t (G_1^{-1})^t.$$

Chamando de T a matriz $G_1^{-1}G_2$, temos então que:

$$G_1^{-1}R_2(G_1^{-1})^t = T\ T^t.$$

Como $T\ T^t$ é uma matriz simétrica, aplicamos o teorema (A.3), decomposição espectral e obtemos:

$$T\ T^t = \Gamma \Lambda \Gamma^t, \tag{A.1.6}$$

onde os elementos de Λ são os auto-valores de $T\ T^t$ e as colunas de Γ seus respectivos auto-vetores.

Portanto, para encontrar os auto-valores de $R_1^{-1}R_2$ basta determinarmos os auto-valores de $G_1^{-1}R_2(G_1^{-1})^t$ usando a expressão (A.1.6).

Finalmente, para determinarmos os auto-vetores de $R_1^{-1}R_2$, basta encontrarmos

$$\nu = (G_1^t)^{-1}\,\nu^*$$

onde ν^* é o auto-vetor de $G_1^{-1}R_2(G_1^{-1})^t$ também determinado através de (A.1.6)

Ressaltamos também que os auto-valores e auto-vetores de $S_1^{-1}S_2$ foram determinandos desta mesma maneira.

A.2. Aproximação para a distribuição Normal

Para determinarmos computacionalmente o valor do ponto x_o tal que

$$Pr(X \leq x_o) = p_o$$

onde p_o é um valor conhecido e X é uma variável aleatória com distribuição normal com média 0 e variância 1, utilizamos o algorítmo de Beasley e Springer (1977).

Este algoritmo diz que devemos utilizar as seguintes aproximações para o valor x_0:

$$\text{se } 0.5 \leq p_0 \leq 0.92 \text{ então}$$

$$x_0 = q * \frac{(((a_3 * q_2 + a_2) * q_2 + a_1) * q_2 + a_0)}{((((b_4 * q_2 + b_3) * q_2 + b_2) * q_2 + b_1) * q_2 + 1)}$$

onde

$q = p_0 - 0.5$	$q_2 = q * q$
$a_0 = 2.50002823884$	$a_1 = -18.61500062529$
$a_2 = 41.39119773534$	$a_3 = -25.44106049637$
$b_1 = -8.47351093090$	$b_2 = 23.08336743743$
$b_3 = -21.06224101826$	$b_4 = 3.13082909833$
$c_0 = -2.78718931138$	$c_1 = -2.29796479134$
$c_2 = 4.85014127135$	$c_3 = 2.32121276858$
$d_1 = 3.54388924762$	$d_2 = 1.63706781897.$

$$\text{Se } 0.92 \leq p_0 \leq 1.0 \text{ então}$$

$$x_0 = \text{sinal}(q) * \frac{(((c_3 * r + c_2) * r + c_1) * r + c_0)}{((d_2 * r + d_1) * r + 1)}$$

onde:

$$r = \sqrt{-\ln(q)}$$

$$\text{sinal}(q) = \begin{cases} 1 & \text{se} \quad q \geq 0 \\ -1 & \text{se} \quad q < 0 \end{cases}$$

Mais detalhes desta aproximação podem ser vistos em Dachs (1988) , p.129.

A.3. Aproximação para a distribuição F

Para determinarmos o valor da distribuição F num ponto dado, utilizaremos a aproximação dada por Peizer e Pratt (1968).

Dachs (1988) recomenda que essa aproximação seja utilizada quando os graus de liberdade do denominador e do numerador sejam ambos maiores ou iguais a 10. Quando um dos graus de liberdade está entre 1 e 10, ele sugere que seja utilizada a aproximação de Mujander e Bhattacharjee (1973)

Como os erros de precisão são da ordem de 10^{-3} para a aproximação de Peizer e Pratt (1968) e o tempo de processamento é bem inferior ao da aproximação de Mujander e Bhattacharjee (1973), optamos pelo uso da primeira.

A aproximação de Peizer e Pratt (1968) para a distribuição F com μ e ν graus de liberdade é dada por:

$$S = \frac{1}{2} (\nu - 1)$$

$$T = \frac{1}{2} (\mu - 1)$$

$$n = \frac{\mu + \nu}{2} - 1$$

$$p = \frac{\nu}{\mu * x_o + \nu}$$

$$q = 1 - p$$

$$d_1 = S + \frac{1}{6} - (n + \frac{1}{3}) p$$

$$d_2 = d_1 + 0.02 \left[\frac{q}{S + 0.5} - \frac{p}{T + 0.5} + \frac{q - 5}{n + 1} \right]$$

$$z_i = d_i \left[\frac{1 + q * g \left(\frac{S}{n*p}\right) + p * g \left(\frac{T}{n*q}\right)}{\left(n + \frac{1}{6} \right) p * q} \right]^{\frac{1}{2}} \qquad i=1,2$$

onde z_1 e z_2 são os valores limites tal que $z_1 \leq p_o \leq z_2$.

A sub-rotina computacional utilizada foi adaptada de Dachs (1988).

A.4. Aproximação para a distribuição χ^2

Desejamos encontrar o valor p_o tal que

$$Pr\ (X \leq x_o) = p_o$$

onde X é uma variável aleatória com distribuição χ^2 com n graus de liberdade.

A rotina utilizada foi adaptada de Maindonald (1984), p.294, e utiliza a aproximação de Peizer e Pratt (1968) dada por:

$$d_1 = x_o - n + \frac{2}{3}$$

$$d_2 = d_1 - \frac{0.08}{n}$$

$$S = n - 1$$

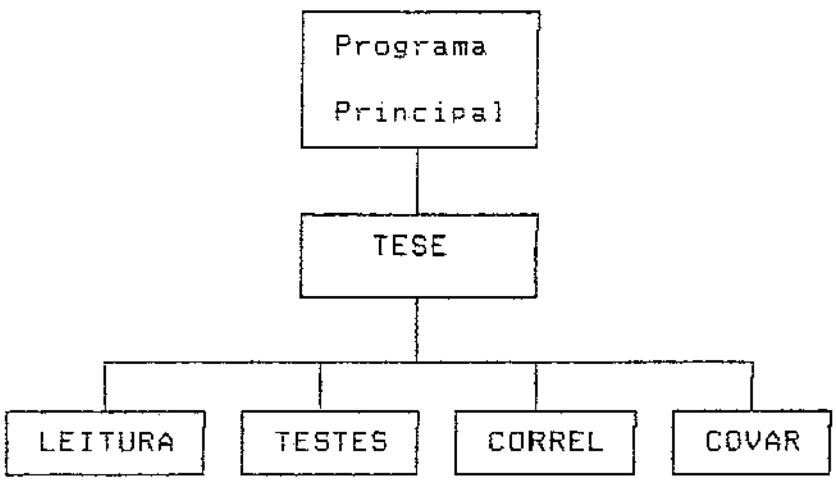
$$z_i = d_i + \left[\frac{1 + g(s \,/\, x_o)}{2\, x_o} \right]^{1/2} ,$$

onde z_i's são os valores limites tal que $z_1 \leq p_o \leq z_2$.

B. DESCRIÇÃO DO PROGRAMA COMPUTACIONAL E SUAS SUB-ROTINAS

No programa principal definimos a área necessária para o armazenamento dos dados utilizados nas sub-rotinas. As unidades lógicas de entrada e saída também estão definidas nesta parte do programa.

O sistema tem a seguinte estrutura:



A seguir apresentamos uma descrição das sub-rotinas utilizadas.

1) TESE: distribue a área alocada no programa principal entre todos os vetores e matrizes de dimensões variáveis.

Coordena a chamada das sub-rotinas necessárias para a leitura dos dados, realização dos testes de igualdade de médias, covariâncias e correlações e também chama as sub-rotinas necessárias para obter os resultados pelo método de Análise Discriminante baseada no quociente das matrizes de covariância (COVAR) e pelo método de discriminação baseado no quociente das matrizes de correlações (CORREL).

2) <u>LEITURA</u>: lê o título, o formato de leitura dos dados e as matrizes de dados das populações 1 e 2

<u>Entrada</u>:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

IN1 = unidade da unidade lógica de entrada que contém os dados referentes a população 1

IN2 = unidade da unidade lógica de entrada que contém os dados referentes a população 2

<u>Saída</u>:

X1 = matriz de dados da população 1, N1 x IP

X2 = matriz de dados da população 2, N2 x IP

NPOP = número de populações (sempre igual a 2)

POP1 = vetor de dimensão N1 contendo a classificação inicial das observações da população 1

POP2 = vetor de dimensão N2 contendo a classificação inicial das observações da população 2

TIT = vetor de dimensão 10 contendo o título a ser impresso nas listagens

FMT = vetor de dimensão 20 contendo o formato de leitura dos dados. Este mesmo formato será utilizado na impressão dos dados

3) TESTES: testa a igualdade dos vetores de médias, a igualdade das matrizes de covariâncias e de correlações

Entrada:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

X1 = matriz de dados da população 1, N1 x IP

X2 = matriz de dados da população 2, N2 x IP

XMED1 = vetor de médias da população 1, IP x 1

XMED2 = vetor de médias da população 2, IP x 1

S1 = matriz de covariância da população 1, IP x IP

S2 = matriz de covariância da população 2, IP x IP

S = matriz de covariância combinada, IP x IP

R1 = matriz de correlação da população 1, IP x IP

R2 = matriz de correlação da população 2, IP x IP

Saída:

TQUAD = estatística T^2

FTAB = valor da distribuição F com IGL1 e IGL2 graus de liberdade

IGL1 = número de graus de liberdade do numerador

IGL2 = número de graus de liberdade do denominador

F = valor da estaística F^*

VC = valor da estatística V * C

IGLQUI = número de graus de liberdade

QUI = valor da distribuição χ^2 com IGLQUI graus de liberdade

T3 = valor da estatística T_3

XO1 = valor de b_α com $\alpha = 0.05$

XO2 = valor de b_α com $\alpha = 0.01$

<u>Área de trabalho</u>:

AUX2 = vetor IP × 1

AUX3 = matriz IP × IP

AUX4 = matriz IP × IP

G1 = matriz IP × IP

4) <u>CORREL</u>: coordena a chamada das sub-rotinas necessárias para o método de discriminação baseado no quociente das matrizes de correlações. Coordena também a chamada das sub-rotinas de impressão e gráfico.

<u>Entrada</u>:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

N12 = número total de observações (= N1 + N2)

NPOP = número de populações (sempre igual a 2)

X1 = matriz de dados da população 1, N1 × IP

X2 = matriz de dados da população 2, N2 × IP

XMED1 = vetor de médias da população 1, IP × 1

XMED2 = vetor de médias da população 2, IP × 1

S1 = matriz de covariância da população 1, IP × IP

S2 = matriz de covariância da população 2, IP × IP

R1 = matriz de correlação da população 1, IP × IP

R2 = matriz de correlação da população 2, IP x IP

POP1 = vetor contendo a classificação inicial das observações da população 1, N1 x 1

POP2 = vetor com a classificação inicial das observações da população 2, N2 x 1

TQUAD = estatística T^2

FTAB = valor da distribuição F com IGL1 e IGL2 graus de liberdade

IGL1 = número de graus de liberdade do numerador

IGL2 = número de graus de liberdade do denominador

F = valor da estaística F^*

VC = valor da estatística $V * C$

IGLQUI = número de graus de liberdade

QUI = valor da distribuição χ^2 com IGLQUI graus de liberdade

T3 = valor da estatística T_3

XO1 = valor de b_α com $\alpha = 0.05$

XO2 = valor de b_α com $\alpha = 0.01$

TIT = vetor de dimensão 10, contendo o título a ser impresso nas listagens

FMT = vetor de dimensão 20 contendo o formato dos dados

IO = unidade lógica de saída

<u>Saída</u>:

Z1 = matriz de dados da população 1 padronizada, N1 x IP

Z2 = matriz de dados da população 2 padronizada, N2 x IP

Y1 = matriz transformada ${}^1y$, N1 x 2

Y2 = matriz transformada ${}^2y$, N2 x 2

VALOR = vetor dos auto-valores de $R_1^{-1}R_2$: IP x 1

VETOR = matriz dos auto-vetores: IP x IP

NN = matriz 2x2 contendo os números de observações na primeira linha e os símbolos a serem usados no gráfico na segunda linha

D1 = inversa da matriz de covariância de ${}^2y$: 2x2

YMED1 = vetor de médias de ${}^1y$: 2x1

YMED2 = vetor de médias de ${}^2y$: 2x1

YS1 = matriz de covariância de ${}^1y$: 2 x 2

YS2 = matriz de covariância de ${}^2y$: 2 x 2

YAUX1 = vetor de observações padronizadas segundo π_1: 1x2

YAUX2 = vetor de observações padronizadas segundo π_2: 1x2

<u>Área de trabalho</u>:

Y1EXT = matriz de dados de π_1 padronizada segundo π_2

Y2EXT = matriz de dados de π_2 padronizada segundo π_1

DP1 = matriz de desvio-padrão de π_1

DP2 = matriz de desvio-padrão de π_2

AUX3 = matriz IP x IP

AUX4 = matriz IP xIP

G1 = matriz IP x IP

YG = vetor (N1+N2) x 1

XGR = vetor com os elementos correspondentes a y_{min} nas duas populações: N1+N2 x 1

YGR = vetor com os elementos correspondentes a y_{max} nas duas populações: N1+N2 x 1

POPFIM = matriz com os resultados das classificações das observações: N1 x 2

INDORD = vetor (N1+N2) x 1

5) <u>COVAR</u>: coordena a chamada das sub-rotinas necessárias para o método de discriminação baseado no quociente das matrizes de covariâncias. Coordena também a chamada das sub-rotinas de impressão e gráfico.

<u>Entrada</u>:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

N12 = número total de observações (= N1 + N2)

NPOP = número de populações (sempre igual a 2)

X1 = matriz de dados da população 1, N1 x IP

X2 = matriz de dados da população 2, N2 x IP

S1 = matriz de covariância da população 1, IP x IP

S2 = matriz de covariância da população 2, IP x IP

POP1 = vetor contendo a classificação inicial das observações da população 1, N1 x 1

POP2 = vetor com a classificação inicial das observações da população 2, N2 x 1

TIT = vetor de dimensão 10, contendo o título a ser impresso nas listagens

FMT = vetor de dimensão 20 contendo o formato dos dados

IO = unidade lógica de saída

<u>Saída</u>:

Y1 = matriz transformada ${}^{1}y$, N1 x 2

Y2 = matriz transformada ${}^{2}y$, N2 x 2

VALOR = vetor dos auto-valores de $R_1^{-1}R_2$. IP x 1

VETOR = matriz dos auto-vetores: IP x IP

NN = matriz 2x2 contendo os números de observações na primeira linha e os símbolos a serem usados no gráfico na segunda linha

D1 = inversa da matriz de covariância de ${}^2y$: 2x2

YMED1 = vetor de médias de ${}^1y$: 2x1

YMED2 = vetor de médias de ${}^2y$: 2x1

YS1 = matriz de covariância de ${}^1y$: 2 x 2

YS2 = matriz de covariância de ${}^2y$: 2 x 2

YAUX1 = vetor de observações padronizadas segundo π_1: 1x2

YAUX2 = vetor de observações padronizadas segundo π_2: 1x2

Área de trabalho:

AUX3 = matriz IP x IP

AUX4 = matriz IP xIP

G1 = matriz IP x IP

YG = vetor (N1+N2) x 1

XGR = vetor com os elementos correspondentes a y_{min} nas duas populações: N1+N2 x 1

YGR = vetor com os elementos correspondentes a y_{max} nas duas populações: N1+N2 x 1

POPFIM = matriz com os resultados das classificações das observações: N1 x 2

INDORD = vetor (N1+N2) x 1

6) MEDIA: calcula o vetor de médias

Entrada:

NL = número de linhas da matriz XX

NC = número de colunas da matriz XX

XX = matriz de dados NL x NC

Saída:

XMED = vetor de médias NC x 1

7) COV: calcula a matriz de covariância amostral

Entrada:

NL = número de linhas de XX

NC = número de colunas de XX

XX = matriz de dados NL x NC

XMED = vetor de médias NC x 1

Saída:

SS = matriz de covariância amostral: NC x NC

8) TESTMED: teste de igualdade dos vetores de médias

Entrada:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

XMED1 = vetor de médias da população 1, IP x 1

XMED2 = vetor de médias da população 2, IP x 1

S = matriz de covariância combinada, IP x IP

Saída:

TQUAD = estatística T^2

F = valor da estatística F^*

FTAB = valor da distribuição F, com IGL1 e IGL2 graus de liberdade, no ponto F

IGL1 = número de graus de liberdade do numerador

IGL2 = número de graus de liberdade do denominador

Área de trabalho:

AUX2 = vetor IP x 1

AUX3 = matriz IP x IP

AUX4 = matriz IP x IP

G1 = matriz IP x IP

9) EFE1: calcula a distribuição F

Entrada:

GN = número de graus de liberdade do numerador

GD = número de graus de liberdade do denominador

XIS = valor no qual será calculada a distribuição F

Saída:

EFE1 = valor da distribuição F no ponto dado

10) APN4: calcula a distribuição normal padrão

Entrada:

XIS = valor no qual desejamos calcular a distribuição

<u>Saída</u>:

APN4 = valor da distribuição no ponto XIS

11) <u>GE</u>: calcula a função g de Peizer e Pratt (1968)

<u>Entrada</u>:

XIS = valor no qual desejamos calcular a função g

<u>Saída</u>:

GE = valor da função g no ponto XIS

12) <u>ST2</u>: determina a distribuição t

<u>Entrada</u>:

GL = número de graus de liberdade

XIS = valor no qual desejamos calcular a distribuição

<u>Saída</u>:

ST2 = valor da distribuição t com GL graus de liberdade no ponto XIS

13) <u>TCOV</u>: testa a igualdade de duas matrizes de covariâncias

Entrada:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

S1 = matriz de covariância da população 1: IP x IP

S2 = matriz de covariância da população 2: IP x IP

S = matriz de covariância combinada: IP x IP

Saída:

VC = valor da estatística V * C

IGLQUI = número de graus de liberdade

QUI = valor da distribuição χ^2 com IGLQUI graus de liberdade no ponto VC

Área de trabalho:

AUX3 = matriz IP x IP

G = matriz IP x IP

14) QUI2: calcula a distribuição χ^2

Entrada:

D2 = número de graus de liberdade da distribuição χ^2

X2 = ponto onde calcular o valor da distribuição χ^2

Saída:

QUI2 = valor da χ^2, com D2 graus de liberdade no ponto X2

15) <u>FN1</u>: calcula a aproximação do logarítmo neperiano da função gama

<u>Entrada</u>:

XIS = ponto onde desejamos calcular a função

<u>Saída</u>:

FN1 = valor da função no ponto XIS

16) <u>TESTCOR</u>: teste de igualdade das matrizes de correlações

<u>Entrada</u>:

N1 = número de observações da população 1
N2 = número de observações da população 2
IP = número de variáveis
R1 = matriz de correlação da população 1
R2 = matriz de correlação da população 2

<u>Saída</u>:

T3 = valor da estatística T_3

17) <u>INVNOR</u>: calcula a distribuição normal inversa, isto é, encontra o valor de x tal que $Pr(Z \leq x) \doteq q$, onde $Z \sim N(0,1)$ e q

é um valor dado

Entrada:

P0 = valor da distribuição, isto é, P0=Pr $(Z \leq X0)$

Saída:

X0 = valor tal que $Pr(Z \leq X0) = P0$

18) MULT: calcula o produto de duas matrizes compatíveis

Entrada:

NLX = número de linhas de X

NCX = número de colunas de X

NCY = número de colunas de Y

XX = matriz NLX x NCX

YY = matriz NCX x NCY

Saída:

ZZ = matriz produto de XX por YY, NLX x NCY

19) CHO: calcula a decomposição de Cholesky de uma matriz

Entrada:

NLC = número de linhas e colunas de XX

A = matriz NLC x NLC

Saída.

T = matriz triangular superior NLC x NLC

20) TRIANG: encontra a inversa da matriz triangular superior

Entrada:

NLC = número de linhas e colunas de XX

XX = matriz triangular superior: NLC x NLC

Saída:

XI = matriz inversa de XX: NLC x NLC

21) TRANSP: encontra a transposta de uma matriz

Entrada:

NL = número de linhas de XX

NC = número de colunas de XX

XX = matriz de dados NL x NC

Saída:

XXL = matriz transposta de XX: NC x NL

22) DVS: faz a decomposição do valor singular (DVS) de uma matriz

Entrada:

IP = número de linhas e colunas de U

U = matriz na qual aplicaremos a DVS: IP x IP

Saída:

D = vetor contendo os auto-valores de U: IP x 1

U = matriz dos auto-vetores (em colunas): IP x IP

23) YCALC: calcula as matrizes ${}^{i}y$, i=1,2

Entrada:

*N*1 = número de observações da população 1

*N*2 = número de observaçoes da população 2

IP = número de variáveis

*ZZ*1 = matriz padronizada: N1 x IP

*ZZ*2 = matriz padronizada: N2 x IP

VETOR = matriz IP x IP dos auto-vetores

Saída:

*YY*1 = matriz transformada ${}^{1}y$: N1 x 2

*YY*2 = matriz transformada ${}^{2}y$: N2 x 2

24) DISCOR: calcula a função discriminante para o método baseado no quociente das correlações

Entrada:

*YAUX*1 = vetor de observação padronizada segundo π_1

1x2

YAUX2 = vetor de observação padronizada segundo π_2: 1x2

YMED1 = vetor de médias de ${}^1y$: 2 x 1

YMED2 = vetor de médias de ${}^2y$: 2 x 1

D1 = inversa da matriz de covariância de ${}^2y$: 2 x 2

DET = valor de $\ln |\hat{\Lambda}_2|$

<u>Saída</u>:

FUNCAO = valor da função discriminante

25) <u>IMPCOR</u>: imprime as matrizes de dados de entrada, os vetores de médias, as matrizes de covariâncias e de correlações. Os resultados dos testes de igualdade das matrizes de covariâncias, dos vetores de médias e das matrizes de correlações também são impressos. Esta sub-rotina imprime ainda os resultados da análise discriminante baseada no quociente das matrizes de correlações.

<u>Entrada</u>:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

NPOP = número de populações analisadas (sempre igual a 2)

FMT = vetor de 20 posições, contendo o formato de leitura dos dados. Utilizamos este mesmo formato para a impressão dos dados

X1 =matriz de dados da população 1: N1 x IP

X2 =matriz de dados da população 2: N2 x IP

XMED1 = vetor de médias da população 1: IP x 1

XMED2 = vetor de médias da população 2: IP x 1

S1 = matriz de covariância da população 1: IP x IP

S2 = matriz de covariância da população 2: IP x IP

R1 = matriz (IP x IP) de correlação da população 1

R2 = matriz (IP x IP) de correlação da população 2

VALOR = vetor dos auto-valores de $R_1^{-1}R_2$: IP x 1

VETOR = matriz IP x IP contendo os auto-valores

Y1 = matriz N1 x 2, referente a população 1, armazenando os valores:

$y1(i,1) = \beta(1)\ {}^1z$

$y1(i,2) = \beta(IP)\ {}^1z$

Y2 = matriz N2 x 2, referente a população 2, armazenando os valores:

$y2(i,1) = \beta(1)\ {}^2z$

$y2(i,2) = \beta(IP)\ {}^2z$

YMED1 = vetor 2 x 1 contendo as médias de ${}^1y$

YMED2 = vetor 2 x 1 contendo as médias de ${}^2y$

YS1 = matriz 2 x 2 das covariâncias de ${}^1y$

YS2 = matriz das covariâncias de ${}^2y$: 2 x 2

POP1 = vetor de dimensão N1 . Contém a classificação inicial das observações da população 1

POP2 = vetor (N2 x 1) contendo a classificação inicial dos elementos da população 2

POPFIM = matriz N1 x 2 com os resultados das classificações dos elementos das populações, por coluna

PERC1 = valor da percentagem de classificações corretas na população 1

PERC2 = percentagem de classificação correta na população 2

TIT = vetor de dimensão 10, contendo o título a ser impresso nas listagens

T3 = estatística do teste de igualdade de correlacões

X01 = valor de b_α com $\alpha = 0.05$

X02 = valor de b_α, com $\alpha = 0.01$

TQUAD = valor da estatística T^2 do teste de igualdade das covariâncias

FTAB = valor da estatística F^* do teste de iguladade das matrizes de covariâncias

IGL1 = número de graus de liberdade do denominador

IGL2 = número de graus de liberdade do numerador

F = valor da distribuição F com IGL1 e IGL2 graus de liberdade no ponto FTAB

IO = número da unidade lógica de saída

IGLQUI = número de graus de liberdade da estatística VC

VC = valor da estatística do teste de igualdade das matrizes de correlações

QUI = valor da distribuição χ^2, com IGLQUI graus de liberdade, no ponto VC

26) GRAF: faz um gráfico de duas dimensões.

Entrada:

YGR = vetor de dimensão N1 + N2 contendo os elementos correspondentes a y_{max} nas duas populações

XGR = vetor de dimensão N1 + N2 contendo os elementos correspondentes a y_{min} nas duas populações

NXY = matriz 2 x 2 onde:

NXY(1,1) = número de observações da população 1

NXY(1,2) = número de observações da população 2

NXY(2,1) = '+' = símbolo usado para representar a população 1

NXY(2,2) = 'O' = símbolo usado para representar a população 2

TIT = vetor 10 x 1 contendo o título a ser impresso no gráfico

<u>Área de trabalho</u>:

YG = vetor (N1+N2) x 1

INDORD = vetor (N1+N2) x 1

27) <u>SORTV2</u>: faz a ordenação, em ordem crescente ou decrescente de um vetor

<u>Entrada</u>:

NL = número de elementos

YG = vetor a ser ordenado, NL x 1

ORDER = constante inteira tal que:

se ORDER ≥ 0 então ordem crescente

se ORDER < 0 então ordem decrescente

<u>Saída</u>:

YG = vetor ordenado

INDORD = vetor contendo as posições originais dos elementos, NL x 1

28) <u>DISCOV</u>: calcula o valor da função discriminante, usando o método baseado no quociente das covariâncias

<u>Entrada</u>:

YYAUX = matriz 1 x 2, contendo uma observação transformada

D1 = inversa da matriz de covariância de ${}^{2}w$

YMED1 = vetor (2 x 1) de médias de ${}^{1}w$

YMED2 = vetor (2x1) de médias de ${}^{2}w$

CONST = valor de $\hat{b}$

Saída:

FUNCAO = valor da função discriminante

29) IMPCOV: imprime os resultados do método discriminante baseado no quociente das matrizes de covariâncias

Entrada:

N1 = número de observações da população 1

N2 = número de observações da população 2

IP = número de variáveis

NPOP = número de populações analisadas (sempre igual a 2)

VALOR = vetor dos auto-valores de $S_1^{-1}S_2$: IP x 1

VETOR = matriz IP x IP contendo os auto-valores

Y1 = matriz N1 x 2, referente a população 1, armazenando os valores:

$y1(i,1) = \gamma(1)\ {}^{1}x$

$y1(i,2) = \gamma(IP)\ {}^{1}x$

Y2 = matriz N2 x 2, referente a população 2, armazenando os valores:

$y2(i,1) = \gamma(1)\ {}^{2}x$

$y2(i,2) = \gamma(IP)\ {}^{2}x$

YMED1 = vetor 2 x 1 contendo as médias de ${}^{1}y$

YMED2 = vetor 2 x 1 contendo as médias de ${}^{2}y$

YS1 = matriz 2 x 2 das covariâncias de ${}^{1}y$

YS2 = matriz das covariâncias de ${}^{2}y$

POP1 = vetor de dimensão N1 . Contém a classificação inicial das observações da população 1

POP2 = vetor (N2 x 1) contendo a classificação inicial dos elementos da população 2

POPFIM = matriz N1 x 2 com os resultados das classificações dos elementos das populações, por coluna

PERC1 = valor da percentagem de classificações corretas na população 1

PERC2 = percentagem de classificação correta na população 2

TIT = vetor de dimensão 10, contendo o título a ser impresso nas listagens

IO = número da unidade lógica de saída

ATENÇÃO LEITOR

Caso tenha interesse em obter, em disquete, cópia do programa computacional e suas subrotinas entrar em contato com:

Maria Ivete B. Brugnerotto
CEPLAC - CEPEC - INFES
caixa postal 7
45.600 Itabuna - Ba

Tel.: (073) 214-3221

C. LISTAGENS DO PROGRAMA PRINCIPAL E SUAS SUB-ROTINAS

PROGRAMA PRINCIPAL

```
      dimension a(7000),ymed1(2),ymed2(2),ys1(2,2),ys2(2,2)
      dimension d1(2,2),nn(2,2),yaux1(2),yaux2(2)
      integer fmt(20), tit(10), iv(1000)

      double precision a, ymed1,ymed2,ys1,ys2,d1,yaux1,yaux2

c  programa para o calculo das funcoes discriminante usado os metodos:
c     baseado no quociente das correlacoes e das covariancias


c   definindo as unidades de entrada e saida


      in1 = 3
      in2 = 4
      io = 7

      open(in1, file = ' ')
      open(in2, file = ' ')
      open(io, file = ' ', recl=80)


c  entrada:
c    n1 = numero de observacoes da populacao 1
c    n2 = numero de observacoes da populacao 2
c    ip = numero de variaveis
c    fmt = formata de leitura dos dados
c    tit = titulo a ser impresso


      read(in1,1) n1,n2
 1    format( 2(1x,i4) )

      read(in1,2) ip
 2    format(1x,i4)


c  chamando a subrotina principal que depois de distribuir area
c  faz todos os calculos e impressoes dos resultados
```

```
      n12 = n1 + n2

c  definindo os tamanhos dos vetores e matrizes inteiras

       j1 = 1
      j2 = j1 + n1
      j3 = j2 + n2
      j4 = j3 + (2 * n1)

      jfim = j4 + n12

      if (jfim .gt. 1000) then
        print 50
 50     format(10x,'as dimensoes das matrizes e vetores inteiras',/,
     *5x,'ultrapassaram o limite.',/,
     *10x,'favor redimencionar este limite',//)
        goto 100
      endif
c  definindo os tamanhos das matrizes e vetores reais

      ia1 = n1 * ip
      ia2 = n2 * ip
      ia3 = ip * ip

      i1 = 1
      i2 = i1 + ia1
      i3 = i2 + ia2
      i4 = i3 + ia1
      i5 = i4 + ia2
      i6 = i5 + (n1 * 2)
      i7 = i6 + (n2 * 2)
      i8 = i7 + (n1 * 2)
      i9 = i8 + (n2 * 2)
      i10 = i9 + ip
      i11 = i10 + ip
      i12 = i11 + ia3
      i13 = i12 + ia3
      i14 = i13 + ia3
      i15 = i14 + ia3
      i16 = i15 + ia3
      i17 = i16 + ia3
      i18 = i17 + ia3
      i19 = i18 + ip
      i20 = i19 + ia3
      i21 = i20 + ia3
      i22 = i21 + ip
      i23 = i22 + ia3
      i24 = i23 + ia3
      i25 = i24 + n12
      i26 = i25 + n12
```

```
      ifim = i26 + n12
      if(ifim .gt. 7000) then
       print 55
 55    format(10x,'as dimensoes reais ultrapassaram o limite superior.',
     */,10x,'favor redefinir a dimensao')
       goto 100
      endif

      call tese(n1,n2,ip,n12,a(i1),a(i2),a(i3),a(i4),a(i5),a(i6),
     1a(i7),a(i8),a(i9),a(i10),a(i11),a(i12),a(i13),a(i14),a(i15),
     2a(i16),a(i17),a(i18),a(i19),a(i20),a(i21),a(i22),a(i23),a(i24),
     3a(i25),a(i26), iv(j1),iv(j2),iv(j3),iv(j4),
     4nn,d1,ymed1,ymed2,ys1,ys2,yaux1,yaux2,tit,fmt,in1,in2,io )

  100 stop
      end
```

SUBROTINA: TESE

```
      SUBROUTINE TESE(N1,N2,IP,N12,X1,X2,Z1,Z2,Y1,Y2,Y1EXT,Y2EXT,
     *XMED1,XMED2,S1,S2,S,DP1,DP2,R1,R2,AUX2,AUX3,AUX4,VALOR,VETOR,
     *G1,YG,XGR,YGR, POP1,POP2,POPFIM,INDORD,NN,D1,YMED1,YMED2,YS1,YS2,
     *YAUX1,YAUX2,TIT,FMT,IN1,IN2,IO)

C
C     PROGRAMA   UTILIZANDO   METODO DAS CORRELACOES E DAS COVARIANCIAS
C

      DIMENSION X1(N1,IP), X2(n2,IP), Z1(N1,IP), Z2(n2,IP)
      DIMENSION Y1(N1,2), Y2(n2,2),Y1EXT(N1,2),Y2EXT(n2,2)
      DIMENSION XMED1(IP), XMED2(IP), S1(IP,IP), S2(IP,IP), S(IP,IP)
      DIMENSION DP1(IP,IP), DP2(IP,IP), R1(IP,IP),R2(IP,IP)
      DIMENSION VALOR(IP), VETOR(IP,IP), G1(IP,IP), NN(2,2),D1(2,2)
      DIMENSION YG(N12), INDORD(N12), XGR(N12), YGR(N12)
      DIMENSION YMED1(2), YMED2(2),YS1(2,2),YS2(2,2),YAUX1(2),YAUX2(2)
      DIMENSION AUX2(IP),AUX3(IP,IP), AUX4(IP,IP)
      INTEGER TIT(10), FMT(20),POP1(N1), POP2(N2), POPFIM(N1,2)

      DOUBLE PRECISION X1,X2,XMED1,XMED2,S,S1,S2,R1,R2,Z1,Z2,DP1,DP2,
     *VALOR,VETOR,X01,X02,P0,Y1,Y2,YMED1,YMED2,YS1,YS2,DD,DET,D1,F,
     *FUNCAO,CONST1,CONST2,CONST,YAUX1,YAUX2,AUX2,AUX3,AUX4,XGR,YGR,YG,
     *Y1EXT,Y2EXT,G1,T3,ALFA,XPROB,TQUAD,FTAB,VC,QUI

      NN(2,1) = '+'
```

```
      NN(2,2) = 'O'

C  LEITURA DOS DADOS

      CALL LEITURA(N1,N2,IP,X1,X2,NPOP,POP1,POP2,TIT,FMT,IN1,IN2)


C  TESTES DE IGUALDADE DE MEDIAS, COVARIANCIAS E CORRELACOES

      CALL TESTES(N1,N2,IP,X1,X2,XMED1,XMED2,S1,S2,S,R1,R2,AUX2,AUX3,
     *      AUX4,G1,TQUAD,FTAB,IGL1,IGL2,F,VC,IGLQUI,QUI,T3,X01,X02)


C  METODO DISCRIMINANTE BASEADO NAS MATRIZES DE CORRELACOES

      CALL CORREL(N1,N2,IP,N12,NPOP,X1,X2,Z1,Z2,Y1,Y2,Y1EXT,Y2EXT,XMED1,
     *      XMED2,S1,S2,DP1,DP2,R1,R2,AUX3,AUX4,VALOR,VETOR,G1,YG,XGR,
     *      YGR,POP1,POP2,POPFIM,INDORD,NN,D1,YMED1,YMED2,YS1,YS2,YAUX1,
     *      YAUX2,TQUAD,FTAB,IGL1,IGL2,F,VC,IGLQUI,QUI,T3,X01,X02,
     *      TIT,FMT,IO)


C  METODO DISCRIMINANTE BASEADO NAS MATRIZES DE CAVARIANCIAS

      CALL COVAR(N1,N2,IP,N12,NPOP,X1,X2,Y1,Y2,S1,S2,AUX3,AUX4,VALOR,
     *      VETOR,G1,YG,XGR,YGR,POP1,POP2,POPFIM,INDORD,NN,D1,YMED1,
     *      YMED2,YS1,YS2,YAUX1,YAUX2,TIT,IO)


      RETURN
      END
```

Subrotina: LEITURA

```
      SUBROUTINE LEITURA(N1,N2,IP,X1,X2,NPOP,POP1,POP2,TIT,FMT,IN1,IN2)

C
C     LEITURA DOS DADOS
C
```

```
      DIMENSION X1(N1,IP), X2(N2,IP)
      INTEGER TIT(10), FMT(20),POP1(N1), POP2(N2)

      DOUBLE PRECISION X1,X2


C
C  TIT = TITULO A SER IMPRESSO
C
      READ(IN1,3) (FMT(I), I=1,20)
 3    format(20a4)

      READ(IN1,31) (TIT(I), I=1,10)
 31   FORMAT(10A4)

C
C   X1 = MATRIZ DE DADOS DA POP 1  X1(N1,P)
C
      DO 10 I=1,N1
        READ(IN1,FMT) (X1(I,J), J=1,IP) , POP1(I)
 10   CONTINUE
C
C    X2 = MATRIZ DE DADOS DA POP 2
C
      DO 15 I=1,N2
        READ(IN2,FMT) (X2(I,J), J=1,IP) , POP2(I)
 15   CONTINUE
C
C   NUMERO DE POPULACAOES - NPOP
C
      NPOP = 2


      RETURN
      END
```

SUBROTINA: TESTES

```
      SUBROUTINE TESTES(N1,N2,IP,X1,X2,XMED1,XMED2,S1,S2,S,R1,R2,
     *                  AUX2,AUX3,AUX4,G1,TQUAD,FTAB,IGL1,IGL2,F,
     *                  VC,IGLQUI,QUI,T3,X01,X02)

C
C    ROTINA PARA REALIZAR OS TESTES DE: IGUALDADE DE MEDIAS
C                                       IGUALDADE DE COVARIANCIAS
C                                       IGUALDADE DE CORRELACOES
```

```
C

      DIMENSION X1(N1,IP), X2(N2,IP), R1(IP,IP), R2(IP,IP)
      DIMENSION XMED1(IP), XMED2(IP), S1(IP,IP), S2(IP,IP), S(IP,IP)
      DIMENSION AUX2(IP),AUX3(IP,IP), AUX4(IP,IP), G1(IP,IP)


      DOUBLE PRECISION X1,X2,XMED1,XMED2,S,S1,S2,R1,R2,X01,X02,P0,F,
     *       AUX2,AUX3,AUX4,G1,T3,ALFA,XPROB,TQUAD,FTAB,VC,QUI


C
C  CALCULANDO AS MEDIAS DE X
C
      CALL MEDIA(N1,IP,X1,XMED1)
      CALL MEDIA(N2,IP,X2,XMED2)
C
C  CALCULANDO AS COVARIANCIAS DE X (DIVIDIDA POR N-1 )
C
      CALL COV(N1,IP,X1,XMED1,S1)
      CALL COV(N2,IP,X2,XMED2,S2)


C
C        CALCULANDO AS CORRELACOES DE X:     R = COV(X,Y) / (SX * SY)
C

      DO 32 I=1,IP
        DO 33 J=I,IP
        R1(I,J) = S1(I,J) / ( DSQRT(S1(I,I) * S1(J,J)) )
 33     R2(I,J) = S2(I,J) / ( DSQRT(S2(I,I) * S2(J,J)) )
 32   CONTINUE

       DO 34 I=1,IP
        DO 36 J=I,IP
          R1(J,I) = R1(I,J)
          R2(J,I) = R2(I,J)
 36         CONTINUE
 34      CONTINUE

C
C    TESTE DE IGUALDADE DE MEDIAS
C
      XN1 = FLOAT(N1 - 1)
      XN2 = FLOAT(N2 - 1)
      DO 41 I=1,IP
        DO 42 J=I,IP
          S(I,J) = ( XN1 * S1(I,J) + XN2 * S2(I,J) ) / ( XN1 + XN2 )
 42        CONTINUE
 41       CONTINUE
```

```
      DO 47 I=1,IP
       DO 48 J=I,IP
         S(J,I) = S(I,J)
48         CONTINUE
47       CONTINUE


      CALL TESTMED(N1,N2,IP,XMED1,XMED2,S,TQUAD,FTAB,AUX2,AUX3,AUX4,G1,
     *IGL1,IGL2,F)


C
C   TESTE DE IGUALDADE DAS MATRIZES DE COVARIANCIAS
C

       CALL TCOV(N1,N2,IP,S1,S2,S,VC,IGLQUI,QUI,AUX3,G1)


C
C    TESTANDO IGUALDADE DE CORRELACOES
C

      CALL TESTCOR(N1,N2,IP,R1,R2,T3)

      XIP = FLOAT( IP * (IP - 1) ) / 2.
      P0 = 1.0d0 + (1.0d0 - 0.05d0) ** XIP
      P0 = P0 / 2.0d0

      CALL INVNOR(P0,X01)

      P0 = 1.0d0 + (1.0d0 - 0.01d0) ** XIP
      P0 = P0 / 2.0d0

      CALL INVNOR(P0,X02)


      RETURN
      END
```

Subrotina: CORREL

```
      SUBROUTINE CORREL(N1,N2,IP,N12,NPOP,X1,X2,Z1,Z2,Y1,Y2,Y1EXT,Y2EXT,
     *XMED1,XMED2,S1,S2,DP1,DP2,R1,R2,AUX3,AUX4,VALOR,VETOR,G1,YG,XGR,
     *YGR, POP1,POP2,POPFIM,INDORD,NN,D1,YMED1,YMED2,YS1,YS2,YAUX1,
     *YAUX2,TQUAD,FTAB,IGL1,IGL2,F,VC,IGLQUI,QUI,T3,X01,X02,TIT,FMT,IO)
```

```
C
C    PROGRAMA   UTILIZANDO   METODO DAS CORRELACOES E DAS COVARIANCIAS
C


      DIMENSION X1(N1,IP), X2(n2,IP), Z1(N1,IP), Z2(n2,IP)
      DIMENSION Y1(N1,2), Y2(n2,2),Y1EXT(N1,2),Y2EXT(n2,2)
      DIMENSION XMED1(IP), XMED2(IP), S1(IP,IP), S2(IP,IP)
      DIMENSION DP1(IP,IP), DP2(IP,IP), R1(IP,IP),R2(IP,IP)
      DIMENSION VALOR(IP), VETOR(IP,IP), G1(IP,IP), NN(2,2),D1(2,2)
      DIMENSION YG(N12), INDORD(N12), XGR(N12), YGR(N12)
      DIMENSION YMED1(2), YMED2(2),YS1(2,2),YS2(2,2),YAUX1(2),YAUX2(2)
      DIMENSION AUX3(IP,IP), AUX4(IP,IP)
      INTEGER TIT(10), FMT(20),POP1(N1), POP2(N2), POPFIM(N1,2)

      DOUBLE PRECISION X1,X2,XMED1,XMED2,S1,S2,R1,R2,Z1,Z2,DP1,DP2,
     *VALOR,VETOR,X01,X02,P0,Y1,Y2,YMED1,YMED2,YS1,YS2,DD,DET,D1,F,
     *FUNCAO,CONST1,CONST2,CONST,YAUX1,YAUX2,AUX3,AUX4,XGR,YGR,YG,
     *Y1EXT,Y2EXT,G1,T3,ALFA,XPROB,TQUAD,FTAB,VC,QUI

      NN(2,1) = '+'
      NN(2,2) = 'O'



C
C   METODO DA CORRELACAO:  INICIO
C


C
C         CALCULANDO DINV=DIAG( 1/ SQRT(S(I,I)) )
C
      DO 35 I=1,IP
        DO 40 J=1,IP
          DP1(I,J) = 0.0D0
          DP2(I,J) = 0.0D0
 40   CONTINUE
 35   CONTINUE
      DO 45 I=1,IP
        DP1(I,I) = 1.0D0 / ( DSQRT(S1(I,I)) )
        DP2(I,I) = 1.0D0 / ( DSQRT(S2(I,I)) )
 45   CONTINUE

C
C    CALCULANDO Z = INV(D) X
C

      CALL MULT(N1,IP,IP,X1,DP1,Z1)
      CALL MULT(N2,IP,IP,X2,DP2,Z2)
C
```

```
C  AUTO-VALORES DE R1 -1 * R2
C

      DO 37 I=1,IP
        DO 38 J=1,IP
 38     AUX3(I,J) = R1(I,J)
 37   CONTINUE

      CALL CHO(IP,AUX3,G1)
      CALL TRIANG(IP,G1,AUX3)
      CALL TRANSP(IP,IP,AUX3,G1)
      CALL MULT(IP,IP,IP,G1,R2,AUX4)
      CALL MULT(IP,IP,IP,AUX4,AUX3,G1)

      CALL DVS(IP,G1,VALOR)

      CALL MULT(IP,IP,IP,AUX3,G1,VETOR)
C
C   CALCULANDO  Y

      CALL YCALC(N1,N2,IP,Z1,Z2,VETOR,Y1,Y2)

C
C  CALCULANDO MEDIA E VARIANCIA DE Y
C

C
C  CALCULANDO AS MEDIAS DE Y
C

      CALL MEDIA(N1,2,Y1,YMED1)
      CALL MEDIA(N2,2,Y2,YMED2)
C
C  CALCULANDO AS COVARIANCIAS DE Y
C

      CALL COV(N1,2,Y1,YMED1,YS1)
      CALL COV(N2,2,Y2,YMED2,YS2)

C
C   CALCULANDO A FUNCAO DISCRIMINANTE
C
C
C   CALCULANDO W1 = D2(-1) * X1          W2 = D1(-1) * X2
C

      CALL MULT(N2,IP,IP,X1,DP2,Z1)
      CALL MULT(N1,IP,IP,X2,DP1,Z2)
C
C  CALCULANDO Y1EXT = BETA´ * W1         Y2EXT = BETA´ * W2
C
```

```
      CALL YCALC(N1,N2,IP,Z1,Z2,VETOR,Y1EXT,Y2EXT)
C
C   APLICANDO A FUNCAO DISCRIMINANTE AOS DADOS DA POP 1
C

C
C        CALCULANDO    LN( DET(SIGMAY2) )
C
      DD = 1.0D0
      DO 135 I=1,2
 135  DD = DD * YS2(I,I)
      DET = DLOG(DD)
C
C          CALCULANDO   INV( SIGMAY2 )
C
      DO 140 I=1,2
        DO 145 J=1,2
 145    D1(I,J) = 0.0D0
 140  CONTINUE
      DO 150 I=1,2
 150  D1(I,I) = 1.0D0 / YS2(I,I)

C
C   APLICANDO A FUNCAO DISCRIMINANTE AOS DADOS DA POP 1
C
      DO 160 I=1,N1

        DO 165 J=1,2
          YAUX1(J) = Y1(I,J)
          YAUX2(J) = Y1EXT(I,J)
 165    CONTINUE


      CALL DISCOR(YAUX1,YAUX2,YMED1,YMED2,D1,DET,FUNCAO)

C
C   FUNCAO
C
      IF ( FUNCAO .GE. 0.0D0 ) THEN
        POPFIM(I,1) = 1
      ELSE
        POPFIM(I,1) = 2
      ENDIF

 160  CONTINUE
C
C   APLICANDO A FUNCAO DISCRIMINANTE AOS DADOS DA POP 2
C
      DO 200 I=1,N2

        DO 205 J=1,2
```

```
          YAUX1(J) = Y2EXT(I,J)
          YAUX2(J) = Y2(I,J)
 205     CONTINUE

      CALL DISCOR(YAUX1,YAUX2,YMED1,YMED2,D1,DET,FUNCAO)

C
C   FUNCAO
C
      IF ( FUNCAO .GE. 0.0D0 ) THEN
        POPFIM(I,2) = 1
      ELSE
        POPFIM(I,2) = 2
      ENDIF

 200  CONTINUE
C
C  CALCULANDO AS PERCENTAGENS DE ACERTO NAS CLASSIFICACOES
C
C                POPULACAO 1
      XI1 = 0.0
      DO 170 I=1,N1
        IF(POPFIM(I,1) .EQ. 1) XI1 = XI1 + 1.0

 170  CONTINUE

      PERC1 = XI1 / FLOAT(N1) * 100.0

C                POPULACAO 2
      XI1 = 0.0
      DO 175 I=1,N2
        IF(POPFIM(I,2) .EQ. 2) THEN
        XI1 = XI1 + 1.0
        ENDIF
 175  CONTINUE

      PERC2 = XI1 / FLOAT(N2) * 100.0

C
C   IMPRESSAO DOS RESULTADOS
C
      CALL IMPCOR(N1,N2,IP,NPOP,FMT,X1,X2,XMED1,XMED2,S1,S2,R1,R2,
     *VALOR,VETOR,Y1,Y2,YMED1,YMED2,YS1,YS2,POP1,POP2,POPFIM,PERC1,
     *PERC2,TIT,t3,X01,X02,TQUAD,FTAB,IGL1,IGL2,F,IO,IGLQUI,VC,QUI)
C
C   GRAFICANDO OS DADOS TRANSFORMADOS
C
      NN(1,1) = N1
      NN(1,2) = N2
      DO 210 I=1,N1
       YGR(I) = Y1(I,1)
```

```
        XGR(I) = Y1(I,2)
  210   CONTINUE
        I1 = N1
        DO 215 I=1,N2
          I1 = I1 + 1
          YGR(I1) = Y2(I,1)
          XGR(I1) = Y2(I,2)
  215   CONTINUE

        CALL GRAF(YGR,XGR,NN,YG,INDORD,TIT)

C     METODO DA CORRELACAO: FIM



        RETURN
        END
```

SUBROTINA: COVAR

```
        SUBROUTINE COVAR(N1,N2,IP,N12,NPOP,X1,X2,Y1,Y2,S1,S2,AUX3,AUX4,
     *      VALOR,VETOR,G1,YG,XGR,YGR, POP1,POP2,POPFIM,INDORD,NN,D1,
     *      YMED1,YMED2,YS1,YS2,YAUX1,YAUX2,TIT,IO)

C
C     ROTINA:    METODO DAS  COVARIANCIAS
C

        DIMENSION X1(N1,IP), X2(n2,IP), Y1(N1,2), Y2(N2,2)
        DIMENSION S1(IP,IP), S2(IP,IP)
        DIMENSION VALOR(IP), VETOR(IP,IP), G1(IP,IP), NN(2,2),D1(2,2)
        DIMENSION YG(N12), INDORD(N12), XGR(N12), YGR(N12)
        DIMENSION YMED1(2), YMED2(2),YS1(2,2),YS2(2,2),YAUX1(2),YAUX2(2)
        DIMENSION AUX3(IP,IP), AUX4(IP,IP)
        INTEGER TIT(10), POP1(N1), POP2(N2), POPFIM(N1,2)

        DOUBLE PRECISION X1,X2,S1,S2,VALOR,VETOR,X01,X02,P0,Y1,Y2,YMED1,
     *        YMED2,YS1,YS2,DD,DET,D1,F,FUNCAO,CONST1,CONST2,CONST,
     *        YAUX1,YAUX2,AUX3,AUX4,XGR,YGR,YG,G1,T3.ALFA,XPROB,TQUAD,
     *        FTAB,VC,QUI

        NN(2,1) = '+'
        NN(2,2) = 'O'
```

```
C     INICIO


C    CALCULANDO OS AUTO-VALORES E AUTO-VETORES DE S1(-1) * S2
C
      DO 78 I=1,IP
        DO 79 J=1,IP
 79       AUX3(I,J) = S1(I,J)
 78   CONTINUE

      CALL CHO(IP,AUX3,G1)
      CALL TRIANG(IP,G1,AUX3)
      CALL TRANSP(IP,IP,AUX3,G1)
      CALL MULT(IP,IP,IP,G1,S2,AUX4)
      CALL MULT(IP,IP,IP,AUX4,AUX3,G1)
      CALL DVS(IP,G1,VALOR)
      CALL MULT(IP,IP,IP,AUX3,G1,VETOR)
C
C    CALCULANDO Y = BETA' * X
C

      CALL YCALC(N1,N2,IP,X1,X2,VETOR,Y1,Y2)

C
C   CALCULANDO MEDIA E VARIANCIA DE Y
C

C
C   CALCULANDO AS MEDIAS DE Y
C

      CALL MEDIA(N1,2,Y1,YMED1)
      CALL MEDIA(N2,2,Y2,YMED2)

C
C   CALCULANDO AS COVARIANCIAS DE Y
C

      CALL COV(N1,2,Y1,YMED1,YS1)
      CALL COV(N2,2,Y2,YMED2,YS2)
C
C    CALCULANDO A FUNCAO DISCRIMINANTE
C
C    INVERSA DE SY2

      DO 1135 I=1,NPOP
        DO 1140 J=1,NPOP
 1140   D1(I,J) = 0.0D0
 1135  CONTINUE

      DO 1145 I=1,NPOP
```

```
 1145 D1(I,I) = 1.0D0 / YS2(I,I)
C
C     CALCULANDO A PARTE CONSTANTE
C
C     DETERMINANTE DE COV(Y2)
C
       DD = 1.0D0
       DO 1150 I=1,NPOP
         DD = DD * YS2(I,I)
 1150  CONTINUE

       DET = DLOG(DD)

        CONST1 = 0.0D0
       DO 1131 I=1,2
 1131        CONST1 = CONST1 + YMED1(I) * YMED1(I)

       CONST2 = 0.0D0
       DO 1132 I=1,2
 1132  CONST2 = CONST2 + YMED2(I) ** 2 * D1(I,I)

       CONST = CONST1 - CONST2 - DET

C
C    APLICANDO A FUNCAO DISCRIMINANTE AOS DADOS DA POP 1
C
       DO 1160 I=1,N1

         DO 1165 J=1,NPOP
          YAUX1(J) = Y1(I,J)
 1165    CONTINUE

       CALL DISCOV(YAUX1,D1,YMED1,YMED2,CONST,FUNCAO)

C
C    FUNCAO
C
        IF (FUNCAO .GE. 0.0D0) GOTO 231
        POPFIM(I,1) = 2
        GO TO 1160
 231    POPFIM(I,1) = 1
 1160   CONTINUE

C
C    APLICANDO A FUNCAO DISCRIMINANTE AOS DADOS DA POP 2
C
       DO 1200 I=1,N2

         DO 1205 J=1,2
          YAUX2(J) = Y2(I,J)
 1205    CONTINUE
```

```
      CALL DISCOV(YAUX2,D1,YMED1,YMED2,CONST,FUNCAO)
C
C   FUNCAO
C

       IF (FUNCAO .GE. 0.0D0) GOTO 261
       POPFIM(I,2) = 2
       GO TO 1200
 261   POPFIM(I,2) = 1
 1200 CONTINUE
C
C  CALCULANDO AS PERCENTAGENS DE ACERTO NAS CLASSIFICACOES
C
C                POPULACAO 1
      XI1 = 0.0
      DO 1170 I=1,N1
        IF(POPFIM(I,1) .NE. 1) GOTO 1170
        XI1 = XI1 + 1.0
 1170  CONTINUE

      PERC1 = XI1 / FLOAT(N1) * 100.0

C                POPULACAO 2
      XI1 = 0.0
      DO 1175 I=1,N2
        IF(POPFIM(I,2) .NE. 2) GOTO 1175
        XI1 = XI1 + 1.0
 1175  CONTINUE

      PERC2 = XI1 / FLOAT(N2) * 100.0

C
C   IMPRESSAO DOS RESULTADOS
C
      CALL IMPCOV(N1,N2,IP,NPOP,VALOR,VETOR,Y1,Y2,YMED1,YMED2,
     *YS1,YS2,POP1,POP2,POPFIM,PERC1,PERC2,TIT,IO)
C
C  GRAFICANDO OS DADOS TRANSFORMADOS
C
      NN(1,1) = N1
      NN(1,2) = N2
      DO 1210 I=1,N1
       YGR(I) = Y1(I,1)
       XGR(I) = Y1(I,2)
 1210  CONTINUE
      I1 = N1
      DO 1215 I=1,N2
        I1 = I1 + 1
        YGR(I1) = Y2(I,1)
        XGR(I1) = Y2(I,2)
```

```
 1215  CONTINUE
       CALL GRAF(YGR,XGR,NN,YG,INDORD,TIT)

C    METODO DA COVARIANCIA: FIM

       RETURN
       END
```

SUBROTINA: MEDIAS

```
C
C    CALCULANDO VETOR DE MEDIAS
C
       SUBROUTINE MEDIA(NL,NC,XX,XMED)
       DIMENSION XX(NL,NC), XMED(NL)
       DOUBLE PRECISION XX, XMED, AUX
       DO 10 I=1,NC
         AUX = 0.0D0
         DO 20 J=1,NL
          AUX = AUX + XX(J,I)
 20      CONTINUE
         XMED(I) = AUX / FLOAT(NL)
 10    CONTINUE
       RETURN
       END
```

SUBROTINA: COV

```
       SUBROUTINE COV(NL,NC,XX,XMED,SS)
       DIMENSION XX(NL,NC), XMED(NC), SS(NC,NC)
       DOUBLE PRECISION XX, XMED, SS, SOMA
C
C   CALCULA A MATRIZ DE COVARIANCIA DE UMA POPULACAO
C
       DO 10 I=1,NC
         DO 20 J=1,NC
           SOMA = 0.0D0
           DO 30 K=1,NL
```

```
30        SOMA = SOMA + XX(K,I) * XX(K,J)
          SS(I,J) = SOMA - XMED(I) * XMED(J) * DFLOAT(NL)
          SS(I,J) = SS(I,J) / DFLOAT(NL-1)
20      CONTINUE
10    CONTINUE

      DO 40 I=2,NC
        I1 = I - 1
        DO 50 J=1,I1
50      SS(I,J) = SS(J,I)
40    CONTINUE

      RETURN
      END
```

Subrotina: TESTMED

```
      subroutine testmed(n1,n2,ip,xmed1,xmed2,s,tquad,ftab,aux2,
     *aux3,aux4,g1,igl1,igl2,f)
      dimension xmed1(ip), xmed2(ip), s(ip,ip),aux3(ip,ip),aux4(ip,ip)
      dimension aux2(ip),g1(ip,ip)
      double precision xmed1,xmed2,s,tquad,f,sum,ftab,aux2,aux3,aux4,g1,
     *                   efe1
C
C  TESTA A IGUALDADE DE DOIS VETORES DE MEDIAS
C
c
c entrada: n1, n2, xmed1, xmed2
c          s = matriz pooled = (s1 + s2) / (n1 + n2 - 2)
c SAIDA:   TQUAD = ESTATISTICA DO TESTE
C          FTAB  = VALOR DA DISTRIBUICAO F

c
c calculando inv(s)
c

      do 5 i=1,ip
        do 7 j=1,ip
 7        aux4(i,j) = s(i,j)
 5      continue

      call cho(ip,aux4,g1)

      call triang(ip,g1,aux3)
      call transp(ip,ip,aux3,aux4)
      call mult(ip,ip,ip,aux3,aux4,g1)
```

```
      do 40 i=1,ip
         sum = 0.0d0
         do 50 j=1,ip
50       sum = sum + (xmed1(j) - xmed2(j)) * g1(j,i)
          aux2(i) = sum
40         continue

      sum = 0.0d0
       do 60 i=1,ip
60        sum = sum + (xmed1(i) - xmed2(i)) * aux2(i)

       tquad  = dfloat(n1 * n2) * sum / dfloat(n1 + n2)

      f = dfloat(n1 + n2 - ip - 1) * tquad / dfloat( (n1 + n2 - 2) * ip)

      igl1 = ip
      igl2 = n1 + n2 - ip - 1
      ftab = efe1(igl1,igl2,f)


      return
      end
```

SUBROTINA: EFE1

```
            double precision function  efe1(gn,gd,xis)
      integer gn,gd
      double precision xis,agn,agd,pe,que,ene,esse,te,de,aux1,aux2,
     *    aux3, aux4, ze, apn4, ge, st2

c  calcula a distribuicao F cumulativa
c   adaptada de Estatistica Computacional    p.152
c                 N. Dachs


      agn = float(gn)
      agd = float(gd)

      if ( (gn .eq. 1) .or. (gd .eq.1) ) then
         if ( gn .eq. 1) then
           efe1 = 2.d0 * st2(gd,dsqrt(xis)) - 1.d0
         else
           efe1 = 2.d0 - 2.d0 * st2(gn,dsqrt(xis))
         endif

      else
```

```
          pe = agd / (agn * xis + agd)
          que = 1.d0 - pe
          ene =  (agn + agd - 2.d0) / 2.d0
          esse =  ( agd - 1.d0 ) / 2.d0
          te =  ( agn - 1.d0 ) / 2.d0
          de = esse + 0.16666666667d0 - (ene + 0.33333333333d0) * pe
          de = de + 0.04d0 * (que/agd - pe/agn + (que + 0.5d0)/(agn+agd))
          aux1 = ge(esse/ene/pe)
          aux2 = ge(te/ene/que)
          aux3 = 1.d0 + que * aux1 + pe * aux2
          aux4 = (ene + 0.16666666667d0) * pe * que
          ze = de * dsqrt(aux3/aux4)
          efe1 = apn4(ze)
        endif


        return
        end
```

SUBROTINA: APN4

```
              double precision function apn4(xis)
        double precision xis, aux1, aux2, aux3, aux4, soma

c aproximacao para a distribuicao cumulativa normal
c  Obs: adaptada de Estatistica Computacional   p.125
c        Norberto Dachs

           if ( dabs(xis) .lt. 10.d0 ) then
          if (xis .ge. 0.d0) then
            aux1 = xis
          else
            aux1 = -xis
          endif

          aux2 = 1.4142135624d0 * aux1 / 3.0d0

          soma = 0.d0
          do 10 i=1,13
            i1 = i - 1
            aux3 = i1 + 0.5d0
            soma = soma + dsin(aux2*aux3) * dexp(-aux3*aux3/9.d0) / aux3
 10          continue

          aux4 = 0.5d0 + soma / 3.1415926536d0
```

```
         if (xis .ge. 0.d0) then
            apn4 = aux4
         else
            apn4 = 1.d0 - aux4
         endif

      else

         if (xis .gt. 0.d0) then
            apn4 = 1.d0
         else
            apn4 = 0.d0
         endif

      endif

      return
      end
```

SUBROTINA: GE

```
              double precision function ge(xis)
      double precision xis

c  calculo da funcao g de Peizer e Pratt
c  Obs: adaptada de Estatistica Computacional   p.145
c                       Norberto Dachs

      if (xis .lt. 1.0e-9) then
        ge = 1.
      else
        if ( abs(1.-xis) .lt. 1.0e-3 ) then
           ge = 0.
        else
           ge = (1. - xis*xis + 2.*xis*alog(xis)) / ((1.-xis)*(1.-xis))
        endif
      endif

      return
      end
```

SUBROTINA: ST2

```
           double precision function st2(gl,xis)
      integer gl
      double precision xis, agl, agn, aux, c1,c2,te,v,w,y,z, apn4

c  aproximacao para a distribuicao t de student
c  obs: adaptada de Estatistica Computacional  p.136
c                   Norberto Dachs


      agl = dfloat(gl)

      if (gl .eq. 1) then
         aux = datan(xis) / 3.1415926536d0 + 0.5d0
      else
        if (gl .eq. 2) then
           aux = 0.5d0 * xis / dsqrt(xis*xis+2.d0) + 0.5d0
        else
          if (gl .eq. 3) then
             aux = ( datan(xis/1.7320508076d0) + xis * 1.7320508076d0 /
     *                  (xis*xis+3.d0)) / 3.1415926536d0 + 0.5d0
          else
            if (gl .eq. 4) then
              aux = 0.5d0 * xis * (1.d0 + 2.d0 / (xis*xis+4.d0) ) /
     *                dsqrt( xis * xis + 4.d0 ) + 0.5d0
            else
              c1 = agl - 0.5d0
              c2 = 48.d0 * c1 * c1
              te = xis * xis / agl
              if (te .ge. 0.04d0) then
                v = c1 * dlog(1.d0 + te)
              else
                v = c1 * (( (-te * 0.75d0 + 1.d0) * te/3.d0 - 0.5d0 ) *
     *                te + 1.d0 ) * te
              endif
              w = ((0.4d0 * v + 3.3d0) * v + 24.d0) * v + 85.5d0
              y = 0.8d0 * v * v + 100.d0 + c2
              z = ((-w / y + v + 3.d0) / c2 + 1.d0) * sqrt(v)
              aux = apn4(z)
            endif
          endif
        endif
      endif

      st2 = aux

      return
      end
```

SUBROTINA: TCOV

```
      subroutine tcov(n1,n2,ip,s1,s2,s,vc,iglqui,qui,aux3,g)
      dimension s1(ip,ip), s2(ip,ip), s(ip,ip), aux3(ip,ip),g(ip,ip)
      double precision s1,s2,s,vc,qui,aux3,g,an1,an2,aip,v,c, dets1,
     *                 dets2, dets,qui2

c faz o teste de igualdade das matrizes de covariancias
c   baseado na estatistica vc
c   utiliza a decomposicao de cholesky

      do 10 i=1,ip
        do 20 j=1,ip
          aux3(i,j) = s1(i,j)
20         continue
10       continue

      call cho(ip,aux3,g)

      dets1 = 1.0d0
      do 30 i=1,ip
30       dets1 = dets1 * g(i,i)
      dets1 = dlog( dets1 * dets1 )


        do 40 i=1,ip
        do 50 j=1,ip
          aux3(i,j) = s2(i,j)
50         continue
40       continue

       call cho(ip,aux3,g)

      dets2 = 1.0d0
      do 60 i=1,ip
60       dets2 = dets2 * g(i,i)
      dets2 = dlog( dets2 * dets2 )

       do 70 i=1,ip
        do 80 j=1,ip
          aux3(i,j) = s(i,j)
80         continue
70       continue

       call cho(ip,aux3,g)

      dets = 1.0d0
      do 90 i=1,ip
```

```
90      dets = dets * g(i,i)
       dets = dlog( dets * dets )

      an1 = dfloat(n1 - 1)
      an2 = dfloat(n2 - 1)
      v = (an1 + an2) * dets - ( an1 * dets1 + an2 * dets2 )


      aip = dfloat(ip)

      c = (1.d0 / an1) + (1.d0 / an2) - ( 1.d0 / (an1 + an2) )
      c = 1.0d0 - (2.d0*aip*aip + 3.d0*aip - 1.d0) / (6.d0*(aip + 1.d0))
     *          * c

      vc = v * c

      iglqui = ip * (ip + 1) / 2


      qui = qui2(iglqui,vc)


      return
      end
```

SUBROTINA: QUI2

```
        double precision function qui2(d2,x2)
      integer d2
      double precision x2,z,z2,c,g,d,a,d3,fn1,d1,t2,z1,s9,c2,ai5

c  calcula a distribuicao qui-quadrado num ponto dado
c   adaptada de  Maindonald   p.294


      if (d2 .lt. 11) then
        z = x2 / 2.d0
        z2 = z * z
        c = 1.d0
        g = 1.d0
        d = dfloat(d2) / 2.d0
        a = d
        d3 = d + 2.d0

100        a = a + 1.d0
         c = c * z / a
```

```
      g = g + c

      if ( c/g .gt. 0.5e-6 ) goto 100

      g = g * dexp(d * dlog(z) - d3 * fn1(d3*d3) - (d3 - 0.5d0) *
     *       dlog(d3) + d3 - z ) * (d + 1)

      qui2 = 1.d0 - g / dsqrt(2.d0 * 3.14159d0)


     else

      d1 = dfloat(d2 - 1)
      t2 = d1 / x2
      d3 = x2 - d2 + 2.d0/3.d0 - 0.08d0/d2

      g = 1.d0
      if (t2 .eq. 0.d0) return
      if ( dabs(1.d0 - t2) .gt. 0.1d0 ) then

         g = (1.d0 - t2*t2 + 2.d0*t2*dlog(t2)) / ((1.d0-t2)*(1.d0-t2))

      else
         g = 0.d0
         do 10 j=1,5
10          g = g + 2.d0 * (1.d0 - t2) ** j / dfloat( (j+1) * (j+2) )

       endif

      z1 = d3 * dsqrt( (1.d0 + g) / (2.d0 * x2) )

      s9 = 0.d0
      c2 = dsqrt(2.d0) / 3.d0 * z1

      do 20 i=1,13
        i1 = i - 1
        ai5 = dfloat(i1) + 0.5d0
20       s9 = s9 + dsin(ai5*c2) * dexp(-ai5*ai5/9.d0)/ai5

       qui2 = 0.5d0 - s9 / 3.141593d0

     endif

     qui2 = 1.d0 - qui2

     return
     end
```

SUBROTINA: FN1

```
            double precision function fn1(xis)
      double precision xis

c funcao fn1 usada na aproximacao do log da funcao gama
c  adaptada de Estatistica Computacional    p.144
c               N. Dachs

      fn1 = 1.d0 / (12.d0*xis) * ( 1.d0 - 1.d0/xis *
     *    (1.d0/30.d0 - 1.d0/xis * (1.d0/105.d0 - 1.d0/(140.d0*xis))) )


      return
      end
```

SUBROTINA: TESTCOR

```
            SUBROUTINE TESTCOR(N1,N2,IP,R1,R2,T3)
      DIMENSION R1(IP,IP), R2(IP,IP)
      DOUBLE PRECISION R1,R2,D,T3, TAUX
C
C         CALCULA O TESTE DE IGUALDADE DE DUAS MATRIZES DE CORRELACOES
C
C
C         ENTRADA: N1 E N2 = NUMERO DE OBSERVACOES
C                  R1 E R2 = MATRIZES DE CORRELACOES
C                  IP = NUMERO DE VARIAVEIS
C
C         SAIDA:   T3 = ESTATISTICA DO TESTE
C

      D = DFLOAT( (N1 - 3) * (N2 - 3) ) / DFLOAT( N1 + N2 - 6 )
      D = DSQRT(D)


      A1 = DLOG ( (1.0D0 + R1(1,2)) / (1.0D0 - R1(1,2)) )
      A2 = DLOG ( (1.0D0 + R2(1,2)) / (1.0D0 - R2(1,2)) )
      T3 = DABS ( (A1 - A2) / 2.0D0 )

      IP1 = IP - 1

      DO 10 I=1,IP1
        I1 = I + 1
```

```
      DO 20 J=I1,IP
        A1 = DLOG( (1.0D0 + R1(I,J)) / (1.0D0 - R1(I,J)) )
        A2 = DLOG( (1.0D0 + R2(I,J)) / (1.0D0 - R2(I,J)) )

        TAUX = DABS( (A1 - A2) / 2.0D0 )

        IF (T3 .LT. TAUX) THEN
          T3 = TAUX
        ENDIF
20      CONTINUE
10    CONTINUE

      T3 = T3 * D

      RETURN
      END
```

SUBROTINA: INVNOR

```
      subroutine invnor(p0,x0)
      double precision p0,x0,a0,a1,a2,a3,b1,b2,b3,b4,c0,c1,c2,c3,d1,d2,
     *q,r,x
c
c   calcula a distribuicao normal inversa, isto e,
c   encontra x0 tal que prob(z <= x0) = alfa
c   z tem dist normal padrao
c   alfa valor dado(p0)
c
c      adaptada de Maindonald,  p.293

      a0 = 2.5066282d0
      a1 = -18.6150006d0
      a2 = 41.3911977d0
      a3 = -25.4410605d0
      b1 = -8.4735109d0
      b2 = 23.0833674d0
      b3 = -21.0622410d0
      b4 = 3.1308291d0
      c0 = -2.7871893d0
      c1 = -2.2979648d0
      c2 = 4.8501413d0
      c3 = 2.3212128d0
      d1 = 3.5438892d0
```

```
      d2 = 1.6730678d0

      q = p0 - 0.5d0
      if ( dabs(q) .gt. 0.42d0 ) then

        r = p0
          if ( q .gt. 0.0d0 ) r = 1.0d0 - p0
        if ( r .le. 0.0d0 ) then
          print 10, p0
 10          format(5x,f10.4,'nao esta entre 0 e 1')
           goto 100
        endif

        r = dsqrt( - dlog(r) )
        x = (((c3 * r + c2) * r + c1) * r + c0)
        x0 = x / ((d2 * r + d1) * r + 1.0d0)
        if ( q .lt. 0.0d0 )  x0 = -x0

      else

        r = q * q
        x = q * (((a3 * r + a2) * r + a1) * r + a0)
        x0 = x / ((((b4 * r + b3) * r + b2) * r + b1) * r + 1)

      endif

 100      return
      end
```

SUBROTINA: MULT

```
       SUBROUTINE MULT(NLX,NCX,NCY,XX,YY,ZZ)
      DIMENSION XX(NLX,NCX), YY(NCX,NCY) , ZZ(NLX,NCY)
      DOUBLE PRECISION XX, YY, ZZ, SOMA
C
C   CALCULA PRODUTO DE DUAS MATRIZES
C      ZZ(NLX,NCY) = XX(NLX,NCX) * YY(NCX,NCY)
C

      DO 10 I = 1,NLX
        DO 20 J = 1,NCY
           SOMA = 0.0D0
           DO 30 K = 1,NCX
              SOMA = SOMA + XX(I,K) * YY(K,J)
 30        CONTINUE
           ZZ(I,J) = SOMA
```

```
 20       CONTINUE
 10     CONTINUE
        RETURN
        END
```

SUBROTINA: CHO

```
      SUBROUTINE CHO(NLC,A,T)
      DIMENSION A(NLC,NLC), T(NLC,NLC)
      DOUBLE PRECISION A,T,SOMA,DELTA,EPS

C     DECOMPOSICAO DE CHOLESKY DA MATRIZ A
C     GUARDA O RESULTADO NA MATRIZ TRIANGULAR SUPERIOR T

      EPS = 0.00001D0

      DO 10 I=1,NLC
        SOMA = 0.0D0
        I1 = I - 1
        DO 20 K=1,I1
 20     SOMA = SOMA + A(K,I) * A(K,I)
        DELTA = A(I,I) - SOMA
        IF (DELTA .LT. 0.D0) THEN
          PRINT 11
 11       FORMAT(10X,'MATRIZ A NAO E POSITIVA DEFINIDA')
          GOTO 100
        ENDIF
        A(I,I) = DSQRT(DELTA)
        I2 = I + 1
        DO 30 J=I2,NLC
          SOMA = 0.D0
          DO 40 K=1,I1
 40       SOMA = SOMA + A(K,I) * A(K,J)
          IF ( DABS(A(I,I)) .LE. EPS ) THEN
            PRINT 11
            GOTO 100
          ENDIF
          A(I,J) = ( A(I,J) - SOMA ) / A(I,I)
 30     CONTINUE
 10   CONTINUE

      DO 50 I=1,NLC
        DO 60 J=1,NLC
          IF ( I .GT. J) THEN
             T(I,J) = 0.0D0
          ELSE
```

```
            T(I,J) = A(I,J)
          ENDIF
 60     CONTINUE
 50   CONTINUE

 100  RETURN
      END
```

Subrotina: TRIANG

```
      SUBROUTINE TRIANG(NLC,XX,XI)
      DIMENSION XX(NLC,NLC), XI(NLC,NLC)
      DOUBLE PRECISION XX, XI, SOMA

C     CALCULO DA INVERSA DE UMA MATRIZ TRIANGULAR SUPERIOR

C       XX = MATRIZ A SER INVERTIDA (DE ENTRADA)
C       XI = MATRIZ INVERSA DE XX (NA SAIDA)

      DO 10 I=1,NLC
 10   XI(I,I) = 1.0D0 / XX(I,I)

      DO 20 I=2,NLC
        I1 = I - 1
        DO 30 J=1,I1
 30     XI(I,J) = 0.0D0
 20   CONTINUE

      NLC1 = NLC - 1

      DO 40 I=NLC1,1,-1

        I1 = I + 1
        DO 50 J=I1,NLC
          SOMA = 0.0D0
          DO 60 K=I1,NLC
 60       SOMA = SOMA + XX(I,K) * XI(K,J)
          XI(I,J) = - SOMA / XX(I,I)
 50     CONTINUE

 40   CONTINUE

      RETURN
      END
```

Subrotina: TRANSP

```
      SUBROUTINE TRANSP(NL,NC,XX,XXL)
C
C     CALCULA MATRIZ TRANSPOSTA
C
C     NL = NUMERO DE LINHAS - MATRIZ ORIGINAL
C     NC = NUMERO DE COLUNAS - MATRIZ ORIGINAL
C     XX = MATRIZ ORIGINAL
C     XXL = MATRIZ TRANSPOSTA
C
      DIMENSION XX(NL,NC), XXL(NC,NL)
      DOUBLE PRECISION XX,XXL
      DO 10 I = 1,NL
         DO 20 J = 1,NC
 20      XXL(J,I) = XX(I,J)
 10   CONTINUE

      RETURN
      END
```

Subrotina: DVS

```
      SUBROUTINE DVS(IP,U,D)
      DIMENSION U(IP,IP), D(IP)
      DOUBLE PRECISION U,D, E0,E2,X2,H,Y2,X0,Y0,C0,S0,C2,Q,H2,R,D1,D2

C   CALCULO DA DECOMPOSICAO DE VALORES SINGULARES DE UMA MATRIZ REAL


C   U - MATRIZ DE ENTRADA. L X IP

C   U, V MATRIZES ORTOGONAIS, D MATRIZ DIAGONAL
C   U' Z V = D
C   NESTE PROGRAMA - CALCULA: V' Z' U = D

C   OBS:
C   D1 ** 2 = AUTO-VALOR DE Z'Z
C   COLUNAS DE V AUTO-VETORES DE Z'Z
C   A FORMA EXPLICITA DE U NAO E FORNECIDA

C       ADAPTADA DE MAINDONALD   P. 339
```

```
      L = IP
      M = IP

C   CALCULANDO E0 = EPSILON

      E0 = 1.0D0
      DO 10 I=1,50
        E0 = E0 / 2.0D0
        E = 1.0D0 + E0
        IF ( E .EQ. 1.0D0 ) GOTO 110
 10   CONTINUE

 110  E0 = E0 * 2.0D0

      L0 = 1
      M0 = 1
      L2 = L0 + L - 1
      M2 = M0 + M - 1
      E2 = DFLOAT(L) * E0 ** 2

C  CALCULANDO A DVS

 7010 M3 = M * (IP-1) / 2

      DO 20 J=M0, M2-1
        DO 30 K=J+1,M2
          X2 = 0.0D0
          H = 0.0D0
          Y2 = 0.0D0
          DO 40 I=L0,L2
            X0 = U(I,J)
            Y0 = U(I,K)
            H = H + X0 * Y0
            X2 = X2 + X0 ** 2
            Y2 = Y2 + Y0 ** 2
 40       CONTINUE

          C0 = 0.0D0
          S0 = 1.0D0
          IF (X2 .LT. Y2) GOTO 7200
          C2 = X2 * Y2
          IF (C2 .LE. E2) GOTO 7220
          IF ( H**2 .LT. (E2*C2) ) GOTO 7220
          Q = X2 - Y2
          H2 = 2.0D0 * H
          R = DSQRT(Q**2 + H2**2)
          C0 = DSQRT( (R+Q) / (2.0D0*R) )
          S0 = H2 / ( 2.0D0*R*C0 )
 7200     DO 50 I=L0,L2
          D1 = U(I,J)
          D2 = U(I,K)
```

```
         U(I,J) = D1 * C0 + D2 * S0
         U(I,K) = -D1 * S0 + D2 * C0
50       CONTINUE
         GOTO 30
7220     M3 = M3 - 1
30     CONTINUE
20    CONTINUE

      IF (M3 .GT. 0) GOTO 7010

      DO 60 J=M0,M2
        X2 = 0.0D0
        DO 70 I=L0,L2
          X2 = X2 + U(I,J) ** 2
70      CONTINUE
        D(J) = DSQRT(X2)
60    CONTINUE

      DO 80 J=M0,M2
        IF ( D(J) .EQ. 0.0D0 ) GOTO 80
        DO 90 I=L0,L2
          U(I,J) = U(I,J) / D(J)
90      CONTINUE
80    CONTINUE

      RETURN
      END
```

SUBROTINA: YCALC

```
      SUBROUTINE YCALC(N1,N2,IP,ZZ1,ZZ2,VETOR,YY1,YY2)
      DIMENSION YY1(N1,2),YY2(N2,2),ZZ1(N1,IP),ZZ2(N2,IP),VETOR(IP,IP)
      DOUBLE PRECISION YY1,YY2,AA1,AA2,ZZ1,ZZ2,VETOR

C   CALCULANDO Y = BETA' * Z

C          Y = BETA(MAX) * Z
C              BETA(MIN) * Z

C         OS AUTO-VALORES E VETORES VEM EM ORD DECRESCENTE



      DO 80 J=1,2
       DO 85 I=1,N1
```

```
85      YY1(I,J) = 0.0D0
        DO 90 I=1,N2
90      YY2(I,J) = 0.0D0
80     CONTINUE

       DO 95 I=1,N1
         AA1 = 0.0D0
         AA2 = 0.0D0
         DO 100 J=1,IP
           AA1 = AA1 + ZZ1(I,J) * VETOR(J,1)
           AA2 = AA2 + ZZ1(I,J) * VETOR(J,IP)
100      CONTINUE
         YY1(I,1) = AA1
         YY1(I,2) = AA2
95     CONTINUE

       DO 110 I=1,N2
         AA1 = 0.0D0
         AA2 = 0.0D0
         DO 115 J=1,IP
           AA1 = AA1 + ZZ2(I,J) * VETOR(J,1)
           AA2 = AA2 + ZZ2(I,J) * VETOR(J,IP)
115      CONTINUE
         YY2(I,1) = AA1
         YY2(I,2) = AA2
110    CONTINUE

       RETURN
       END
```

SUBROTINA: DISCOR

```
       SUBROUTINE DISCOR(YAUX1,YAUX2,YMED1,YMED2,D1,DET,FUNCAO)
       DIMENSION YAUX1(2), YAUX2(2), YMED1(2),YMED2(2),TAUX(2),D1(2,2)
       DOUBLE PRECISION D1,T1,T2,DET,FUNCAO,TAUX,YAUX1,YAUX2,YMED1,YMED2

C CALCULANDO (Y2 - YMED2)' * INV(YS2) * (Y2 - YMED2)

       DO 10 I=1,2
   YAUX1(I) = YAUX1(I) - YMED1(I)
   YAUX2(I) = YAUX2(I) - YMED2(I)
10     CONTINUE

       TAUX(1) = YAUX2(1) * D1(1,1)
```

```
      TAUX(2) = YAUX2(2) * D1(2,2)

 T1 = ( TAUX(1) * YAUX2(1) ) + ( TAUX(2) * YAUX2(2) )

C  CALCULANDO (Y1 - YMED1)' * (Y1 - YMED1)

      T2 = ( YAUX1(1) * YAUX1(1) ) + ( YAUX1(2) * YAUX1(2) )

C  CALCULANDO:
C  FUNCAO = LN ( DET(YS2) ) + (Y2 - YMED2)' * INV(YS2) * (Y2 - YMED2) +
C           + (Y1 - YMED1)' * (Y1 - YMED1)


      FUNCAO = T1 - T2 + DET


      RETURN
      END
```

SUBROTINA: IMPCOR

```
      SUBROUTINE IMPCOR(N1,N2,IP,NPOP,FMT,X1,X2,XMED1,XMED2,S1,S2,R1,
     *R2,VALOR,VETOR,Y1,Y2,YMED1,YMED2,YS1,YS2,POP1,POP2,POPFIM,PERC1,
     *PERC2,TIT,T3,X01,X02,TQUAD,FTAB,IGL1,IGL2,F,IO,IGLQUI,VC,QUI)
      DIMENSION X1(N1,IP),X2(N2,IP),XMED1(IP),XMED2(IP),S1(IP,IP)
      DIMENSION S2(IP,IP),R1(IP,IP), R2(IP,IP),VALOR(IP),VETOR(IP,IP)
      DIMENSION Y1(N1,2),Y2(N2,2),YMED1(NPOP),YMED2(NPOP)
      DIMENSION YS1(NPOP,NPOP), YS2(NPOP,NPOP)
      INTEGER POP1(N1), POP2(N2), POPFIM(N1,2), FMT(20), TIT(10)
      INTEGER P
      DOUBLE PRECISION X1, X2, XMED1, XMED2, S1, S2, R1, R2,VALOR,
     *VETOR, Y1, Y2, YMED1,YMED2,YS1,YS2,T3,TQUAD,FTAB,F,X01,X02,VC,QUI

      P = IP
C
C          IMPRESSAO DOS RESULTADOS: METODO DA CORRELACAO
C
C
C    IMPRIMINDO OS DADOS ORIGINAIS
C
      WRITE(IO,1) (TIT(I), I=1,10)
 1    FORMAT(5X,10A4,5X,'METODO DAS CORRELACOES',
     */,30X,'POPULACAO 1',/)
      I1  =  0
      DO 10 I=1,N1
        I1 = I1 + 1
```

```
        IF (I1 .LT. 51) GO TO 101
          WRITE(IO,31) (TIT(KK), KK=1,10)
 31     FORMAT('1',5X,10A4,5X,'METODO DAS CORRELACOES',/)
          WRITE(IO,200)
 200      FORMAT(15X,'POPULACAO 1 - CONTINUACAO',/)
          I1 = 0
 101    WRITE(IO,FMT) (X1(I,J), J=1,P) , POP1(I)
 10   CONTINUE

      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,2)
 2    FORMAT(30X,'POPULACAO 2',//)
      I1 = 0
      DO 20 I=1,N2
        I1 = I1 + 1
        IF (I1 .LT. 51) GOTO 102
          WRITE(IO,31) (TIT(KK), KK=1,10)
          WRITE(IO,201)
 201      FORMAT(15X,'POPULACAO 2 - CONTINUACAO',/)
          I1 = 0
 102    WRITE(IO,FMT) (X2(I,J), J=1,P) , POP2(I)
 20   CONTINUE
C
C  MEDIAS DE X
C
      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,3)
 3    FORMAT(10X,'VETORES DE MEDIAS DAS POPULACOES 1 E 2',//)
      DO 30 I=1,P
        WRITE(IO,4) XMED1(I),XMED2(I)
 4      FORMAT(3X,2(2X,F15.4))
 30   CONTINUE
C
C  COVARIANCIAS DE X
C
      WRITE(IO,5)
 5    FORMAT(////,10X,'MATRIZ DE COVARIANCIA DA POPULACAO 1',/)
      DO 40 I=1,P
       WRITE(IO,6) (S1(I,J), J=1,P)
 6     FORMAT(7(G10.4),/,2G10.4)
 40   CONTINUE

      WRITE(IO,36)
 36   FORMAT(///,10X,'MATRIZ DE COVARIANCIA DA POPULACAO 2',/)
      DO 50 I=1,P
       WRITE(IO,6) (S2(I,J), J=1,P)
 50   CONTINUE
C
C  CORRELACOES DE X
C
```

```
      WRITE(IO,7)
 7    FORMAT(////,10X,'MATRIZ DE CORRELACAO DA POPULACAO 1',/)
      DO 60 I=1,P
       WRITE(IO,6) (R1(I,J), J=1,P)
 60   CONTINUE
      WRITE(IO,8)
 8    FORMAT(///,10X,'MATRIZ DE CORRELACAO DA POPULACAO 2',/)
      DO 65 I=1,P
       WRITE(IO,6) (R2(I,J), J=1,P)
 65   CONTINUE
C
C  TESTE DE IGUALDADE DAS MATRIZES DE COVARIANCIAS
C

      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,54) VC,IGLQUI,QUI
 54   FORMAT(///,15X,'TESTE DE IGUALDADE DAS MATRIZES DE COVARIANCIAS',
     *//,10X,'ESTATISTICA DO TESTE (VC) = ',G10.4,//,
     *10X,'NUMERO DE GRAUS DE LIBERDADE = ',I4,//,
     *10X,'PROB ( QUI < QUI CALCULADO ) = ',F4.2,//)

      IF ( QUI .LT. 0.95 ) THEN
         WRITE(IO,72)
 72      FORMAT(10X,'CONCLUSAO:',/,12X,
     *'ACEITO A IGUALDADE DAS COVARIANCIAS AOS NIVEIS DE 5%  E 1%')
      ENDIF

      IF ( (QUI .GE. 0.95) .AND. (QUI .LT. 0.99) ) THEN
          WRITE(IO,73)
 73       FORMAT(10X,'CONCLUSAO:',/,12X,
     *'REJEITO A IGUALDADE DAS COVARIANCIAS AO NIVEL DE 5%  E',/,12X,
     *'ACEITO A IGUALDADE AO NIVEL DE 1%' )
      ENDIF

      IF ( QUI .GE. 0.99 ) THEN
         WRITE(IO,74)
 74      FORMAT(10X,'CONCLUSAO:',/,12X,
     *'REJEITO A IGUALDADE DAS COVARIANCIAS AOS NIVEIS DE 5% E 1%')
      ENDIF

C
C  TESTE DE IGUALDADE DE MEDIAS
C

      WRITE(IO,61) TQUAD,IGL1,IGL2,F,FTAB
 61      FORMAT(////,15X,'TESTE DE IGUALDADE DE MEDIAS',//,
     *10X,'ESTATISTICA DO TESTE(T QUADRADO) = ',G10.4,//,
     *10X,'NUMEROS DE GRAUS DE LIBERDADE: ',I4,' E ',I4,//,
     *10X,'ESTATISTICA F (CALCULADO) = ',G12.6,//,
     *10X,'PROB ( F < F CALCULADO ) = ',F4.2,//)
```

```
      IF ( FTAB .LT. 0.95 ) THEN
         WRITE(IO,62)
 62      FORMAT(10X,'CONCLUSAO:',/,12X,
     *'ACEITO A IGUALDADE DE MEDIAS AOS NIVEIS DE 5%  E 1%')
      ENDIF

      IF ( (FTAB .GE. 0.95) .AND. (FTAB .LT. 0.99) ) THEN
          WRITE(IO,63)
 63       FORMAT(10X,'CONCLUSAO:',/,12X,
     *'REJEITO A IGUALDADE DE MEDIAS AO NIVEL DE 5%  E',/,12X,
     *'ACEITO A IGUALDADE AO NIVEL DE 1%' )
      ENDIF

      IF ( FTAB .GE. 0.99 ) THEN
    WRITE(IO,64)
 64      FORMAT(10X,'CONCLUSAO:',/,12X,
     *'REJEITO A IGUALDADE DE MEDIAS AOS NIVEIS DE 5% E 1%')
      ENDIF

C
C  TESTE DE IGUALDADE DE CORRELACOES
C

      WRITE(IO,361) T3,X01,X02
 361  FORMAT(////,15X,'TESTE DE IGUALDADE DAS MATRIZES DE CORRELACOES',
     *//,10X,'ESTATISTICA DO TESTE = ',G10.4,//,
     *10X,'B(ALFA=0.05) = ',G10.4,//,
     *10X,'B(ALFA=0.01) = ',G10.4,//)

      IF ( (T3 .GT. X01) .AND. (T3 .GT. X02) ) THEN
        WRITE(IO,363)
 363        FORMAT(10X,'CONCLUSAO: ',/,12X,
     *'REJEITO A IGUALDADE DAS CORRELACOES AOS NIVEIS DE 5% E 1%')

      ELSE
        IF ( (T3 .LE. X01) .AND. (T3 .LE. X02) ) THEN
          WRITE(IO,364)
 364        FORMAT(10X,'CONCLUSAO: ',/,12X,
     *'ACEITO A IGUALDADE DAS CORRELACOES AOS NIVEIS DE 5% E 1%')

        ELSE
          IF ( (T3 .GT. X01) .AND. (T3 .LE. X02) ) THEN
            WRITE(IO,365)
 365            FORMAT(10X,'CONCLUSAO: ',/,12X,
     *'REJEITO A IGUALDADE DAS CORRELACOES AO NIVEL DE 5% E',/,12X,
     *'ACEITO A IGUALDADE AO NIVEL DE 1%')

          ELSE
            IF ( (T3 .LE. X01) .AND. (T3 .GT. X02) ) THEN
              WRITE(IO,366)
 366              FORMAT(10X,'CONCLUSAO: ',/,12X,
```

```
     *'ACEITO A IGUALDADE DAS CORRELACOES AO NIVEL DE 5% E',/,12X,
     *'REJEITO A IGUALDADE AO NIVEL DE 1%')
            ENDIF
          ENDIF
        ENDIF
      ENDIF

C
C   AUTO-VALORES E AUTO-VETORES
C
      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,9)
 9    FORMAT(10X,'AUTO-VALORES DE INV(R1) * R2',/)
      WRITE(IO,6) (VALOR(I), I=1,P)
      WRITE(IO,11)
 11   FORMAT(////,10X,'AUTO-VETORES DE INV(R1) * R2',/)
      DO 70 I=1,P
 70   WRITE(IO,6) (VETOR(I,J), J=1,P)
C
C   IMPRIMINDO Y
C

      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,12)
 12   FORMAT(25X,'POPULACAO 1',/,15X,'MATRIZ Y')
      I1 = 0
      DO 80 I=1,N1
        I1 = I1 + 1
        IF (I1 .LT. 51) GOTO 81
           WRITE(IO,31) (TIT(KK), KK=1,10)
           WRITE(IO,202)
 202       FORMAT(25X,'POPULACAO 1 - CONTINUACAO',/,15X,'MATRIZ Y')
           I1 = 0
 81     WRITE(IO,13) (Y1(I,J), J=1,NPOP) , POP1(I)
 13     FORMAT(2(5X,G12.4), I4 )
 80   CONTINUE

      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,14)
 14   FORMAT(25X,'POPULACAO 2',/,15X,'MATRIZ Y')
      I1 = 0
      DO 85 I=1,N2
        I1 = I1 + 1
        IF (I1 .LT. 51) GOTO 86
          WRITE(IO,31) (TIT(KK), KK=1,10)
          WRITE(IO,203)
 203      FORMAT(25X,'POPULACAO 2 - CONTINUACAO',/,15X,'MATRIZ Y')
          I1 = 0
 86     WRITE(IO,13) (Y2(I,J), J=1,NPOP) , POP2(I)
 85   CONTINUE
C
```

```
C  MEDIAS DE Y
C
      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,16)
 16   FORMAT(10X,'MEDIAS DE Y - POPULACAO 1 E 2',/)
      DO 90 I=1,2
        WRITE(IO,23) YMED1(I),YMED2(I)
 23     FORMAT(2(5X,F10.4))
 90   CONTINUE
C
C  COVARIANCIAS DE Y
C
      WRITE(IO,17)
 17   FORMAT(////,10X,'COVARIANCIA DE Y - POPULACAO 1',/)
      DO 100 I=1,2
       WRITE(IO,23) (YS1(I,J), J=1,2)
 100   CONTINUE
      WRITE(IO,18)
 18   FORMAT(///,10X,'COVARIANCIA DE Y - POPULACAO 2',/)
      DO 105 I=1,2
       WRITE(IO,23) (YS2(I,J), J=1,2)
 105  CONTINUE
C
C  FUNCAO DISCRIMINANTE
C
      WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,19)
 19   FORMAT(10X,'CLASSIFICACAO DOS DADOS DA POPULACAO 1',/)
      I1 = 0
      NAUX = (N1 / 3) * 3
      NAUX1 = N1 - NAUX
      DO 110 I=1,NAUX,3
       I1 = I1 + 1
       IF (I1 .LT. 51) GO TO 27
          WRITE(IO,31) (TIT(KK), KK=1,10)
          WRITE(IO,200)
          I1 = 0
 27    WRITE(IO,21) POP1(I), POPFIM(I,1), POP1(I+1), POPFIM(I+1,1),
     *POP1(I+2), POPFIM(I+2,1)
 21    FORMAT(3(4X,I6,4X,I6,4X))
 110  CONTINUE

      IF (NAUX1 .EQ. 0) GOTO 41
      IF (NAUX1 .EQ. 1) GOTO 42
      IF (NAUX1 .EQ. 2) GOTO 43
      GOTO 41

 42   WRITE(IO,29) POP1(N1), POPFIM(N1,1)
 29   FORMAT(4X,I6,4X,I6)
      GOTO 41
```

```
43    WRITE(IO,28) POP1(N1-1), POPFIM(N1-1,1), POP1(N1), POPFIM(N1,1)
28    FORMAT(2(4X,I6,4X,I6,4X))

41    WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,22)
22    FORMAT(10X,'CLASSIFICACAO DOS DADOS DA POPULACAO 2',/)
      I1 = 0
      NAUX = (N2 / 3) * 3
      NAUX1 = N2 - NAUX
      DO 115 I=1,NAUX,3
       I1 = I1 + 1
       IF (I1 .LT. 51) GO TO 116
          WRITE(IO,31) (TIT(KK), KK=1,10)
          WRITE(IO,201)
          I1 = 0
116    WRITE(IO,21) POP2(I), POPFIM(I,2), POP2(I+1), POPFIM(I+1,2),
     *POP2(I+2), POPFIM(I+2,2)
115   CONTINUE

      IF (NAUX1 .EQ. 0) GOTO 51
      IF (NAUX1 .EQ. 1) GOTO 52
      IF (NAUX1 .EQ. 2) GOTO 53
      GOTO 51

52    WRITE(IO,29) POP2(N2), POPFIM(N2,2)
      GOTO 51

53    WRITE(IO,28) POP2(N2-1), POPFIM(N2-1,2), POP2(N2), POPFIM(N2,2)
C
C  PERCENTAGENS DE ACERTO
C
51    WRITE(IO,31) (TIT(KK), KK=1,10)
      WRITE(IO,231)
231   FORMAT(10X,'PERCENTAGENS DE CLASSIFICACAO CORRETAS',/)
      WRITE(IO,24) PERC1
24    FORMAT(15X,F6.2,1X,'POR CENTO NA POPULACAO 1')
      WRITE(IO,26) PERC2
26    FORMAT(15X,F6.2,1X,'POR CENTO NA POPULACAO 2')

      RETURN
      END
```

SUBROTINA: GRAF

```
      SUBROUTINE GRAF(YGR,XGR,NXY,YG,INDORD,TIT)
      DIMENSION YGR(1), XGR(1), NXY(1), YG(1), INDORD(1), XL(6)
```

```
      INTEGER R(68), A(11), CHAR, BRANCO, TRACO, PONTO, DPONTO
      INTEGER IYCHAR, IXCHAR, TIT(10)
      DOUBLE PRECISION XGR, YGR, YG, XMX, XMN, XMIN, XMAX, XL

      DATA A(1),A(2),A(3),A(4),A(5) /2H  ,2H+ ,2H2 ,2H3 ,2H4 /
      DATA A(6),A(7),A(8),A(9),A(10) /2H5 ,2H6 ,2H7 ,2H8 ,2H9 /
      DATA A(11) /2H$ /
      DATA BRANCO,PONTO,TRACO,DPONTO / 2H  , 2H. , 2H- , 2H: /
      DATA NAXES,HAXIS,VAXIS /3,0.,0./
      DATA MAXMIN,YMAX,YMIN,XMAX,XMIN /0,0.,0.,0.,0./
C
C    IMPRIME O GRAFICO
C

      WRITE(7,1) (TIT(KK), KK=1,10)
 1    FORMAT('1',10X,10A4,/)
      WRITE(7,2)
 2    FORMAT(10X,'POPULACAO 1 = +',4X,'POPULACAO 2 = O',/)
      EPS = 1.0E-10
      NG = 2
      CHAR = A(2)
      NPOINT = 0

      DO 10 J=1,NG
 10   NPOINT = NPOINT + NXY(2*J-1)

      IF (NPOINT .LE. 1) THEN
        WRITE(7,3) NPOINT
 3      FORMAT(///,1H ,'ERRO NO GRAFICO, NUMERO DE PONTOS = ',I5,///)
        GOTO 450
      ENDIF

      NHORIZ = 68
      NVERT = 51

      DO 20 KI = 1,NPOINT
 20   YG(KI) = YGR(KI)
C
C    ORDENACAO DE Y - DECRESCENTE
C
      CALL SORTV2(NPOINT,YG,INDORD,-5)
      YMX = YG(1)
      YMN = YG(NPOINT)
C
C    MAXIMO E MINIMO DE X
C
      XMX = XGR(1)
      XMN = XGR(1)

      DO 30 I=2,NPOINT
        IF (XMX .LT. XGR(I)) XMX = XGR(I)
```

```
         IF (XMN .GT. XGR(I)) XMN = XGR(I)
   30  CONTINUE

       YMAX = YMX
       YMIN = YMN
       XMAX = XMX
       XMIN = XMN

       IF( YMIN .LT. YMAX  .AND.  XMIN .LT. XMAX) THEN
C
C  INTERVALO DE X  E   Y
C
          XQ = (XMAX - XMIN) / (NHORIZ - 2.)
          YQ = (YMAX - YMIN) / (NVERT - 1.)

          YY1 = YMAX + 0.5 * YQ
          YY2 = YMIN - 0.5 * YQ
          XZ1 = XMAX + 0.5 * XQ
          XZ2 = XMIN - 0.5 * XQ

          IF ( (YMX .LT. YY2) .OR. (YMN .GE. YY1) ) GOTO 116
          IF ( (XMX.LT.XZ2) .OR. (XMN.GE.XZ1) ) GOTO 116
          GOTO 118
 116   WRITE(7,4) YMIN,YMAX,XMIN,XMAX
       WRITE(7,6) YMN,YMX,XMN,XMX
 6     FORMAT(1H ,'TODOS OS PONTOS ESTAO FORA DOS DOMINIOS',/,
      *1H ,'YMIN = ',E12.5,'YMAX = ',E12.5,/,
      *1H ,'XMIN = ',E12.5,'XMAX = ',E12.5,///)
       GOTO 440
       ELSE
       WRITE(7,4) YMIN,YMAX,XMIN,XMAX
 4     FORMAT(///,1H ,'ERRO NO DOMINIO DO GRAFICO',/,
      *'VALORES CALCULADOS SAO ',/, 'YMIN = ',E12.5, 'YMAX = ',E12.5,/,
      *'XMIN = ',E12.5,'XMAX = ',E12.5,/)
       GOTO 440
       ENDIF
C
C  PONTOS A SEREM IMPRESSOS NO EIXO X
C
 118   XL(1) = XMIN
       NXL = 6

       DO 40 I=2,NXL
  40   XL(I) = XL(I-1) + 20. * XQ

       XMN = XZ2
       IF ( XL(1) .LE. 1.E-10 ) THEN
          IF ( XL(NXL) .GE. -1.E-10 ) THEN
             DO 50 I=1,NXL
                IF( ABS(XL(I)) .LE. 1.E-10 )  XL(I) = 0
  50         CONTINUE
```

```
         ENDIF
      ENDIF

C    IMPRIMINDO A ESCALA DE X

      WRITE(7,7) (XL(I), I=1,6)
 7    FORMAT( 2X,F10.3,5(3X,f10.3) )
      WRITE(7,8)
 8    FORMAT(1H ,8X,1H.,5(12X,1H.))

      DO 60 J=1,NHORIZ
 60   R(J) = PONTO
C
C    IMPRIMINDO UMA LINHA DE PONTOS
      WRITE(7,9) (R(J), J=1,NHORIZ)

      NP = 0
      IYCHAR = BRANCO
      IXCHAR = BRANCO
      JCOL = NHORIZ / 2
      JLINE = NVERT + 1
C
C   CONTANDO QTOS PONTOS ESTAO FORA DO GRAFICO
C
      DO 70 I=1,NPOINT
       IF( YG(I) .LT. YY1) GOTO 214
  70  CONTINUE

      I = NPOINT + 1
 214  IY = I
      NCOUNT = I - 1

      DO 300 N=1,NVERT
        YY2 = YY1
        YY1 = YY2 - YQ

        DO 80 J=1,NHORIZ
  80    R(J) = 0

 220    IF( IY .LE. NPOINT ) THEN
            IF( YG(IY) .GE. YY1 ) THEN
                IX = INDORD(IY)
                JJ = ( XGR(IX) - XMN ) / XQ + 1.
                IF ( JJ .LT. 1  .OR.  JJ .GT. (NHORIZ-1) ) THEN
                   NCOUNT = NCOUNT + 1
                   IY = IY + 1
                   GOTO 220
                ENDIF
                K = R(JJ)
                IF (K .LT. 0) THEN
                   R(JJ) = 2
```

```
              ELSE
                IF ( K  .EQ. 0) THEN
                  R(JJ) = -IX
                ELSE
                  R(JJ) = K + 1
                ENDIF
              ENDIF
              IY = IY + 1
              GOTO 220
          ENDIF
      ENDIF
      DO 240 J=1,NHORIZ
        K = R(J)
        IF (K .LT. 0) THEN
           NPT = 0
           K = -K
           DO 90 I=1,NG
             NPT = NPT + NXY(2 * I - 1)
             IF (K .LE. NPT) THEN
               CHAR = NXY(2 * I)
               GOTO 235
             ENDIF
 90        CONTINUE
             CHAR = A(2)
235        R(J) = CHAR
           GOTO 240
        ELSE
           IF (K .GE. 10) K = 10
           R(J) = A(K+1)
        ENDIF
240   CONTINUE
      IF( R(NHORIZ) .EQ. BRANCO ) R(NHORIZ) = PONTO
      IF( R(JCOL) .EQ. BRANCO ) R(JCOL) = IXCHAR
      IF( N-10*NP .EQ. 1 ) GOTO 270
      IF( N .EQ. NVERT ) GOTO 271
      IF (N .EQ. JLINE) GOTO 272
      WRITE(7,9) (R(J),J=1,NHORIZ)
9     FORMAT(1H ,7X,1H.,68A1 )
      GOTO 300
270   NP = NP + 1
271   IF (N .NE. JLINE) GOTO 275
272   YM = HAXIS
      DO 273 J=1,NHORIZ
       IF(R(J) .EQ. BRANCO) R(J) = IYCHAR
273   CONTINUE
      GOTO 276
275   YM = YY1 + 0.5 * YQ
276   IF ( ABS(YM) .LE. 1.E-10) YM = 0
      WRITE(7,290) YM, (R(J),J=1,NHORIZ)
290   FORMAT(1H ,f7.2,1H.,68A1)
300 CONTINUE
```

```
      NCOUNT = NCOUNT + NPOINT + 1 - IY
      DO 310 J=1,NHORIZ
 310  R(J) = PONTO
      GOTO 350
 350  WRITE(7,9) (R(J),J=1,NHORIZ)
      WRITE(7,8)
      WRITE(7,7) (XL(I),I=1,6)
      WRITE(7,410) XQ, YQ
 410  FORMAT(1H ,'INTERVALO HORIZONTAL = ',G10.3,
     *7X,'INTERVALO VERTICAL = ',G10.3)
      IF (NCOUNT .EQ. 0) GOTO 440
      ICOUNT = NPOINT - NCOUNT
      WRITE(7,412) ICOUNT, NCOUNT
 412  FORMAT(1H ,1X,'NUM. DE PONTOS GRAFICADOS = ',I4,
     *5X,'NUM. DE PONTOS FORA DO GRAFICO = ',I4)
 440  CONTINUE
 450  MAXMIN = 0
      NAXES = 0

      RETURN
      END
```

SUBROTINA: SORTV2

```
      subroutine sortv2(nl,yg,indord,order)
      dimension yg(1), indord(1)
      double precision yg, ya, yb
      logical exch, dir
      integer order
c
c  ordenacao de um vetor
c
      do 15 i=1,nl
 15      indord(i) = i

       dir = (order .ge. 0)
      jgap = nl

 5       if (jgap .le. 1) goto 100
       jgap = jgap / 2
      jmax = nl - jgap
 10       exch = .false.

       do 20 j=1,jmax
        jplusg = j + jgap
        ya = yg(j)
```

```
         yb = yg(jplusg)
        if((dir .and. ya.le.yb).or.((.not. dir) .and. yb.le.ya)) goto 20
         ya = yg(j)
         yg(j) = yg(jplusg)
         yg(jplusg) = ya
         itroc = indord(j)
         indord(j) = indord(jplusg)
         indord(jplusg) = itroc
         exch = .true.
20        continue

        if (exch) goto 10
       goto 5

100        return
        end
```

SUBROTINA: DISCOV

```
      SUBROUTINE DISCOV(YYAUX,D1,YMED1,YMED2,CONST,FUNCAO)
      DIMENSION YYAUX(1,2), YYAUXT(2,1), T1(2,2), D1(2,2), T2(1,2)
      DIMENSION T3(2,1),YMED2(2),YMED1(2), T5(2,1)
      DOUBLE PRECISION YYAUX,YYAUXT,T1,T2,T3,T5,D1,YMED1,YMED2,
     *TERMO1,TERMO2,FUNCAO,CONST

      DO 10 I=1,2
        YYAUXT(I,1) = YYAUX(1,I)
 10   CONTINUE
C
C     TERMO 1  y' * (s2-1 - i2) * y
C
      DO 170 K=1,2
        DO 175 L=1,2
 175    T1(K,L) = D1(K,L)
 170  CONTINUE


      DO 180 K=1,2
 180  T1(K,K) = T1(K,K) - 1.0D0

      CALL MULT(1,2,2,YYAUX,T1,T2)
      CALL MULT(1,2,1,T2,YYAUXT,TERMO1)
C
C     TERMO 2: y' * (ymed1 - sy2-1 * ymed2)
C
```

```
      DO 191 K=1,2
 191  T5(K,1) = YMED2(K)

      CALL MULT(2,2,1,D1,T5,T3)

      DO 185 K=1,2
 185  T3(K,1) = YMED1(K) - T3(K,1)

      CALL MULT(1,2,1,YYAUX,T3,TERMO2)
C
C   FUNCAO: termo1 + 2 * termo2 - constante
C
      FUNCAO = TERMO1 + 2*TERMO2 - CONST

      RETURN
      END
```

SUBROTINA: IMPCOV

```
      SUBROUTINE IMPCOV(N1,N2,IP,NPOP,VALOR,VETOR,Y1,Y2,YMED1,YMED2,
     *YS1,YS2,POP1,POP2,POPFIM,PERC1,PERC2,TIT,IO)
      DIMENSION VALOR(IP),VETOR(IP,IP)
      DIMENSION Y1(N1,2),Y2(N2,2),YMED1(NPOP),YMED2(NPOP)
      DIMENSION YS1(NPOP,NPOP), YS2(NPOP,NPOP)
      INTEGER POP1(N1), POP2(N2), POPFIM(N1,2), FMT(20), TIT(10)
      INTEGER P
      DOUBLE PRECISION VALOR, VETOR, Y1, Y2, YMED1,YMED2,
     *YS1,YS2

      P = IP
C
C        IMPRESSAO DOS RESULTADOS: METODO DA COVARIANCIA
C
C
C   AUTO-VALORES E AUTO-VETORES
C
      WRITE(IO,31) ( TIT(KK), KK=1,10)
 31   FORMAT('1',10X,10A4,5X,'METODO DAS COVARIANCIAS',/)
      WRITE(IO,9)
 9    FORMAT(10X,'AUTO-VALORES DE INV(S1) * S2',/)
      WRITE(IO,6) (VALOR(I), I=1,P)
 6    FORMAT(9(F8.2))
      WRITE(IO,71)
 71   FORMAT(////,10X,'AUTO-VETORES DE INV(S1) * S2',/)
      DO 70 I=1,P
 70   WRITE(IO,6) (VETOR(I,J), J=1,P)
```

```
C
C    IMPRIMINDO Y
C
      ALFA1 = ALFA1 * 90.D0 / 3.1417
      ALFA2 = ALFA2 * 90.D0 / 3.1417

      WRITE(io,31) ( TIT(KK), KK=1,10)
      WRITE(io,72)
 72   FORMAT(25X,'POPULACAO 1',/,15X,'MATRIZ Y',/)
      I1 = 0
      DO 80 I=1,N1
        I1 = I1 + 1
        IF (I1 .LT. 51) GOTO 81
           WRITE(io,31) ( TIT(KK), KK=1,10)
           WRITE(io,202)
 202       FORMAT(25X,'POPULACAO 1 - CONTINUACAO',/,15X,'MATRIZ Y',/)
           I1 = 0
 81     WRITE(io,13) (Y1(I,J), J=1,NPOP) , POP1(I)
 13     FORMAT(2(5X,G12.4), I4)
 80   CONTINUE

      WRITE(io,31) ( TIT(KK), KK=1,10)
      WRITE(io,14)
 14   FORMAT(25X,'POPULACAO 2',/,15X,'MATRIZ Y',/)
      I1 = 0
      DO 85 I=1,N2
        I1 = I1 + 1
        IF (I1 .LT. 51) GOTO 86
           WRITE(io,31) ( TIT(KK), KK=1,10)
           WRITE(io,203)
 203       FORMAT(25X,'POPULACAO 2 - CONTINUACAO',/,15X,'MATRIZ Y',/)
           I1 = 0
 86     WRITE(io,13) (Y2(I,J), J=1,NPOP) , POP2(I)
 85   CONTINUE
C
C  MEDIAS DE Y
C
      WRITE(io,31) ( TIT(KK), KK=1,10)
      WRITE(io,16)
 16   FORMAT(10X,'MEDIAS DE Y - POPULACAO 1 E 2',/)
      DO 90 I=1,2
        WRITE(io,23) YMED1(I),YMED2(I)
 23     FORMAT(2(5X,F10.4))
 90   CONTINUE
C
C  COVARIANCIAS DE Y
C

      WRITE(io,17)
 17   FORMAT(////,10X,'COVARIANCIA DE Y - POPULACAO 1',/)
      DO 100 I=1,2
```

```
       WRITE(io,23) (YS1(I,J), J=1,2)
 100   CONTINUE
      WRITE(io,18)
 18   FORMAT(////,10X,'COVARIANCIA DE Y - POPULACAO 2',/)
      DO 105 I=1,2
       WRITE(io,23) (YS2(I,J), J=1,2)
 105  CONTINUE
C
C  FUNCAO DISCRIMINANTE
C
      WRITE(io,31) ( TIT(KK), KK=1,10)
      WRITE(io,19)
 19   FORMAT(10X,'CLASSIFICACAO DOS DADOS DA POPULACAO 1',/)
      I1 = 0
      NAUX = (N1 / 3) * 3
      NAUX1 = N1 - NAUX
      DO 110 I=1,NAUX,3
       I1 = I1 + 1
       IF (I1 .LT. 51) GOTO 27
          WRITE(io,31) (TIT(KK), KK=1,10)
          WRITE(io,200)
 200      FORMAT(15X,'POPULACAO 1 - CONTINUACAO',/)
          I1 = 0
 27    WRITE(io,21) POP1(I), POPFIM(I,1), POP1(I+1), POPFIM(I+1,1),
     *POP1(I+2), POPFIM(I+2,1)
 21    FORMAT(3(4X,I6,4X,I6,4X))
 110  CONTINUE

      IF (NAUX1 .EQ. 0) GOTO 41
      IF (NAUX1 .EQ. 1) GOTO 42
      IF (NAUX1 .EQ. 2) GOTO 43
      GO TO 41

 42   WRITE(io,29) POP1(N1), POPFIM(N1,1)
 29   FORMAT(4X,I6,4X,I6)
      GOTO 41

 43   WRITE(io,28) POP1(N1-1), POPFIM(N1-1,1),POP1(N1), POPFIM(N1,1)
 28   FORMAT(2(4X,I6,4X,I6,4X))

 41   WRITE(io,31) (TIT(KK), KK=1,10)
      WRITE(io,22)
 22   FORMAT(10X,'CLASSIFICACAO DOS DADOS DA POPULACAO 2',/)
      I1 = 0
      NAUX = (N2 / 3) * 3
      NAUX1 = N2 - NAUX
      DO 115 I=1,NAUX,3
       I1 = I1 + 1
       IF (I1 .LT. 51) GOTO 116
          WRITE(io,31) (TIT(KK), KK=1,10)
          WRITE(io,201)
```

```
201        FORMAT(15X,'POPULACAO 2 - CONTINUACAO',/)
           I1 = 0
116    WRITE(io,21) POP2(I), POPFIM(I,2), POP2(I+1), POPFIM(I+1,2),
     *POP2(I+2), POPFIM(I+2,2)
115   CONTINUE

      IF (NAUX1 .EQ. 0) GOTO 51
      IF (NAUX1 .EQ. 1) GOTO 52
      IF (NAUX1 .EQ. 2) GOTO 53
      GO TO 51

52    WRITE(io,29) POP2(N2), POPFIM(N2,2)
      GOTO 51

53    WRITE(io,28) POP2(N2-1), POPFIM(N2-1,2),POP2(N2), POPFIM(N2,2)

C
C  PERCENTAGENS DE ACERTO
C
51    WRITE(io,31) (TIT(KK), KK=1,10)
      WRITE(io,231)
231   FORMAT(10X,'PERCENTAGENS DE CLASSIFICACAO CORRETAS',/)
      WRITE(io,24) PERC1
24    FORMAT(15X,F6.2,1X,'POR CENTO NA POPULACAO 1')
      WRITE(io,26) PERC2
26    FORMAT(15X,F6.2,1X,'POR CENTO NA POPULACAO 2')

      RETURN
      END
```

D. UM EXEMPLO DE SAIDA DO PROGRAMA COMPUTACIONAL

ISOLADOS METODO DAS CORRELACOÉS

POPULACAO 1

32.0	26.0	2.0	52.0	36.5	101
43.0	31.0	2.2	37.0	58.1	102
44.0	28.0	2.0	73.0	23.6	103
43.0	30.0	2.3	75.0	32.6	104
32.0	25.0	1.8	39.0	26.3	105
51.0	33.0	1.9	94.0	38.0	106
30.0	15.0	1.5	81.0	18.6	107
33.0	15.0	2.6	78.0	23.8	108
47.0	23.0	2.3	46.0	42.0	109
44.0	25.0	2.4	93.0	24.5	110
46.0	30.0	2.2	79.0	32.8	111
53.0	32.0	2.6	81.0	63.8	112
61.0	26.0	2.5	75.0	54.2	113
36.0	20.0	1.8	37.0	73.1	114
34.0	21.0	2.2	58.0	36.8	115
36.0	22.0	2.0	39.0	33.2	116
38.0	24.0	1.7	92.0	27.6	117
42.0	25.0	1.5	76.0	32.0	118
51.0	31.0	2.0	76.0	36.2	119
43.0	27.0	2.2	85.0	20.3	120

ISOLADOS METODO DAS CORRELACOES

POPULACAO 2

53.0	31.0	3.0	2.0	62.8	201
56.0	30.0	3.1	2.4	54.3	202
58.0	32.0	3.2	4.0	61.1	203
50.0	31.0	3.0	4.0	66.3	204
56.0	32.0	3.3	4.5	69.8	205
58.0	33.0	3.2	4.0	73.2	206
56.0	32.0	3.2	3.0	81.6	207
58.0	30.0	3.3	3.5	80.3	208
53.0	31.0	3.0	4.0	58.6	209
51.0	30.0	3.0	4.5	61.6	210
56.0	32.0	3.0	5.0	69.3	211
54.0	30.0	3.1	5.0	74.5	212
55.0	31.0	2.9	4.0	76.3	213
56.0	32.0	3.2	4.0	55.0	214
58.0	33.0	3.2	5.0	54.3	215
52.0	30.0	3.0	3.0	50.6	216
55.0	32.0	3.1	5.0	83.4	217
58.0	32.0	3.2	5.0	81.6	218
43.0	30.0	3.0	3.0	50.3	219
56.0	31.0	3.1	2.5	85.0	220

VETORES DE MEDIAS DAS POPULACOES 1 E 2

41.9500	54.6000
25.4500	31.2500
2.0850	3.1050
68.3000	3.8700
36.7000	67.4950

MATRIZ DE COVARIANCIA DA POPULACAO 1

67.00	28.18	1.183	58.81	46.13
28.18	26.37	.3229	18.96	19.33
1.183	.3229	.1056	.5995	1.147
58.81	18.96	.5995	390.4	-123.1
46.13	19.33	1.147	-123.1	220.8

MATRIZ DE COVARIANCIA DA POPULACAO 2

13.20	2.105	.2600	.8874	18.73
2.105	1.039	.5132E-01	.4079	2.512
.2600	.5132E-01	.1313E-01	.1911E-01	.3548
.8874	.4079	.1911E-01	.9064	2.192
18.73	2.512	.3548	2.192	134.4

MATRIZ DE CORRELACAO DA POPULACAO 1

1.000	.6705	.4450	.3636	.3793
.6705	1.000	.1936	.1869	.2534
.4450	.1936	1.000	.9338E-01	.2376
.3636	.1869	.9338E-01	1.000	-.4194
.3793	.2534	.2376	-.4194	1.000

MATRIZ DE CORRELACAO DA POPULACAO 2

1.000	.5683	.6245	.2565	.4448
.5683	1.000	.4392	.4202	.2126
.6245	.4392	1.000	.1751	.2671
.2565	.4202	.1751	1.000	.1987
.4448	.2126	.2671	.1987	1.000

TESTE DE IGUALDADE DAS MATRIZES DE COVARIANCIAS

ESTATISTICA DO TESTE (VC) = 137.2

NUMERO DE GRAUS DE LIBERDADE = 15

PROB (QUI < QUI CALCULADO) = 1.00

CONCLUSAO:
REJEITO A IGUALDADE DAS COVARIANCIAS AOS NIVEIS DE 5% E 1%

TESTE DE IGUALDADE DE MEDIAS

ESTATISTICA DO TESTE(T QUADRADO) = 485.9

NUMEROS DE GRAUS DE LIBERDADE: 5 E 34

ESTATISTICA F (CALCULADO) = 86.9446

PROB (F < F CALCULADO) = 1.00

CONCLUSAO:
REJEITO A IGUALDADE DE MEDIAS AOS NIVEIS DE 5% E 1%

TESTE DE IGUALDADE DAS MATRIZES DE CORRELACOES

ESTATISTICA DO TESTE = 1.890

B(ALFA=0.05) = .8394

B(ALFA=0.01) = 1.652

CONCLUSAO:
REJEITO A IGUALDADE DAS CORRELACOES AOS NIVEIS DE 5% E 1%

AUTO-VALORES DE INV(R1) * R2

3.439	1.410	1.124	.5851	.4529

AUTO-VETORES DE INV(R1) * R2

-1.345	-.8051	-.6253	-.6179	.3658
.5128	1.111	.5046E-01	-.4335	-.4498
.1097	.5383	-.5320	.8407	.1683E-01
1.286	-.1464	.3580E-02	.2318E-01	.5781
1.176	-.5868	-.1016	.1147	-.5176

ISOLADOS METODO DAS CORRELACOES

POPULACAO 1

MATRIZ Y

4.289	-.4943	101
3.782	-1.622	102
2.863	.9305	103
4.171	.4711	104
2.468	-.4420	105
4.685	.9128	106
3.821	1.826	107
3.915	1.748	108
1.671	.8718E-01	109
4.071	1.768	110
3.920	.7101	111
5.689	-.1531	112
2.591	.8837	113
4.886	-1.514	114
3.943	.2085	115
2.125	-.2305	116
4.901	1.414	117
3.583	.8732	118
3.205	.6293	119
3.514	1.450	120

ISOLADOS METODO DAS CORRELACOES

POPULACAO 2

MATRIZ Y

7.921	-9.490	201
6.081	-8.110	202
9.294	-8.108	203
12.09	-8.734	204
11.69	-8.380	205
11.02	-9.090	206
10.76	-9.832	207
9.657	-8.372	208
10.20	-8.088	209
11.41	-7.679	210
12.03	-8.098	211
12.38	-7.635	212
11.16	-8.692	213
9.415	-8.037	214
10.46	-7.639	215
7.901	-7.998	216
13.92	-8.814	217
12.73	-8.417	218
11.20	-8.890	219
9.835	-9.861	220

MEDIAS DE Y - POPULACAO 1 E 2

3.7046	10.5578
.4728	-8.4982

COVARIANCIA DE Y - POPULACAO 1

1.0000	.0000
.0000	1.0000

COVARIANCIA DE Y - POPULACAO 2

3.4386	.0000
.0000	.4529

CLASSIFICACAO DOS DADOS DA POPULACAO 1

101	1	102	1	103	1
104	1	105	1	106	1
107	1	108	1	109	1
110	1	111	1	112	1
113	1	114	1	115	1
116	1	117	1	118	1
119	1	120	1		

ISOLADOS METODO DAS CORRELACOES

CLASSIFICACAO DOS DADOS DA POPULACAO 2

201	2	202	2	203	2
204	2	205	2	206	2
207	2	208	2	209	2
210	2	211	2	212	2
213	2	214	2	215	2
216	2	217	2	218	2
219	2	220	2		

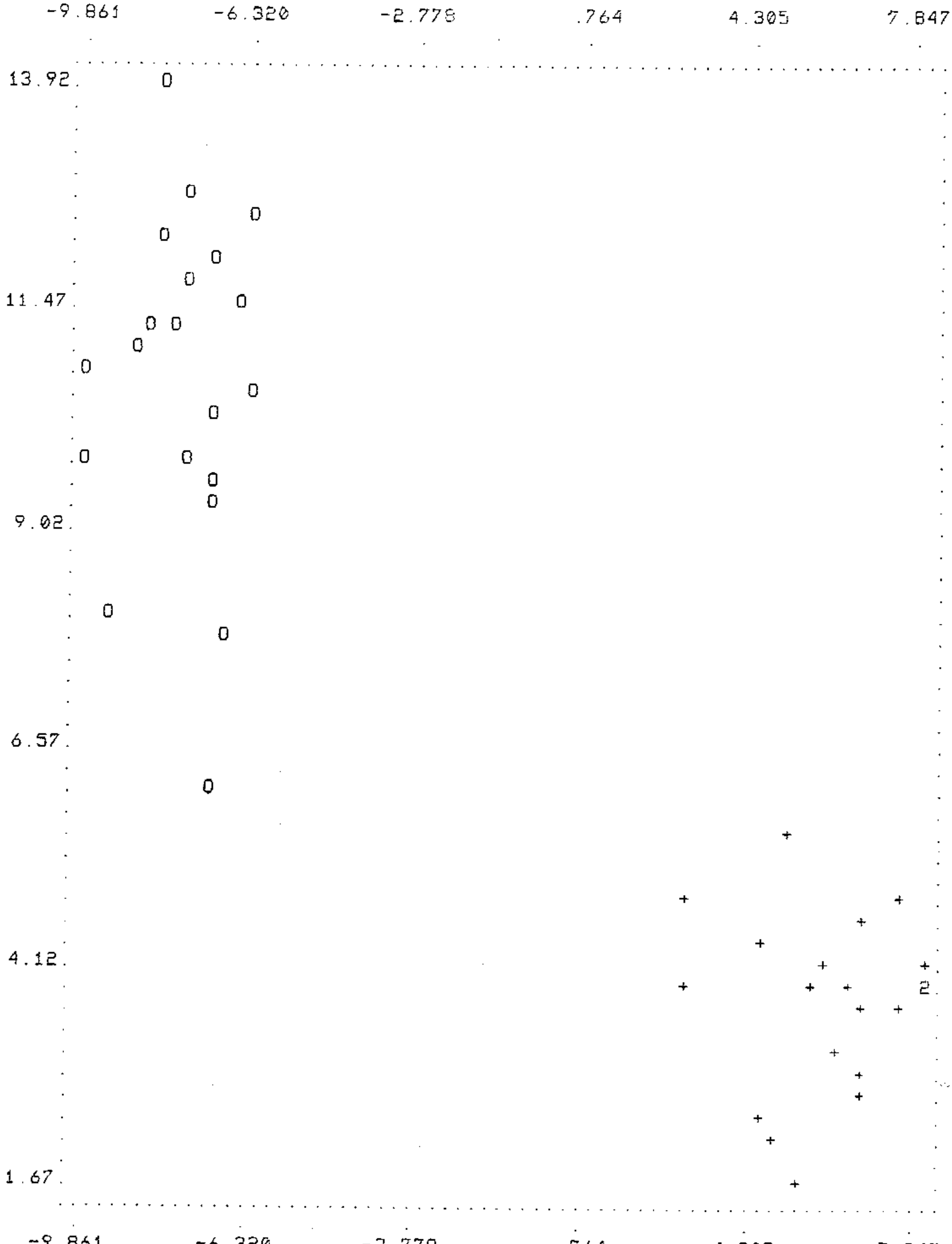
ISOLADOS METODO DAS CORRELACOES

PERCENTAGENS DE CLASSIFICACAO CORRETAS

100.00 POR CENTO NA POPULACAO 1
100.00 POR CENTO NA POPULACAO 2

ISOLADOS

POPULACAO 1 = +   POPULACAO 2 = 0



INTERVALO HORIZONTAL = .177   INTERVALO VERTICAL = .245

AUTO-VALORES DE INV(S1) * S2

1.01	.31	.10	.03	.00

AUTO-VETORES DE INV(S1) * S2

-.10	-.19	.07	-.01	.00
.01	.15	-.07	.21	.02
-.14	.45	-3.40	-.66	-.01
.05	.02	-.00	-.00	-.05
.10	.01	.00	.00	.00

POPULACAO 1

MATRIZ Y

2.562	-2.156	101
2.872	-1.290	102
1.122	-3.207	103
2.135	-3.264	104
1.027	-1.519	105
2.802	-4.174	106
2.371	-3.857	107
2.281	-3.711	108
1.296	-1.921	109
2.036	-4.287	110
2.057	-3.470	111
4.374	-3.520	112
2.391	-3.346	113
4.982	-1.486	114
2.611	-2.563	115
1.250	-1.578	116
2.966	-4.243	117
2.309	-3.407	118
1.793	-3.298	119
1.412	-3.840	120

POPULACAO 2

MATRIZ Y

.6523	.4819	201
-.4520	.4323	202
.7015E-01	.3924	203
1.368	.3853	204
1.103	.3743	205
1.229	.4206	206
2.174	.4597	207
1.853	.3914	208
.3421	.3770	209
.8400	.3359	210
1.121	.3505	211
1.786	.3160	212
1.849	.3884	213
-.3167	.3899	214
-.5277	.3563	215
-.3738	.4036	216
2.548	.3608	217
2.070	.3560	218
.4738	.4118	219
2.485	.4686	220

MEDIAS DE Y - POPULACAO 1 E 2

2.3324	1.0147
-3.0067	.3926

COVARIANCIA DE Y - POPULACAO 1

1.0000	.0000
.0000	1.0000

COVARIANCIA DE Y - POPULACAO 2

1.0097	.0000
.0000	.0019

ISOLADOS METODO DAS COVARIANCIAS

CLASSIFICACAO DOS DADOS DA POPULACAO 1

101	1	102	1	103	1
104	1	105	1	106	1
107	1	108	1	109	1
110	1	111	1	112	1
113	1	114	1	115	1
116	1	117	1	118	1
119	1	120	1		

ISOLADOS METODO DAS COVARIANCIAS

CLASSIFICACAO DOS DADOS DA POPULACAO 2

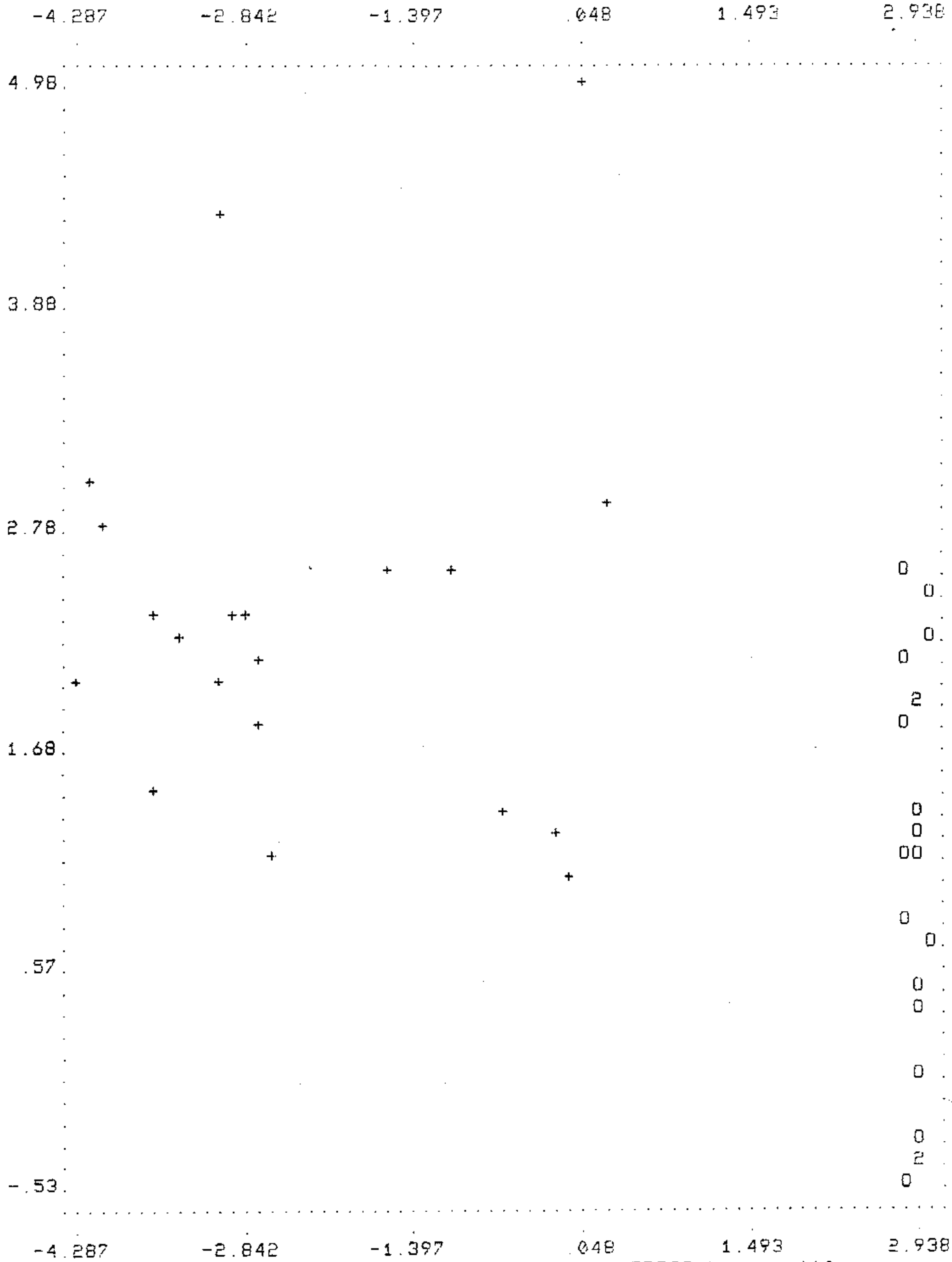
201	2	202	2	203	2
204	2	205	2	206	2
207	2	208	2	209	2
210	2	211	2	212	2
213	2	214	2	215	2
216	2	217	2	218	2
219	2	220	2		

ISOLADOS METODO DAS COVARIANCIAS

PERCENTAGENS DE CLASSIFICACAO CORRETAS

100.00 POR CENTO NA POPULACAO 1
100.00 POR CENTO NA POPULACAO 2

ISOLADOS
POPULACAO 1 = +   POPULACAO 2 = 0



INTERVALO HORIZONTAL = .723E-01   INTERVALO VERTICAL = .110

E. GRÁFICOS OBTIDOS PELOS MÉTODOS DISCRIMINANTES BASEADO NO QUOCIENTE DAS MATRIZES DE CORRELAÇÕES E COVARIÂNCIAS

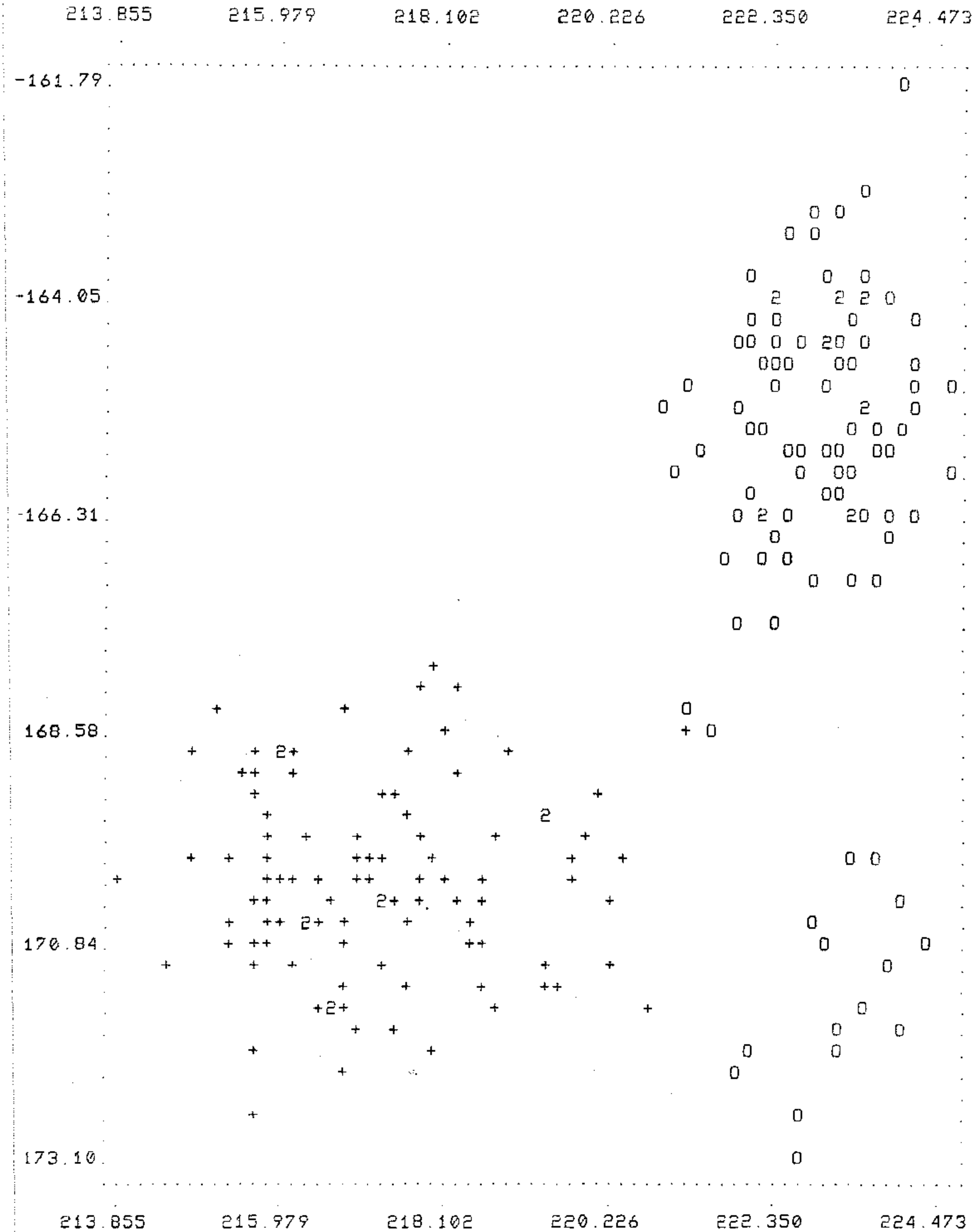
Neste anexo apresentaremos os gráficos obtidos através da execução do programa computacional apresentado no anexo C para os conjuntos de dados analisados.

Esses gráficos permitiram termos uma visualização gráfica da discriminação das populações.

```
     DADOS  DO FLURY - NOTAS FALSAS E VERDAD      METODO DAS CORRELACOES
         POPULACAO 1 = +     POPULACAO 2 = 0

   -155.069     -135.880     -116.691      -97.502      -78.312      -59.123
         .            .            .            .            .            .
        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-259.85.                                                                  4322.
       .                                                                +9$$$9.
       .                                                                     +.
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
-291.92.                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
-323.99.                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
-356.06.                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
-388.12.                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .                                                                      .
       .030                                                                   .
       .$$$                                                                   .
       .2$3                                                                   .
-420.19.0                                                                     .
        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
         .            .            .            .            .            .
   -155.069     -135.880     -116.691      -97.502      -78.312      -59.123
INTERVALO HORIZONTAL =    .959           INTERVALO VERTICAL =    3.21
```

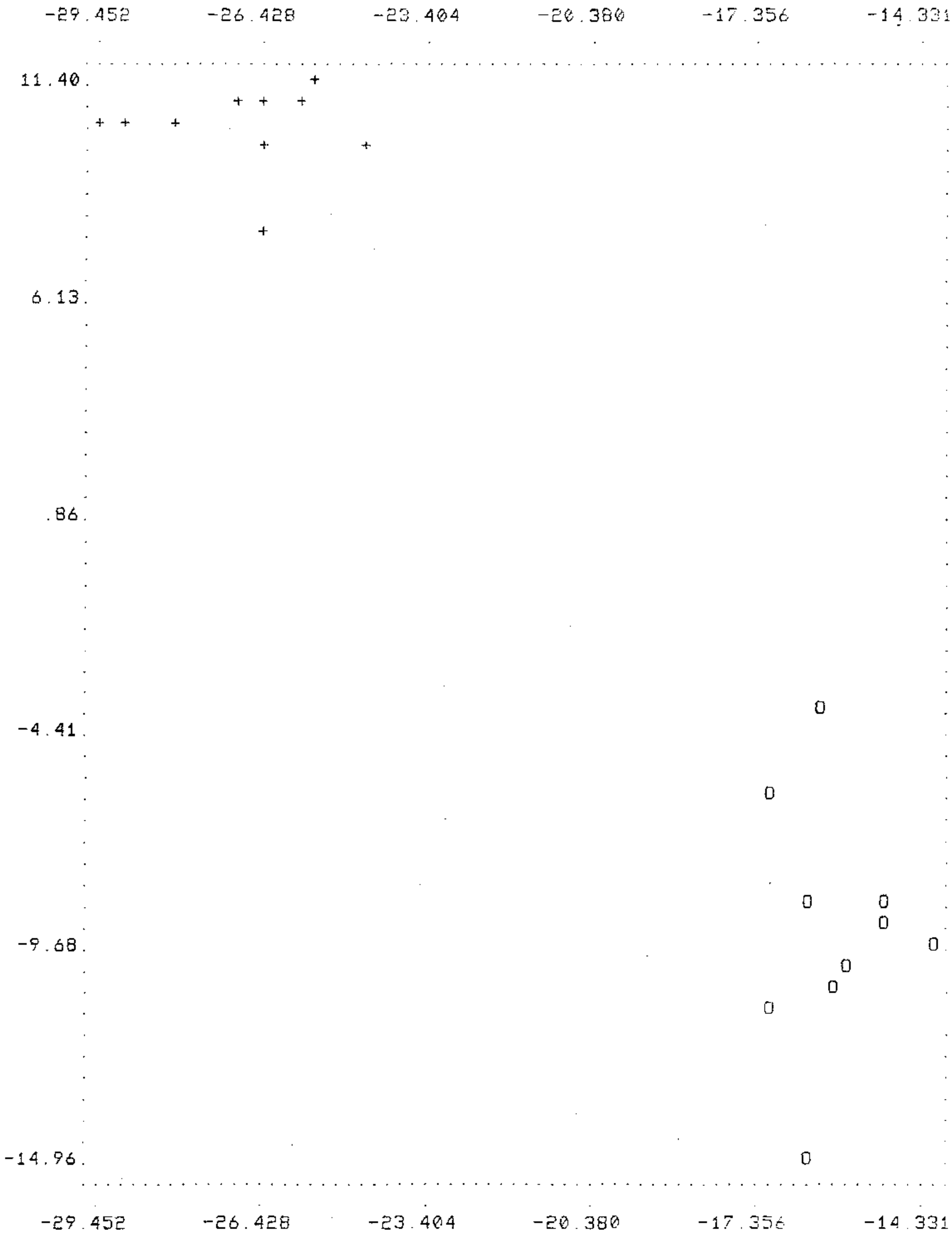
DADOS DO FLURY - NOTAS FALSAS E VERDAD   METODO DAS COVARIANCIAS
POPULACAO 1 = +   POPULACAO 2 = 0



NTERVALO HORIZONTAL = .106   INTERVALO VERTICAL = .226

DADOS DE BOVINO METODO DAS CORRELACOES

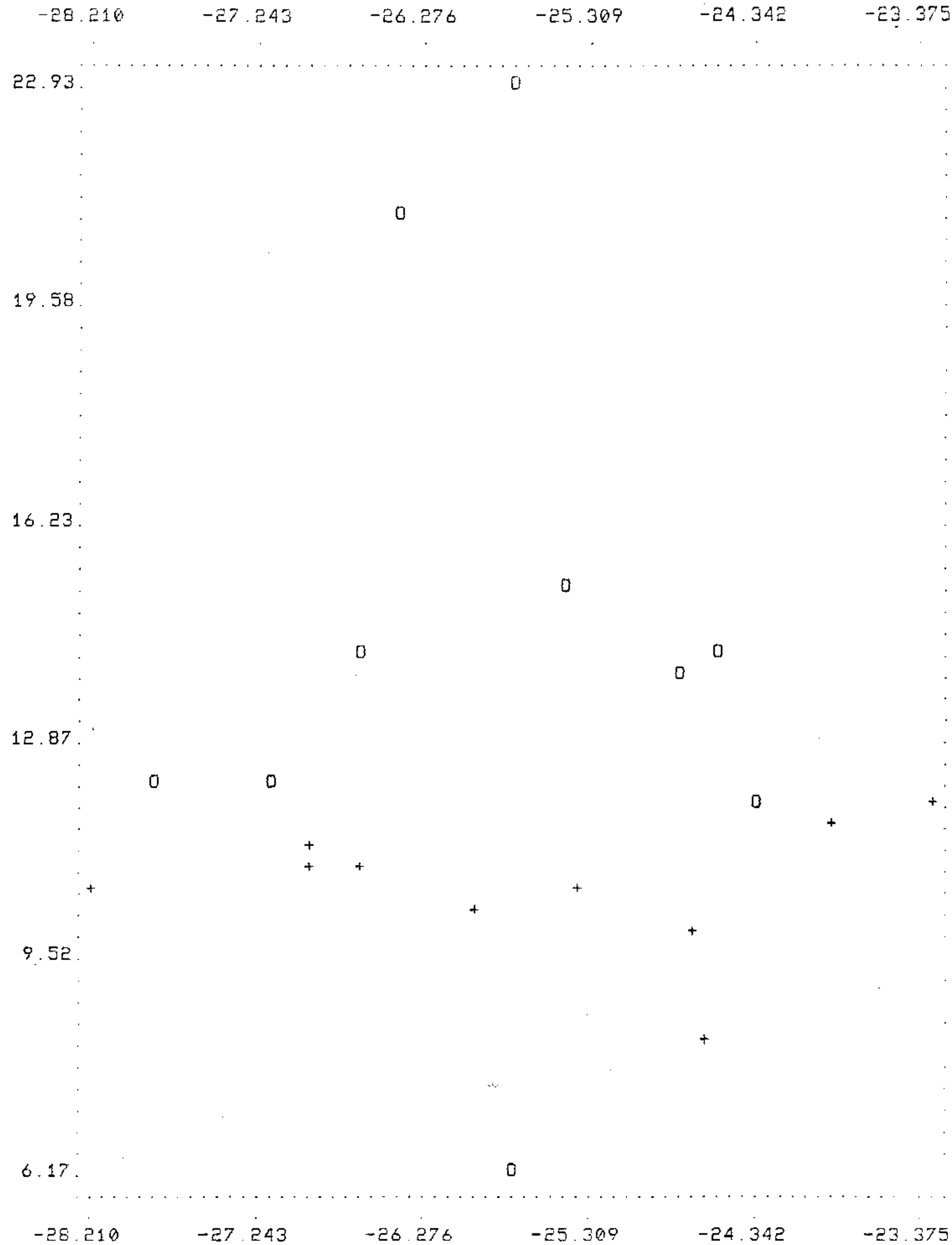
POPULACAO 1 = + POPULACAO 2 = 0



NTERVALO HORIZONTAL = .151 INTERVALO VERTICAL = .527

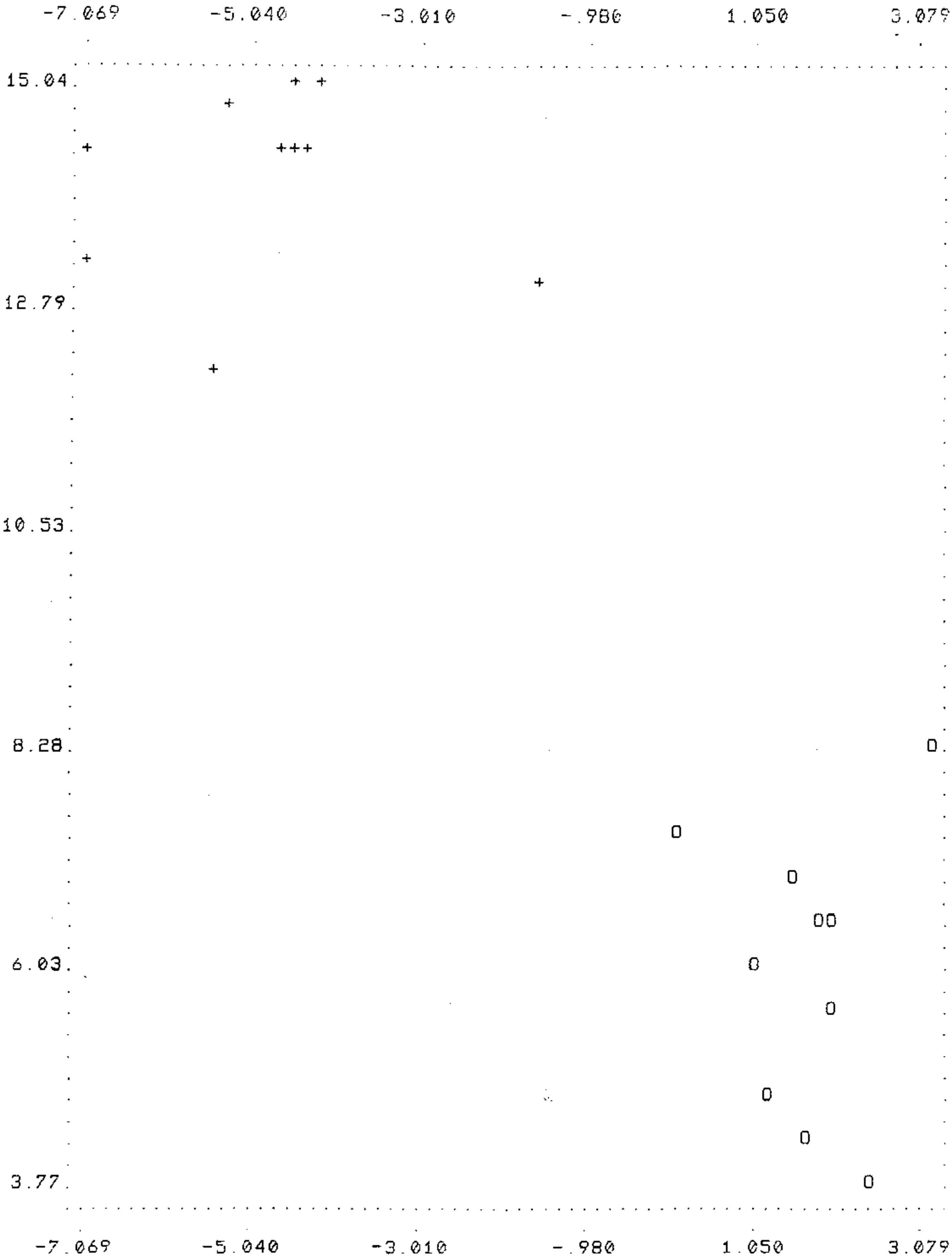
DADOS DO BOVINO METODO DAS COVARIANCIAS

POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .484E-01 INTERVALO VERTICAL = .335

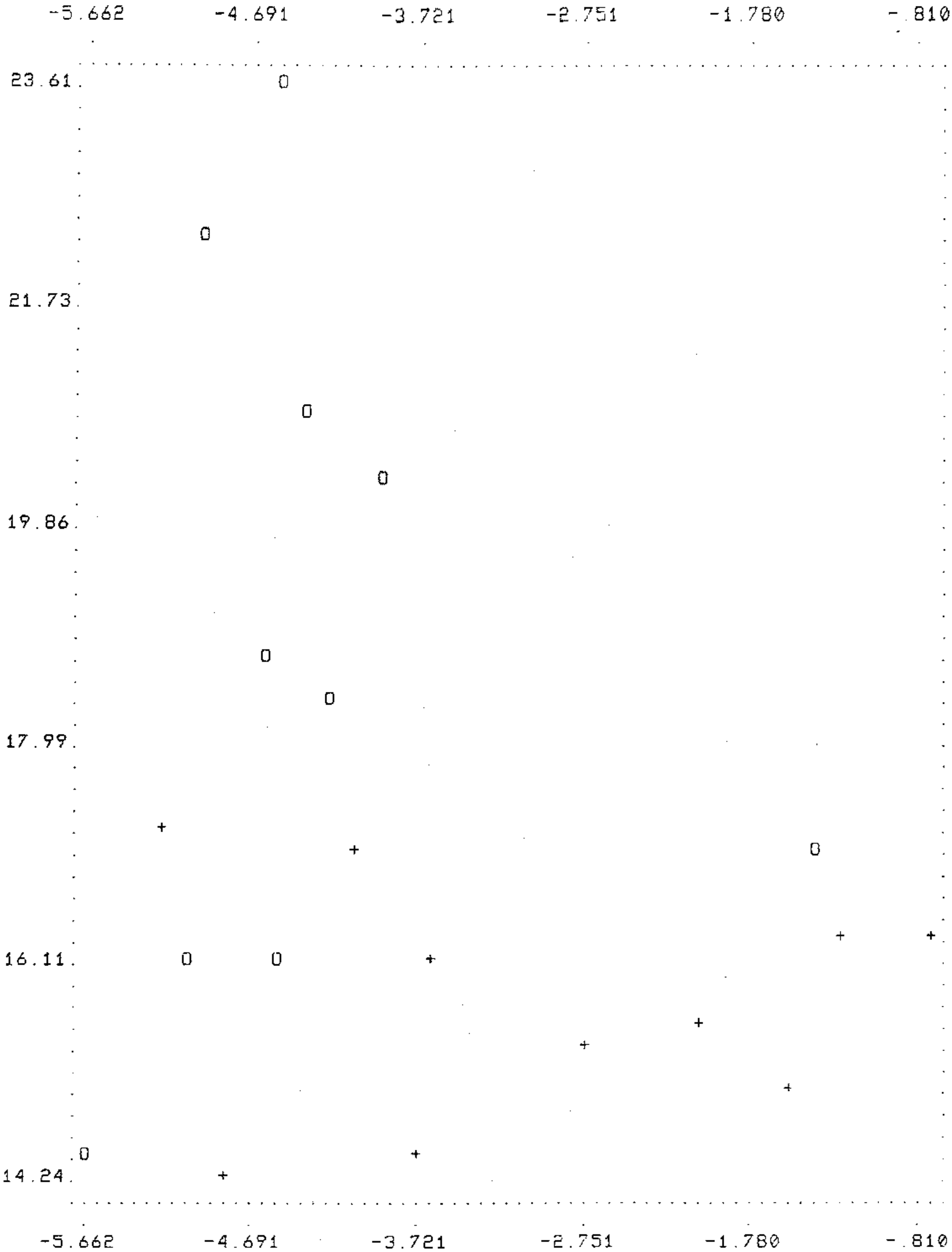
DADOS DO IGI - GORDURA INDUSTRIAL METODO DAS CORRELACOES
POPULACAO 1 = + POPULACAO 2 = O



NTERVALO HORIZONTAL = .101 INTERVALO VERTICAL = .225

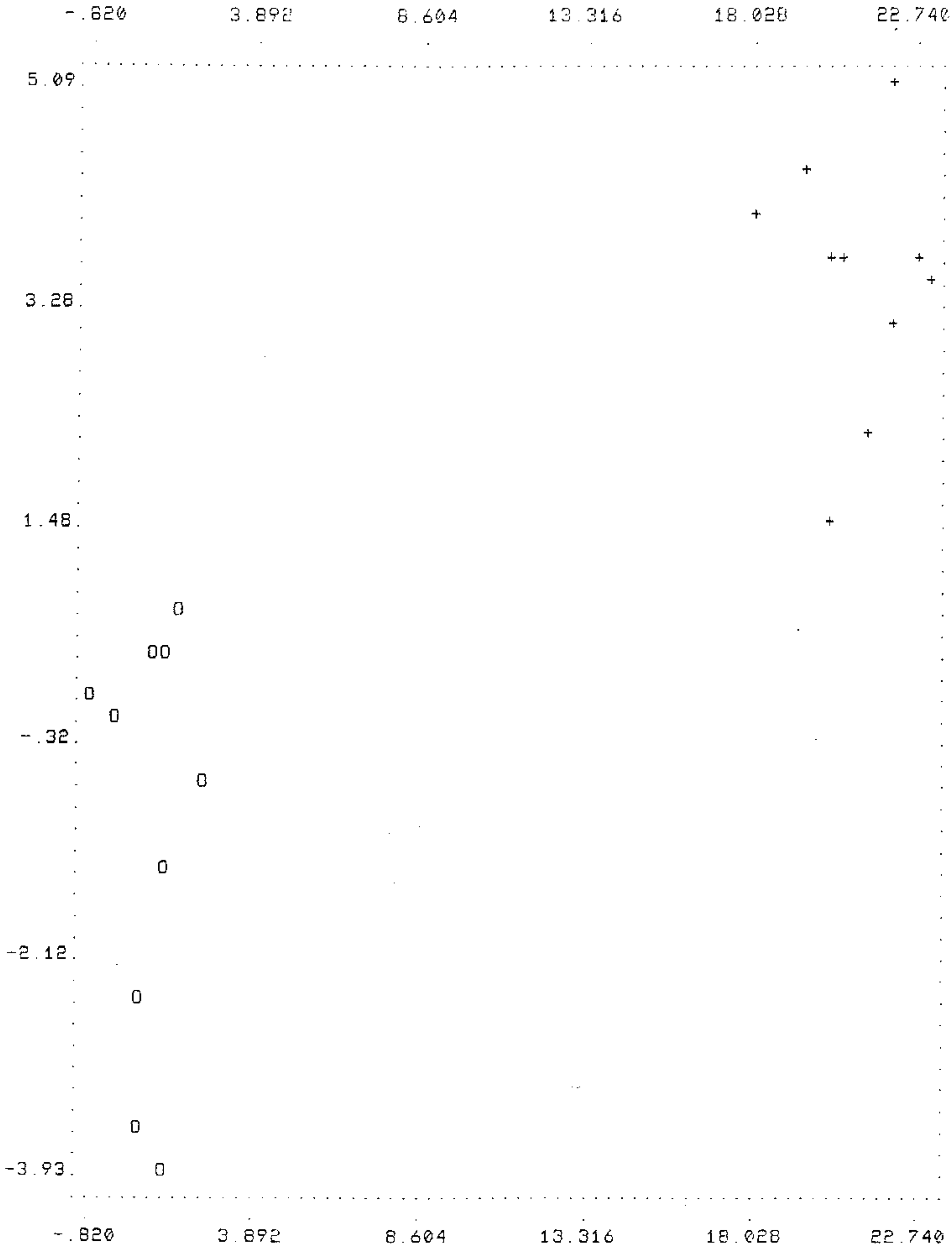
DADOS DO BOI - GORDURA INDUSTRIAL   METODO DAS COVARIANCIAS

POPULACAO 1 = +   POPULACAO 2 = 0



INTERVALO HORIZONTAL = .485E-01   INTERVALO VERTICAL = .187

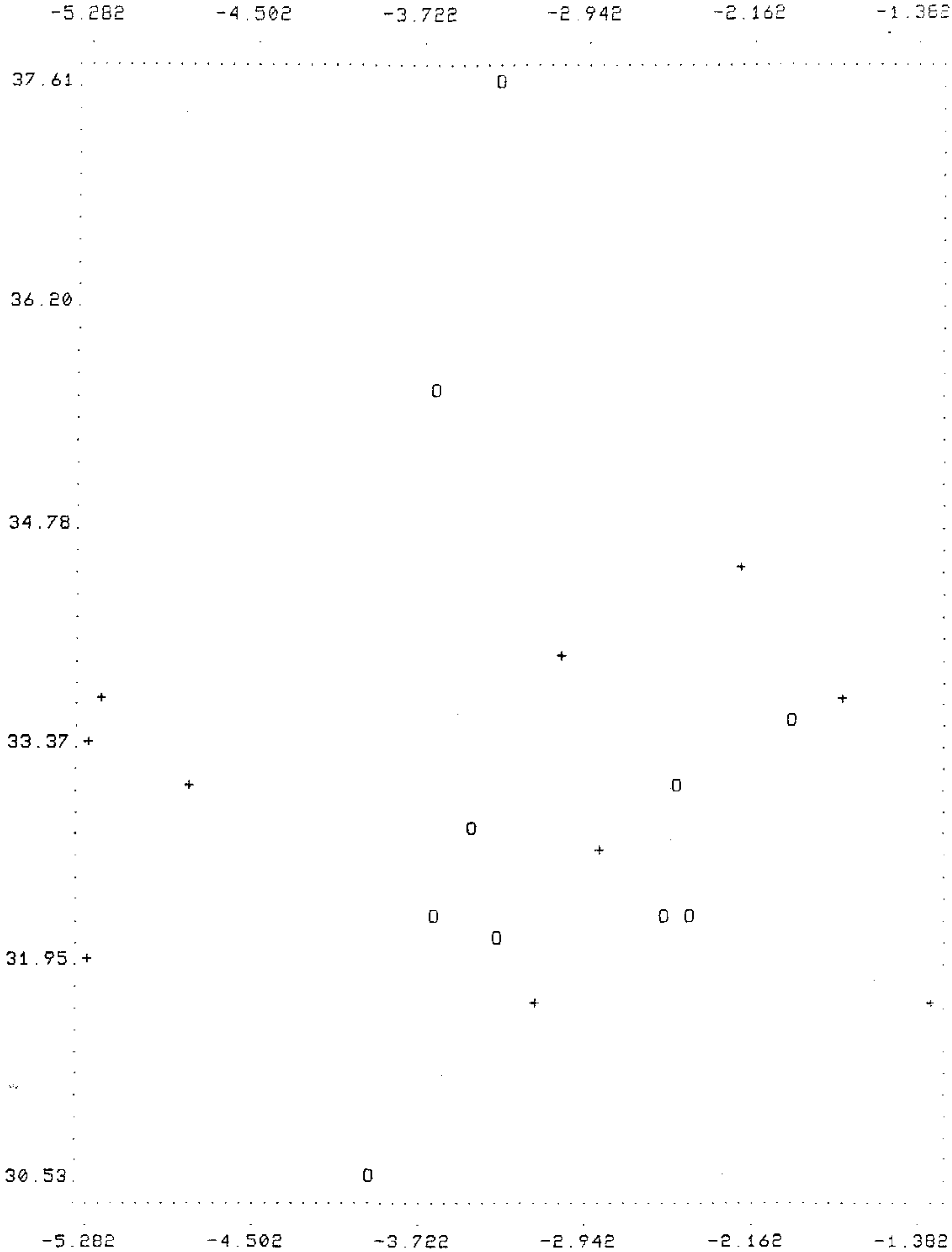
DADOS DO BOI - CARNES APROVEITAVEIS   METODO DAS CORRELACOES
POPULACAO 1 = +   POPULACAO 2 = O



NTERVALO HORIZONTAL = .236   INTERVALO VERTICAL = .180

DADOS DO BOI - CARNES APROVEITAVEIS

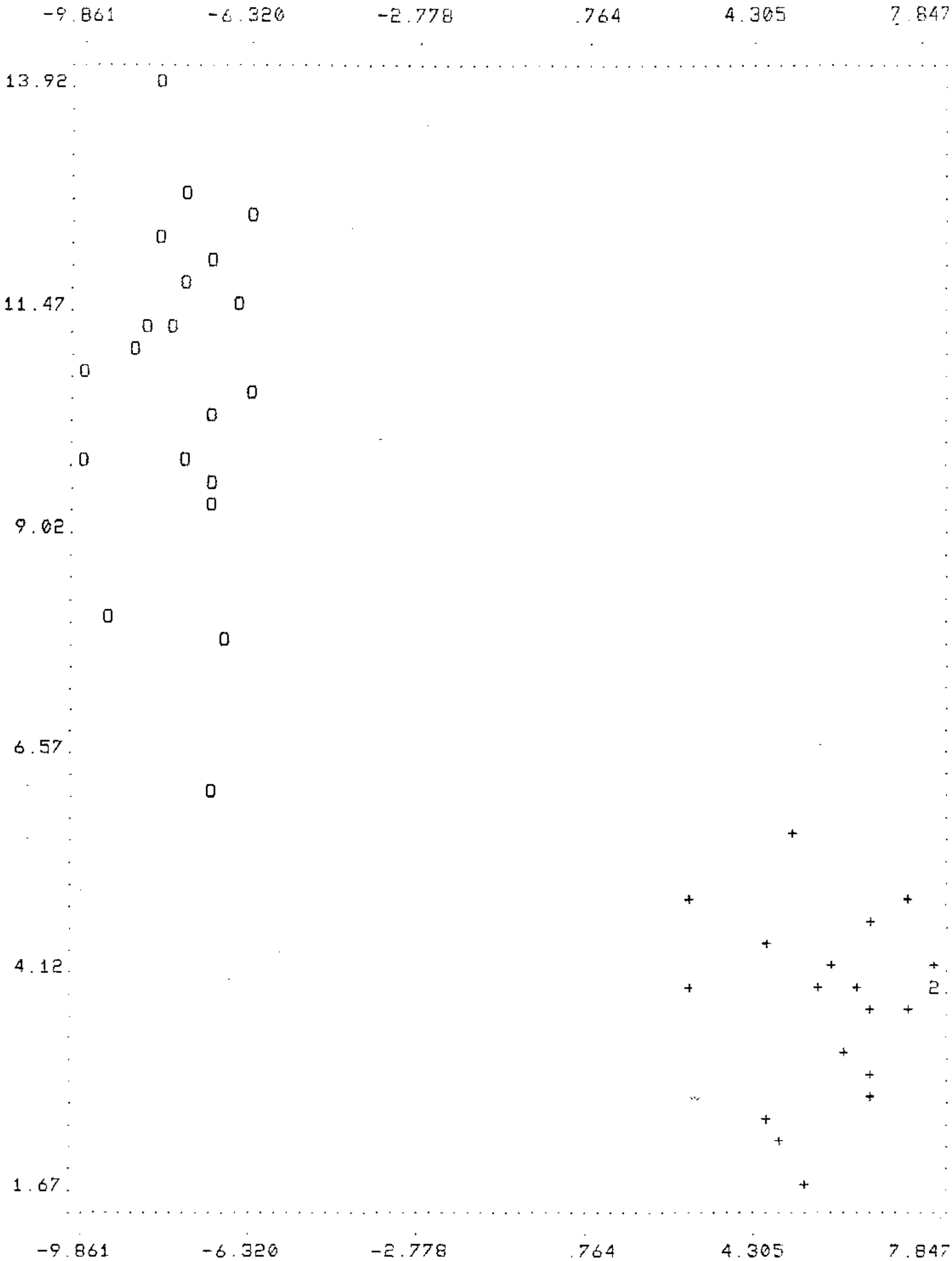
POPULACAO 1 = +   POPULACAO 2 = 0



INTERVALO HORIZONTAL =   .390E-01   INTERVALO VERTICAL =   .142

ISOLADOS METODO DAS CORRELACOES

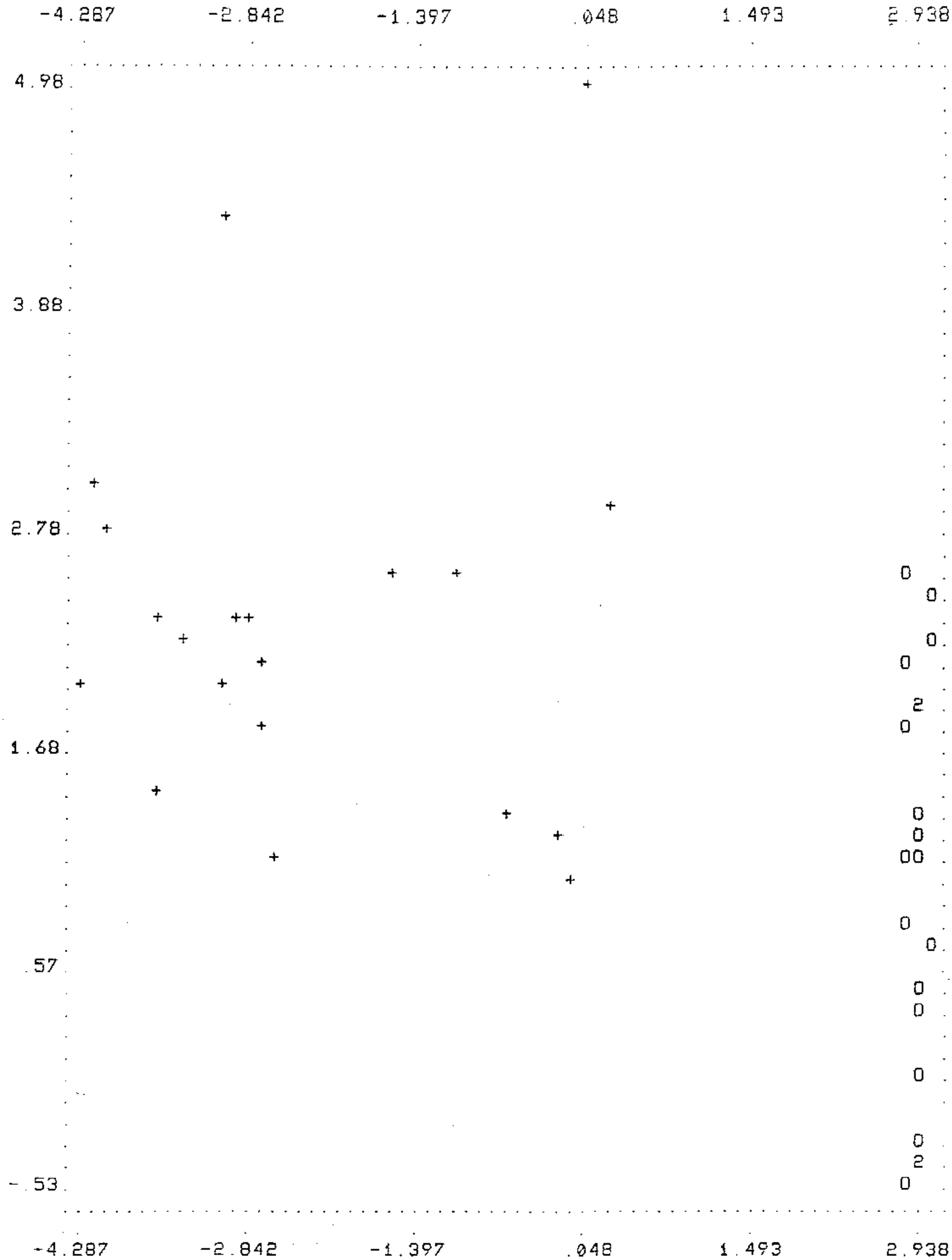
POPULACAO 1 = + POPULACAO 2 = O



NTERVALO HORIZONTAL = .177 INTERVALO VERTICAL = .245

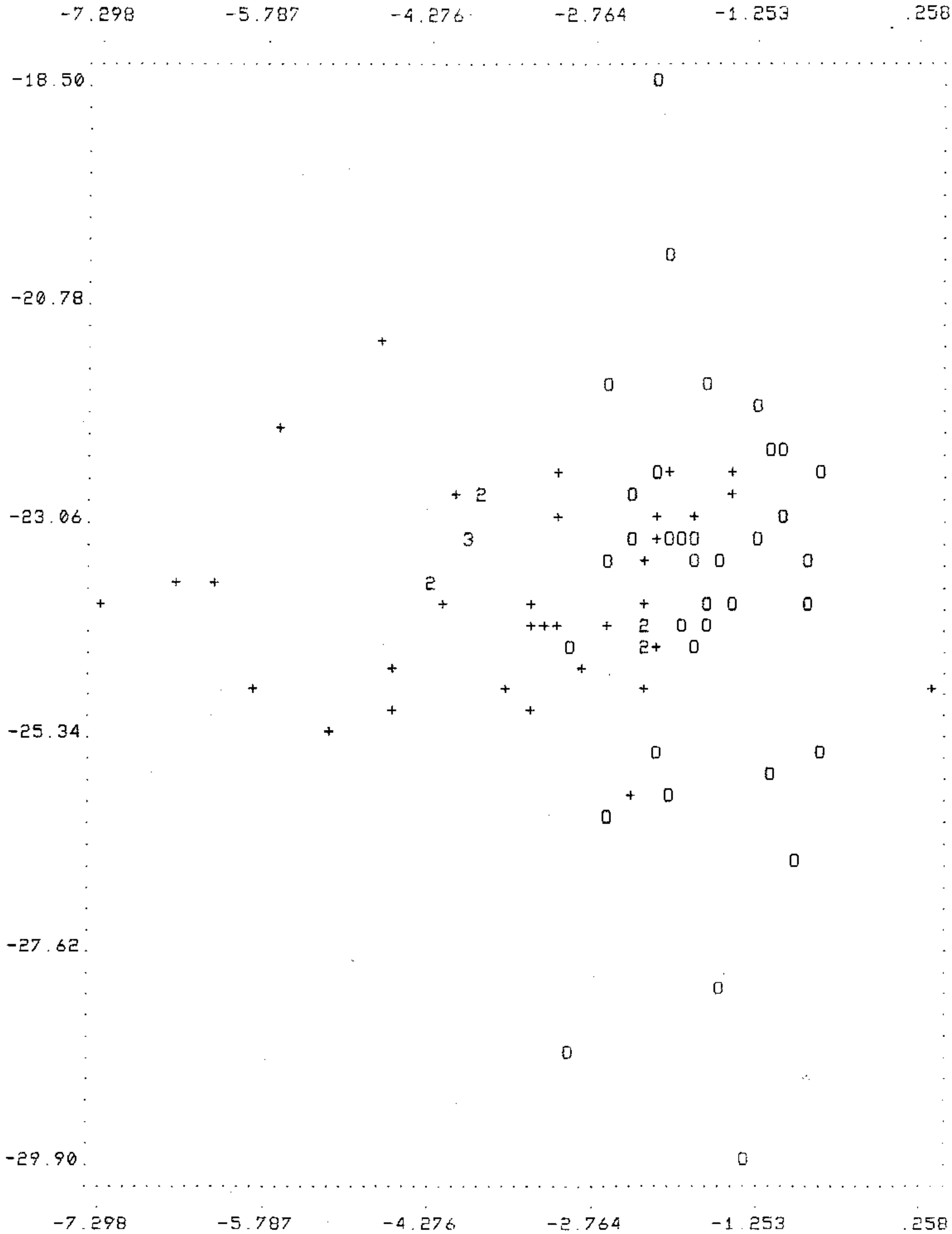
ISOLADOS METODO DAS COVARIANCIAS

POPULACAO 1 = + POPULACAO 2 = 0



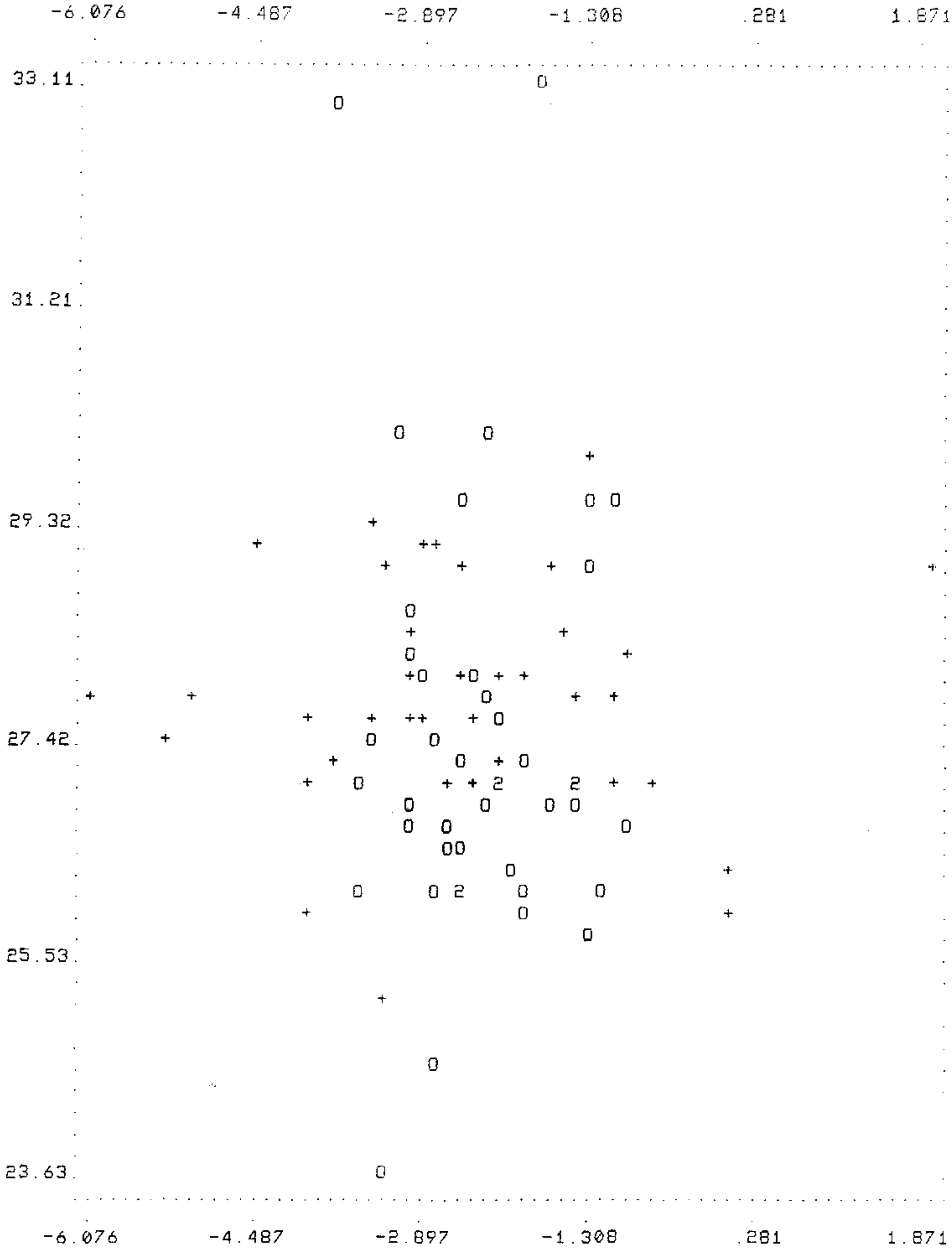
NTERVALO HORIZONTAL = .723E-01 INTERVALO VERTICAL = .110

ASA - LINHAGENS 1 E 2 METODO DAS CORRELACOES
POPULACAO 1 = + POPULACAO 2 = 0



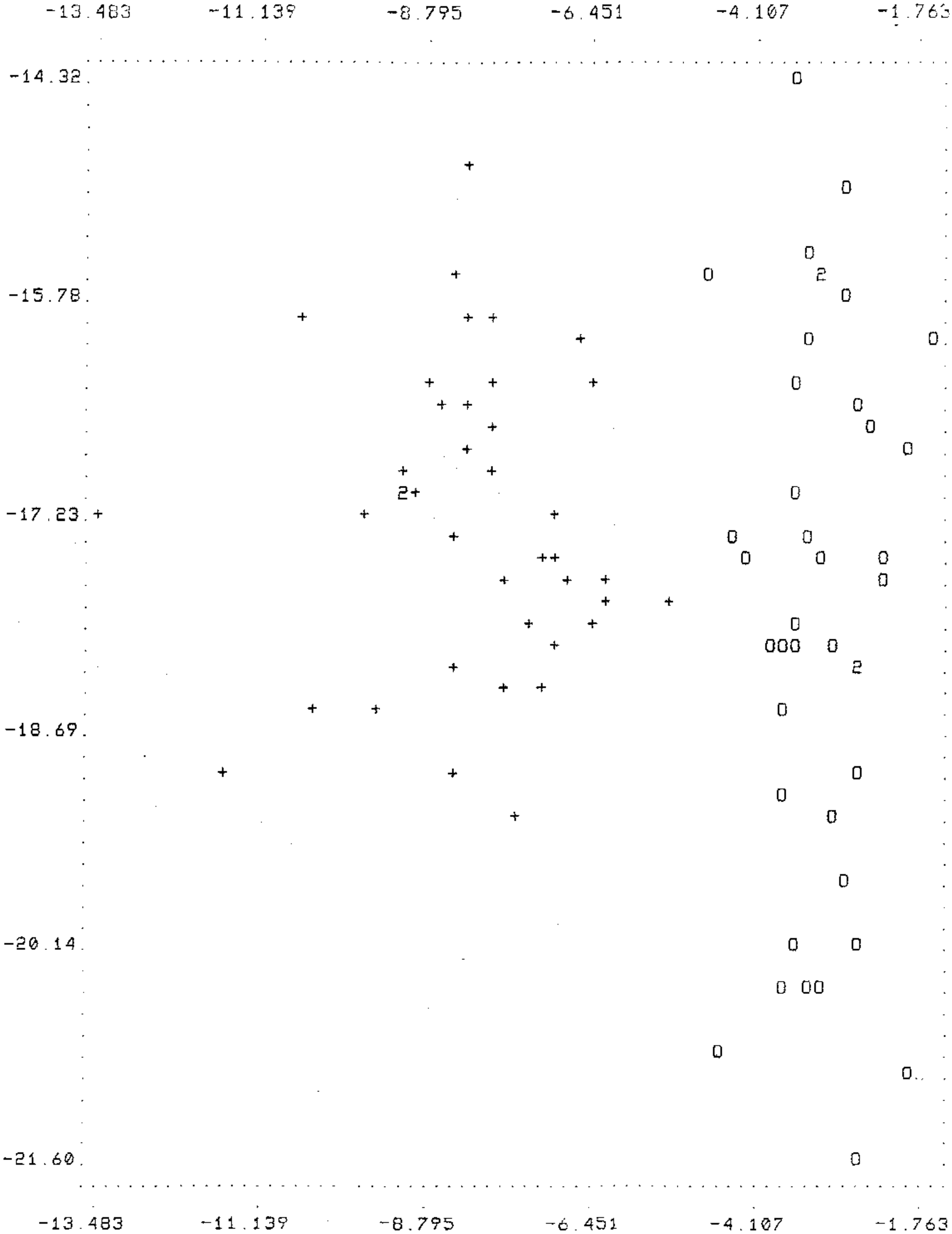
NTERVALO HORIZONTAL = .756E-01 INTERVALO VERTICAL = .228

ASF - LINHAGENS 1 E 2 METODO DAS COVARIANCIAS
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .795E-01 INTERVALO VERTICAL = .190

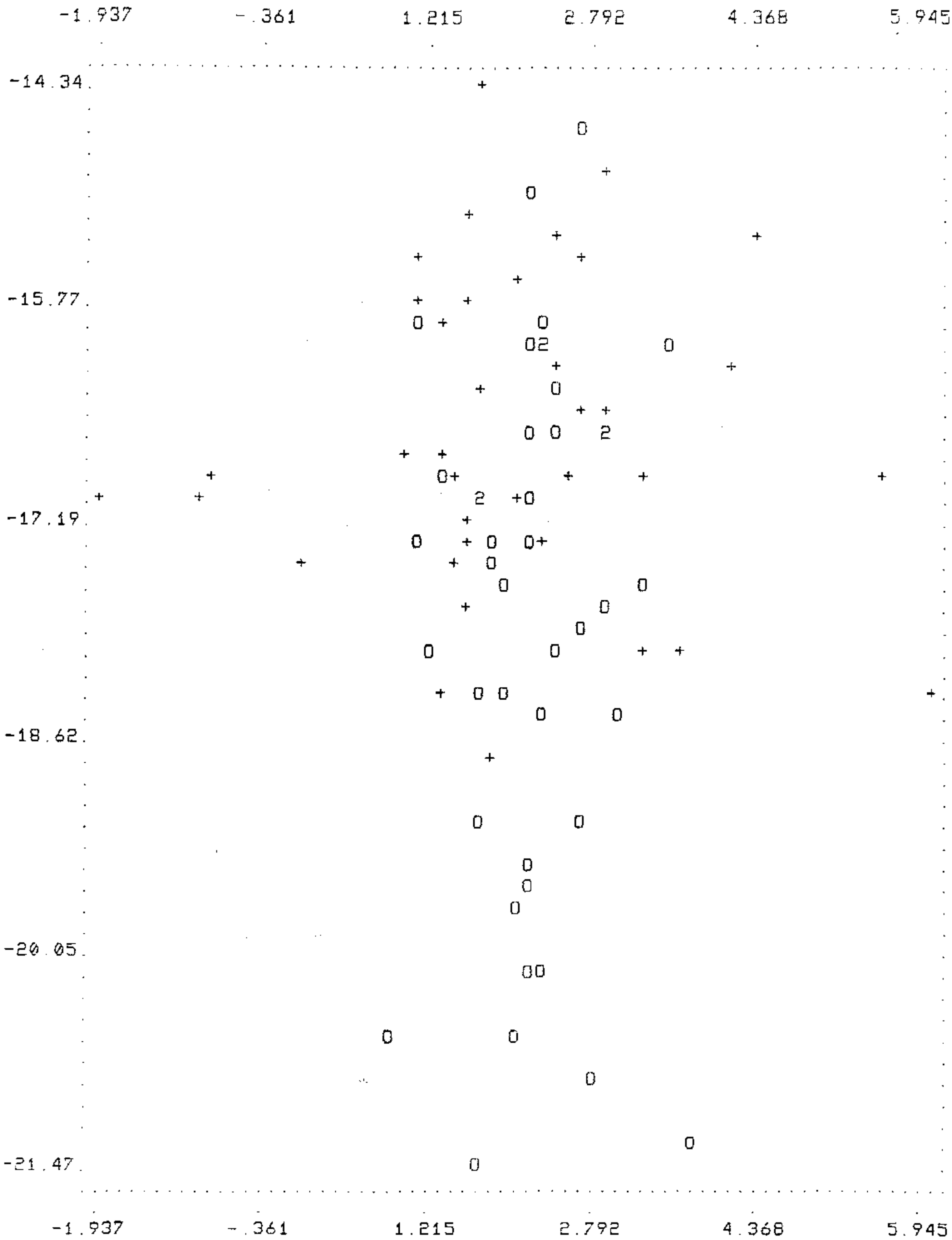
ASA - LINHAGENS 1 E 3 METODO DAS CORRELACOES
POPULACAO 1 = + POPULACAO 2 = 0



NTERVALO HORIZONTAL = .117 INTERVALO VERTICAL = .145

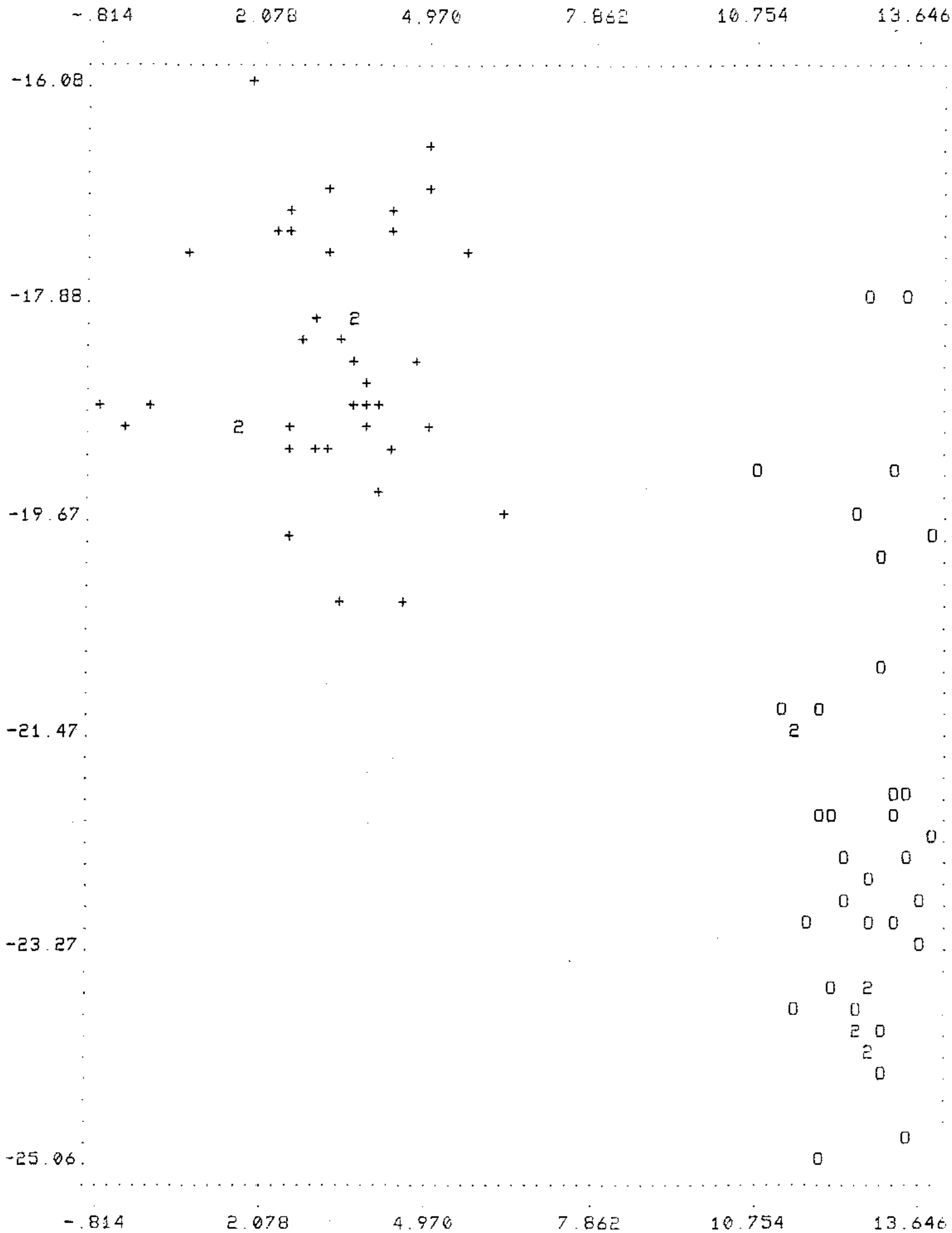
LINHAGENS 1 E 3 METODO DAS

POPULACAO 1 = +   POPULACAO 2 = 0

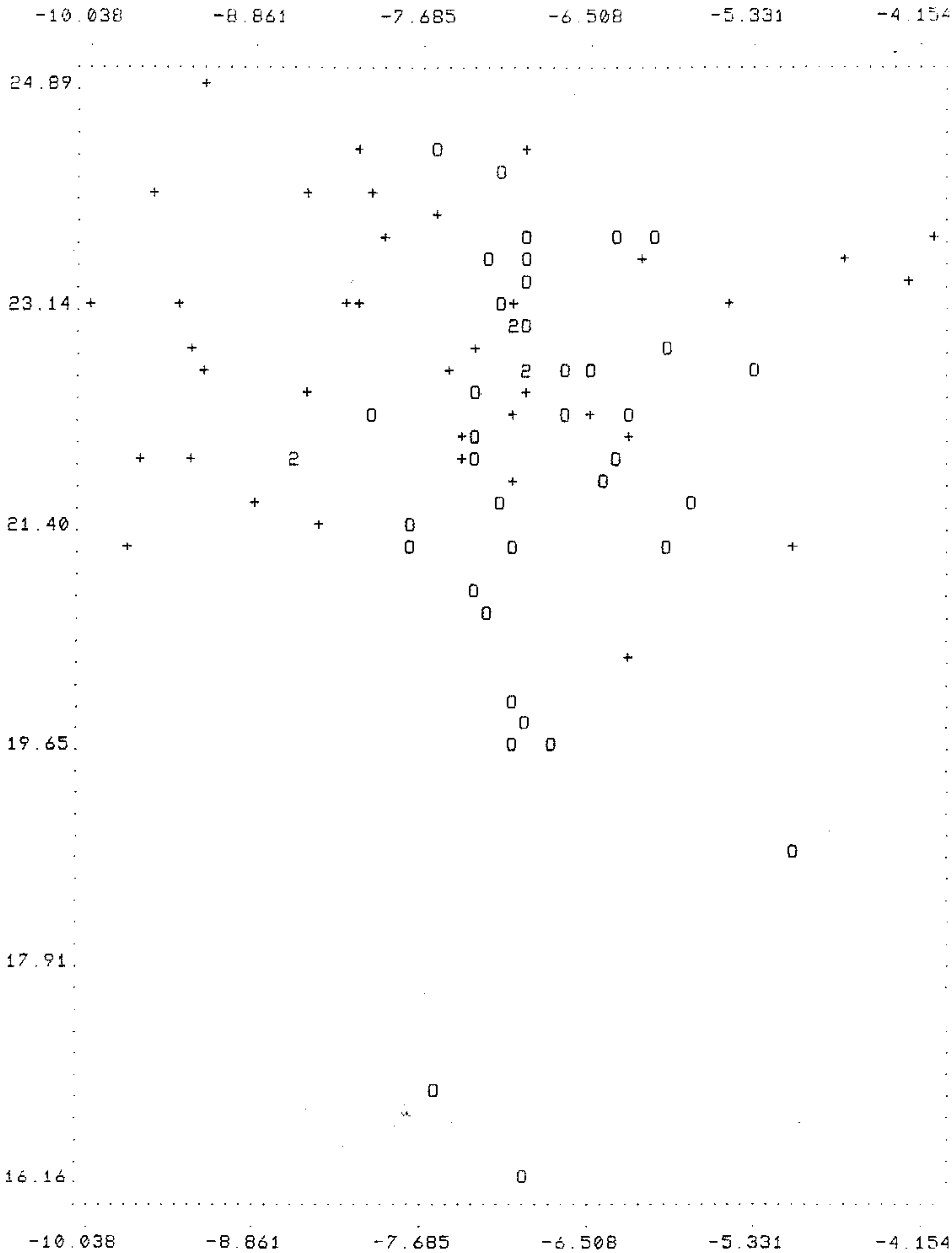


INTERVALO HORIZONTAL = .788E-01   INTERVALO VERTICAL = .143

ASA - LINHAGENS 3 E 4 METODO DAS CORRELACOES
POPULACAO 1 = + POPULACAO 2 = 0



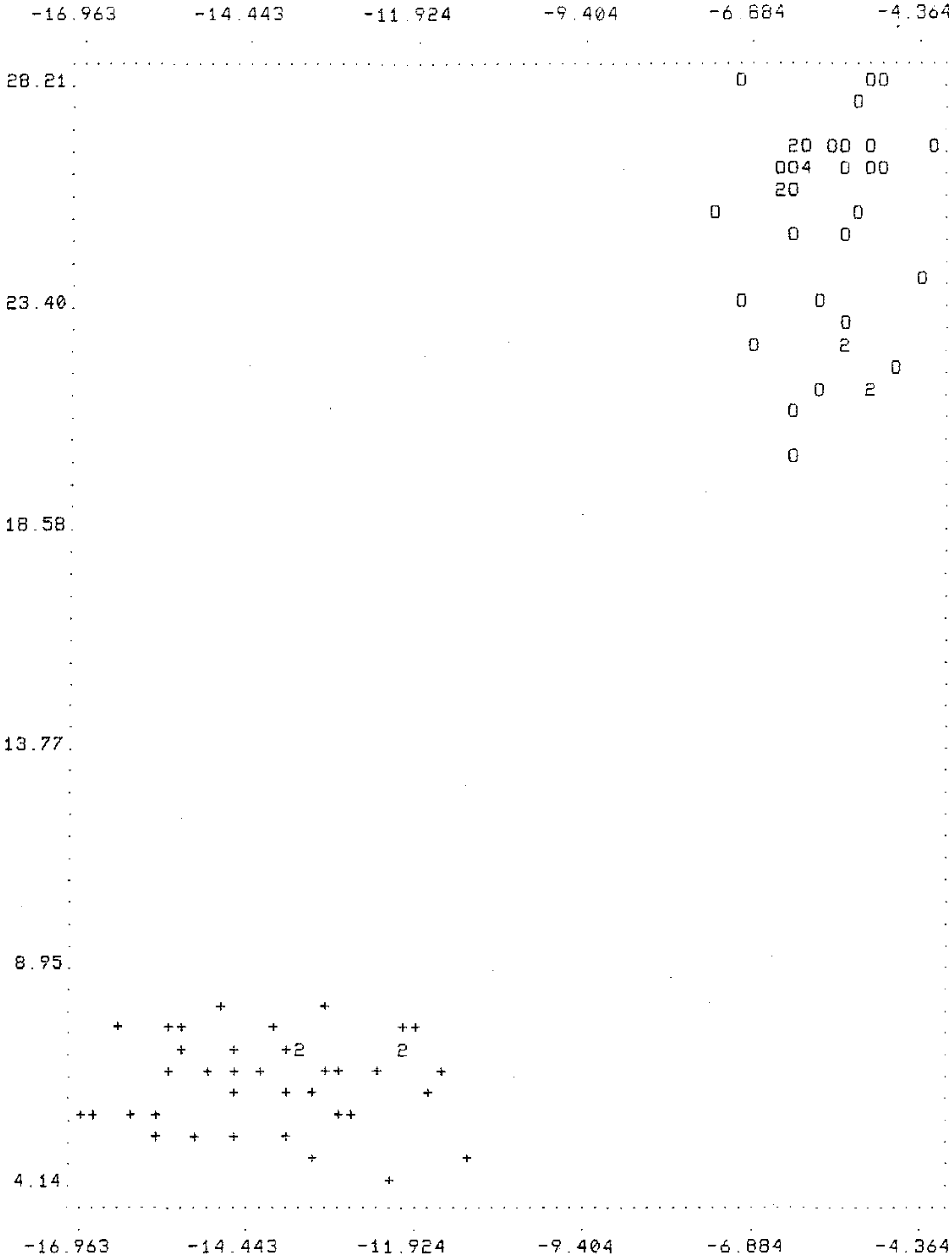
INTERVALO HORIZONTAL = .145 INTERVALO VERTICAL = .180



NTERVALO HORIZONTAL = .588E-01 INTERVALO VERTICAL = 175

ASA - LINHAGENS 2 E 3 METODO DAS CORRELACOES

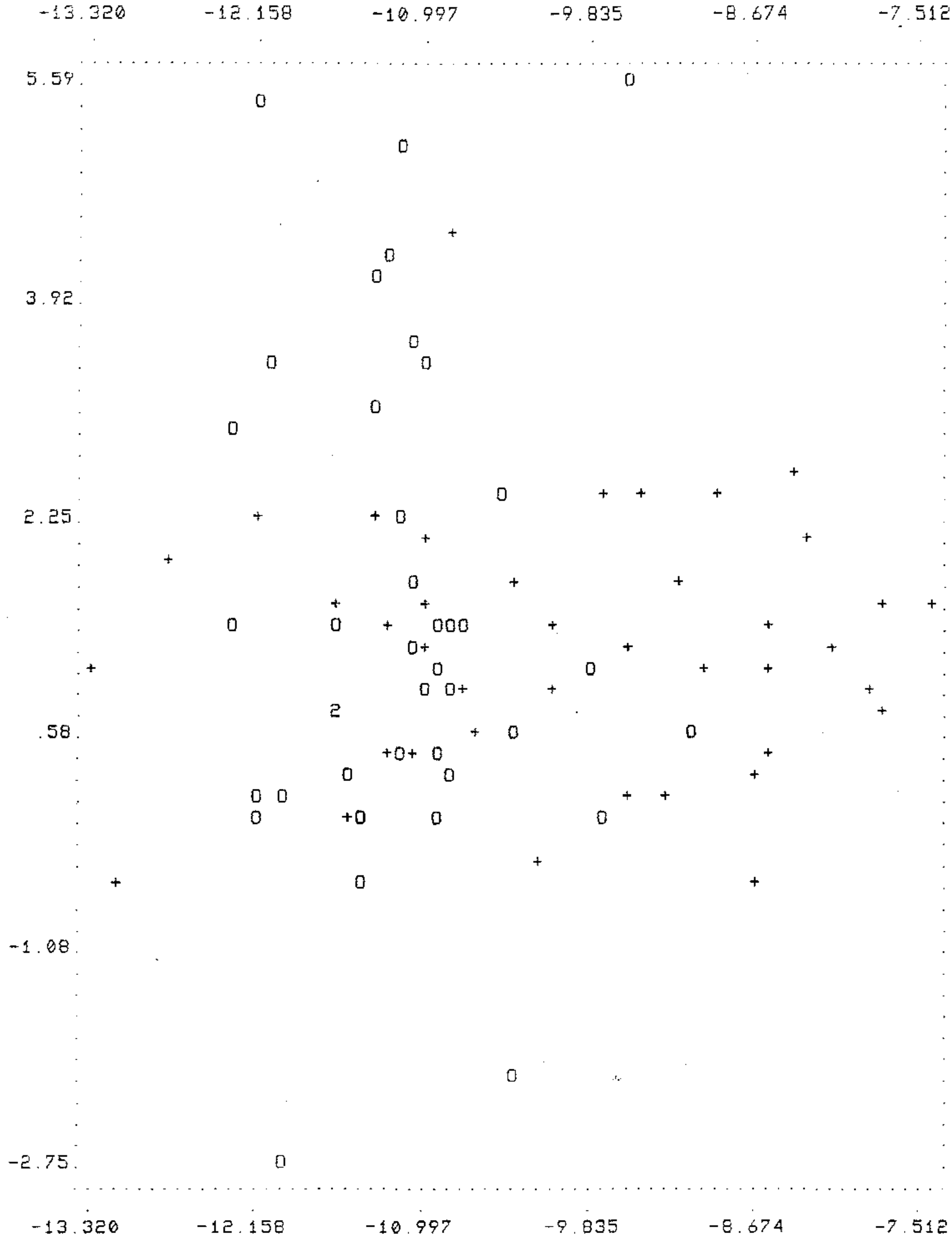
POPULACAO 1 = + POPULACAO 2 = 0



NTERVALO HORIZONTAL = .126 INTERVALO VERTICAL = .482

MSA - LINHAGENS 2 E 3 METODO DAS COVARIANCIAS

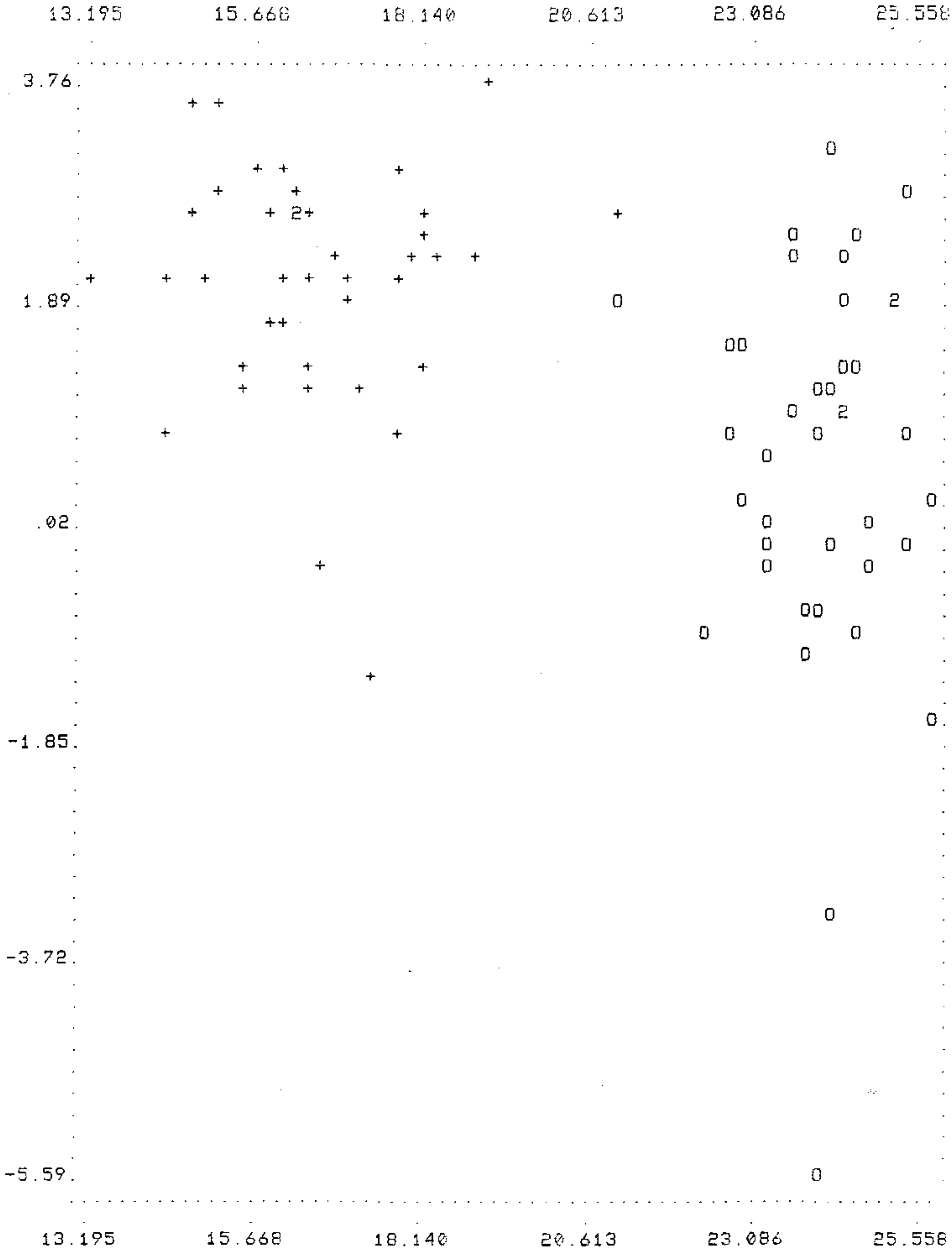
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .581E-01 INTERVALO VERTICAL = .167

LINHAGENS 2 E 4 DAS CORRELACOES

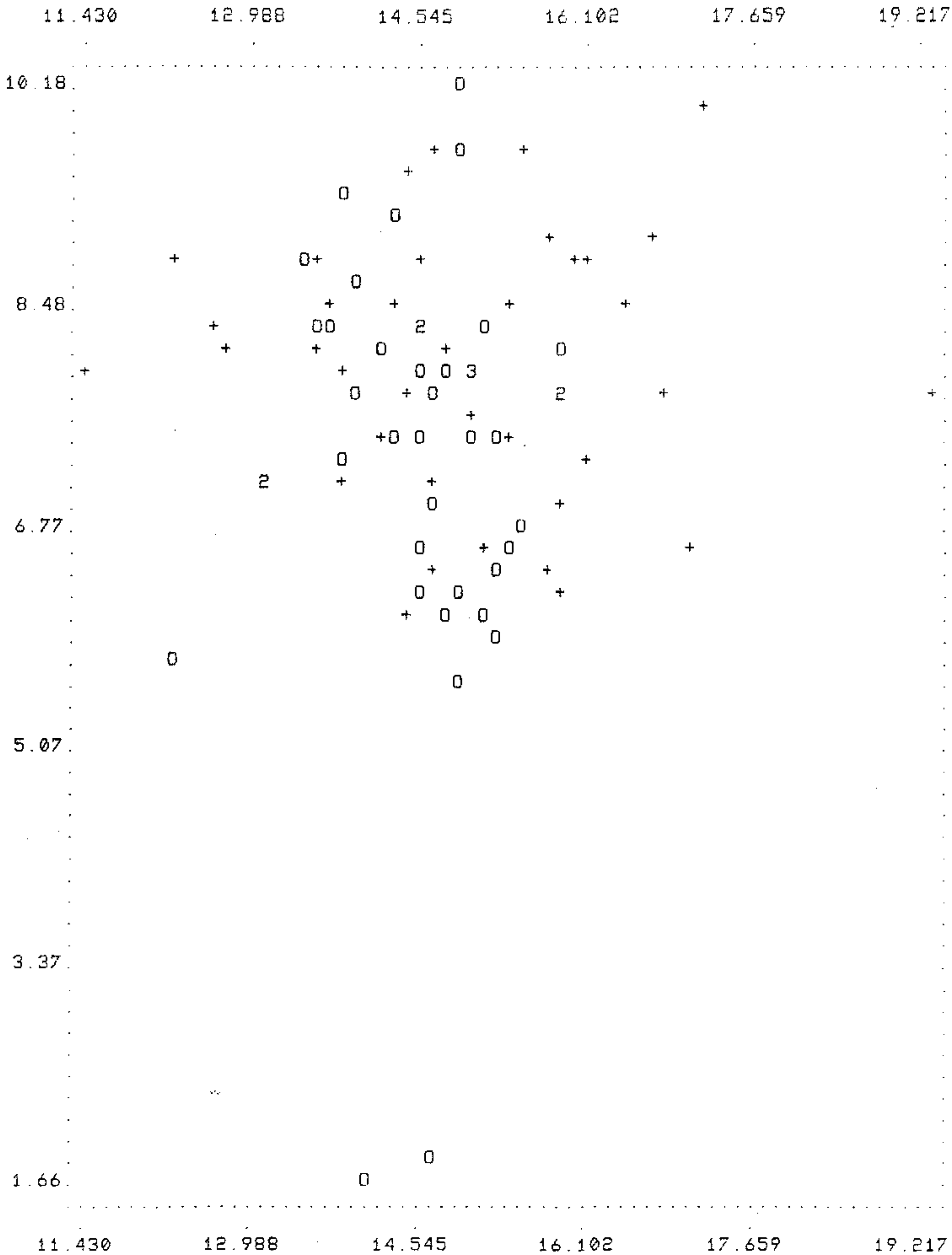
POPULACAO 1 = + POPULACAO 2 = 0



NTERVALO HORIZONTAL = .124 INTERVALO VERTICAL = .187

- LINHAGENS 2 E 4 METODO DAS COVARIANCIAS

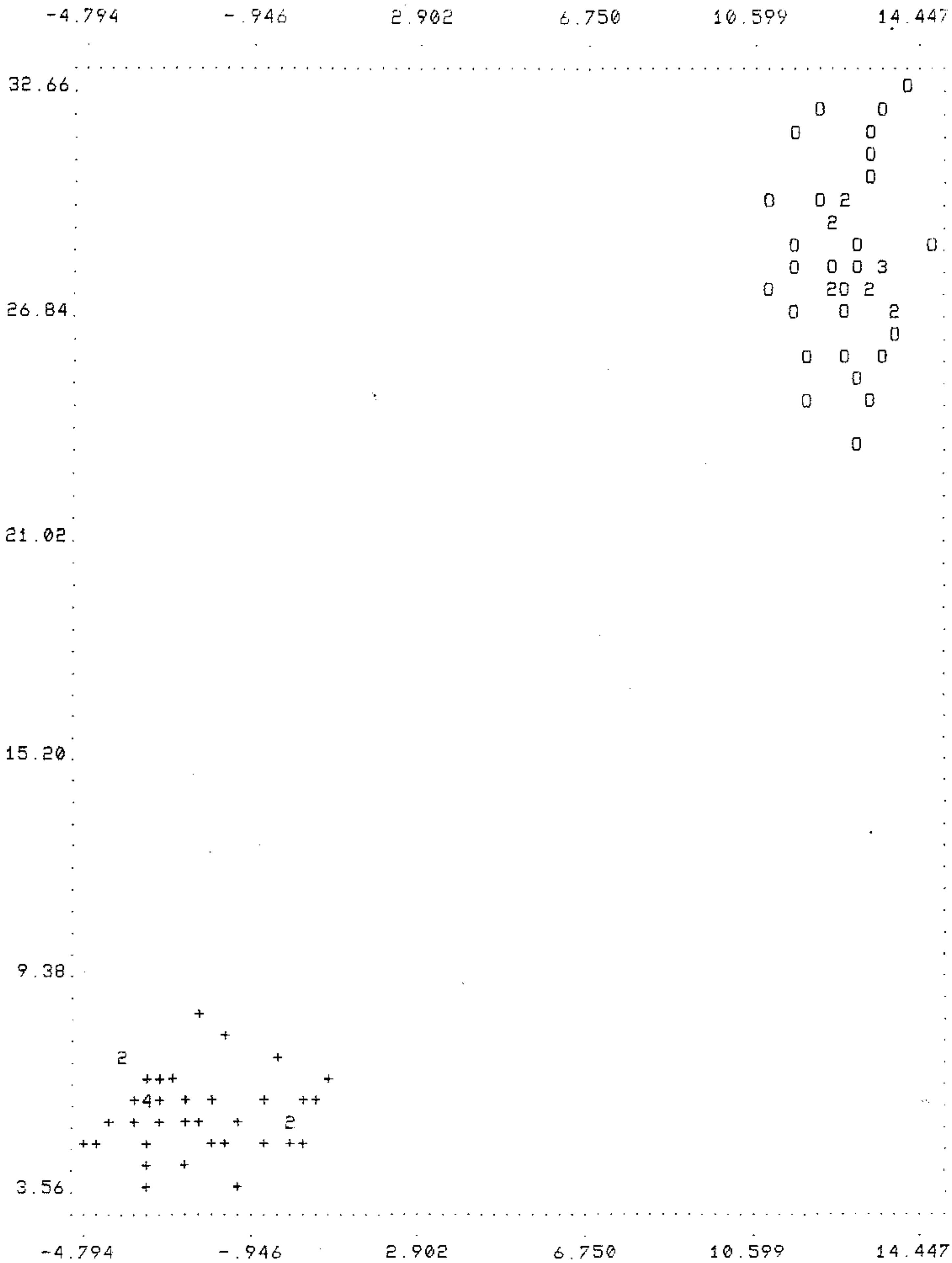
POPULACAO 1 = +   POPULACAO 2 = 0



NTERVALO HORIZONTAL = .779E-01   INTERVALO VERTICAL = .170

45A - LINHAGEM 3 E 4 METODO DAS CORRELACOES

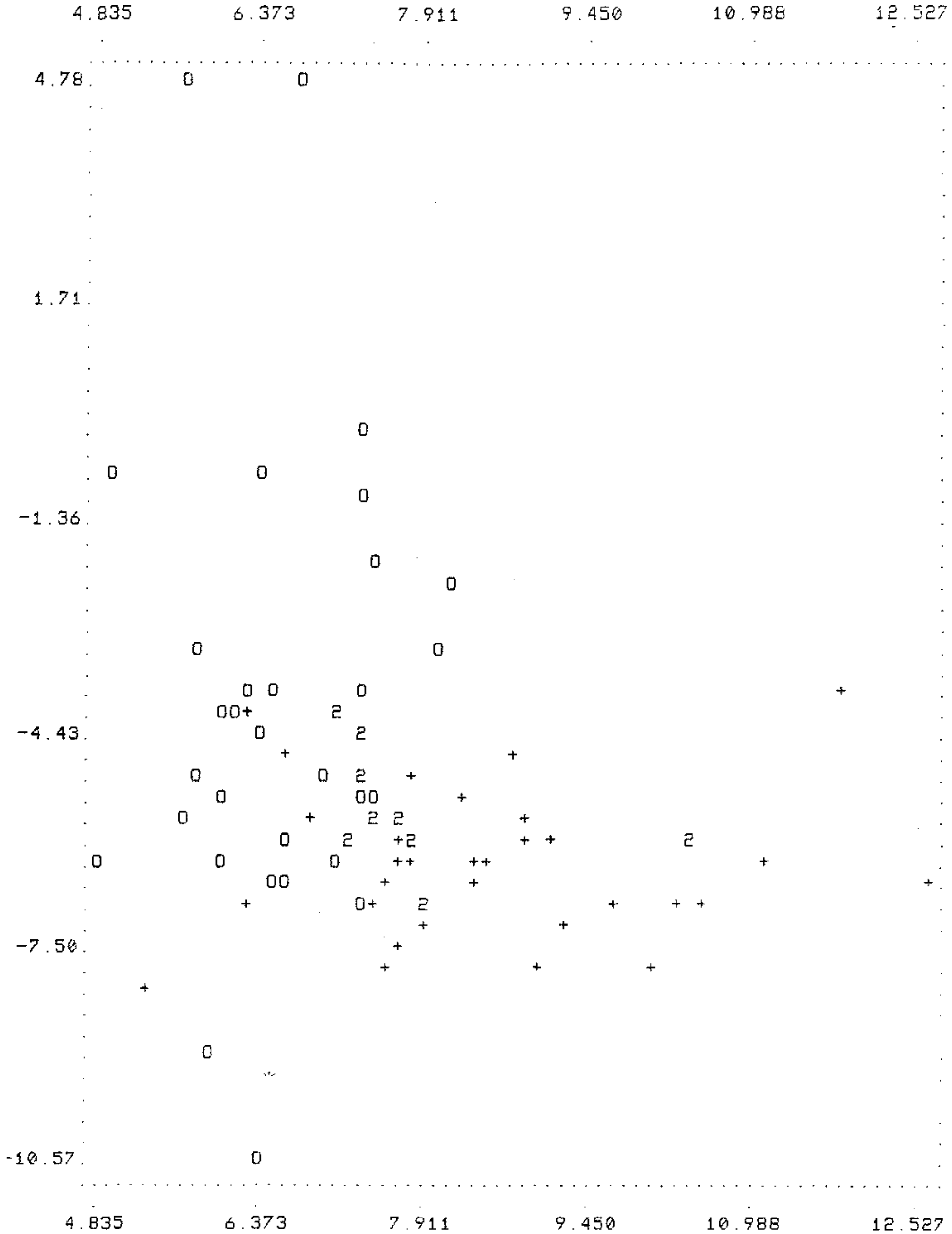
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .192 INTERVALO VERTICAL = .582

ASA - LINHAGENS 3 E 4 METODO DAS COVARIANCIAS

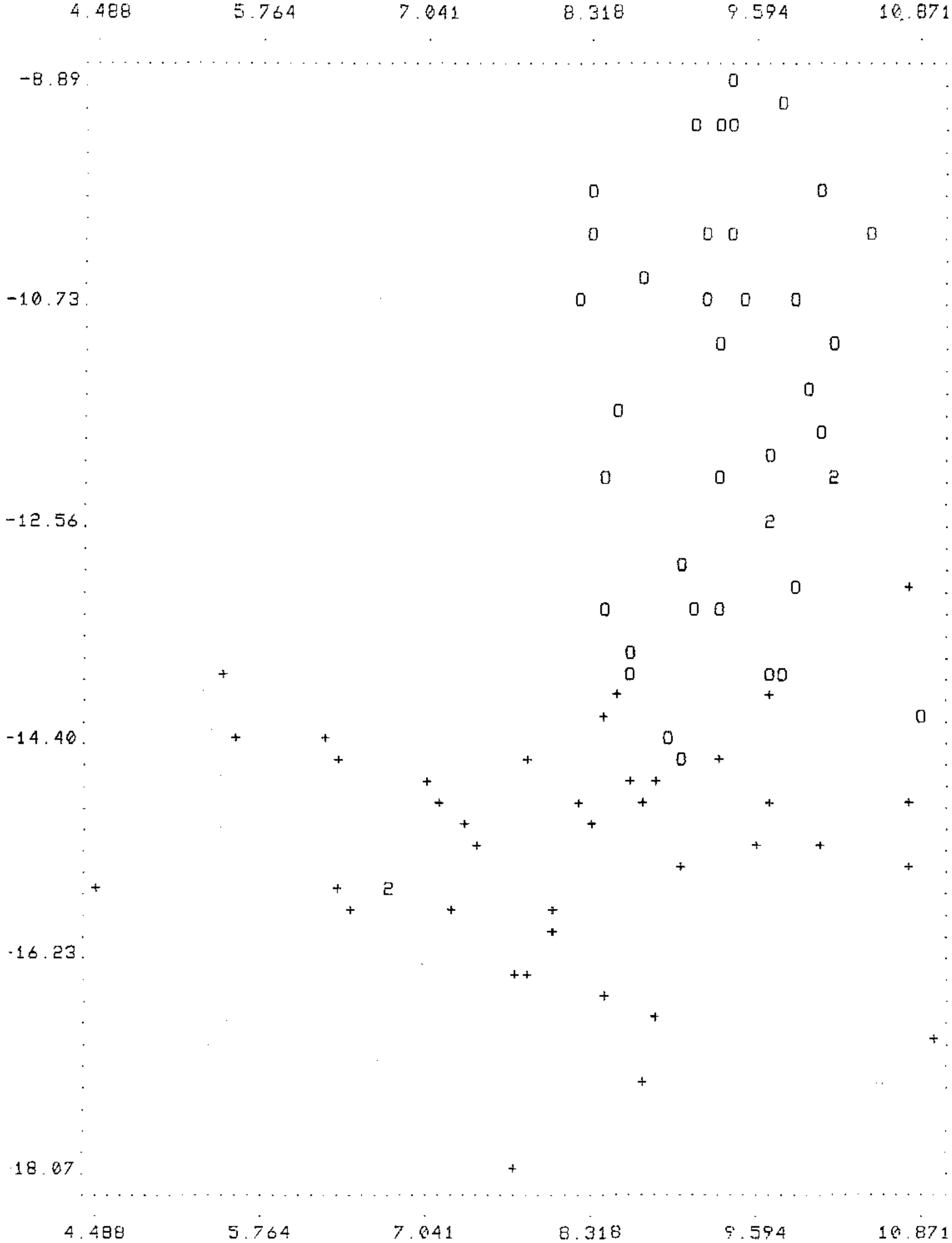
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .769E-01 INTERVALO VERTICAL = .307

CONCOF - LINHAGENS 1 E 2 METODO DAS CORRELACOES

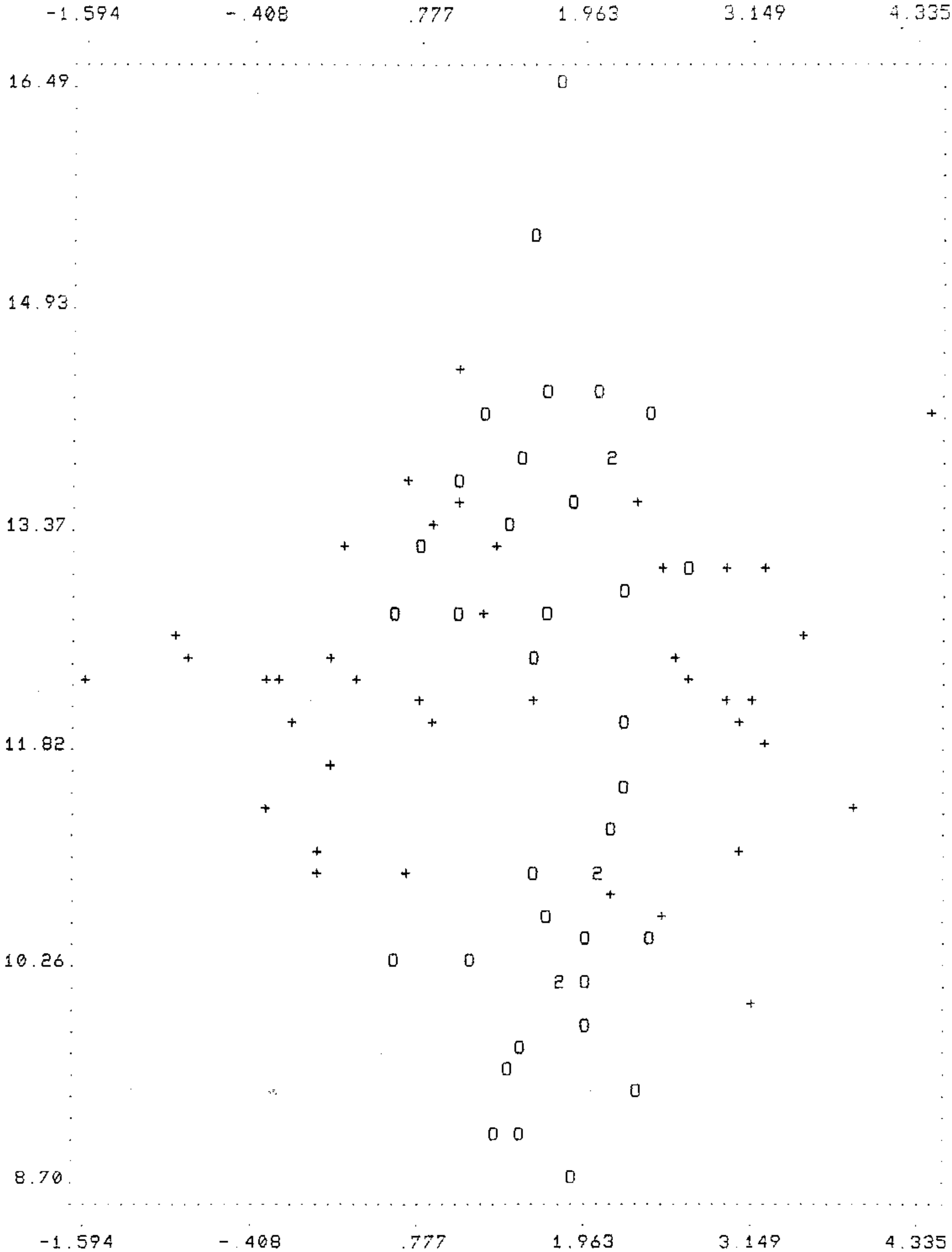
POPULACAO 1 = + POPULACAO 2 = 0



TERVALO HORIZONTAL = .638E-01 INTERVALO VERTICAL = .184

CONCOR - LINHAGENS 1 E 2

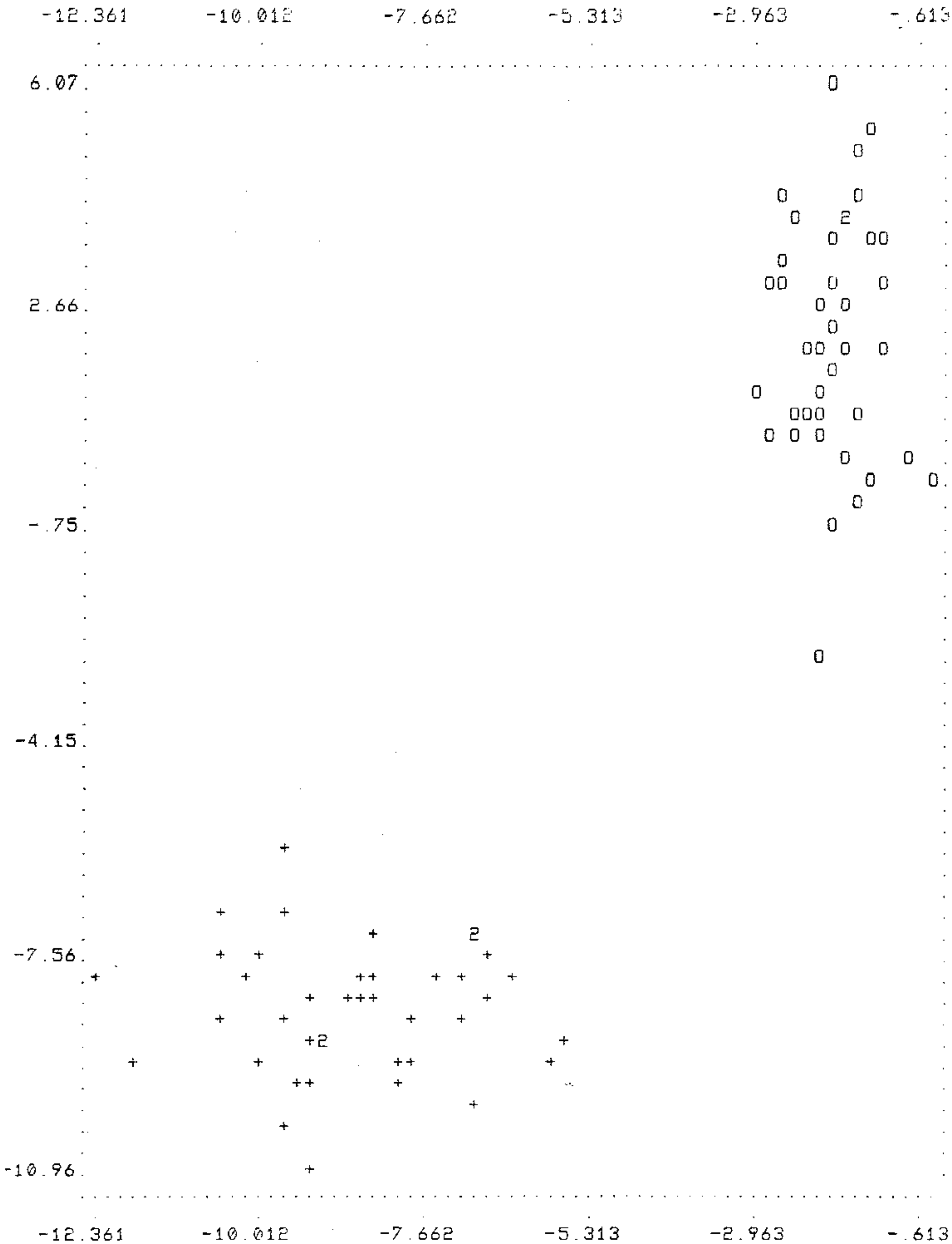
POPULACAO 1 = +   POPULACAO 2 = O



INTERVALO HORIZONTAL = .593E-01   INTERVALO VERTICAL = .156

LINHAGENS 1 E 2 METODO DAS CORRELACOES

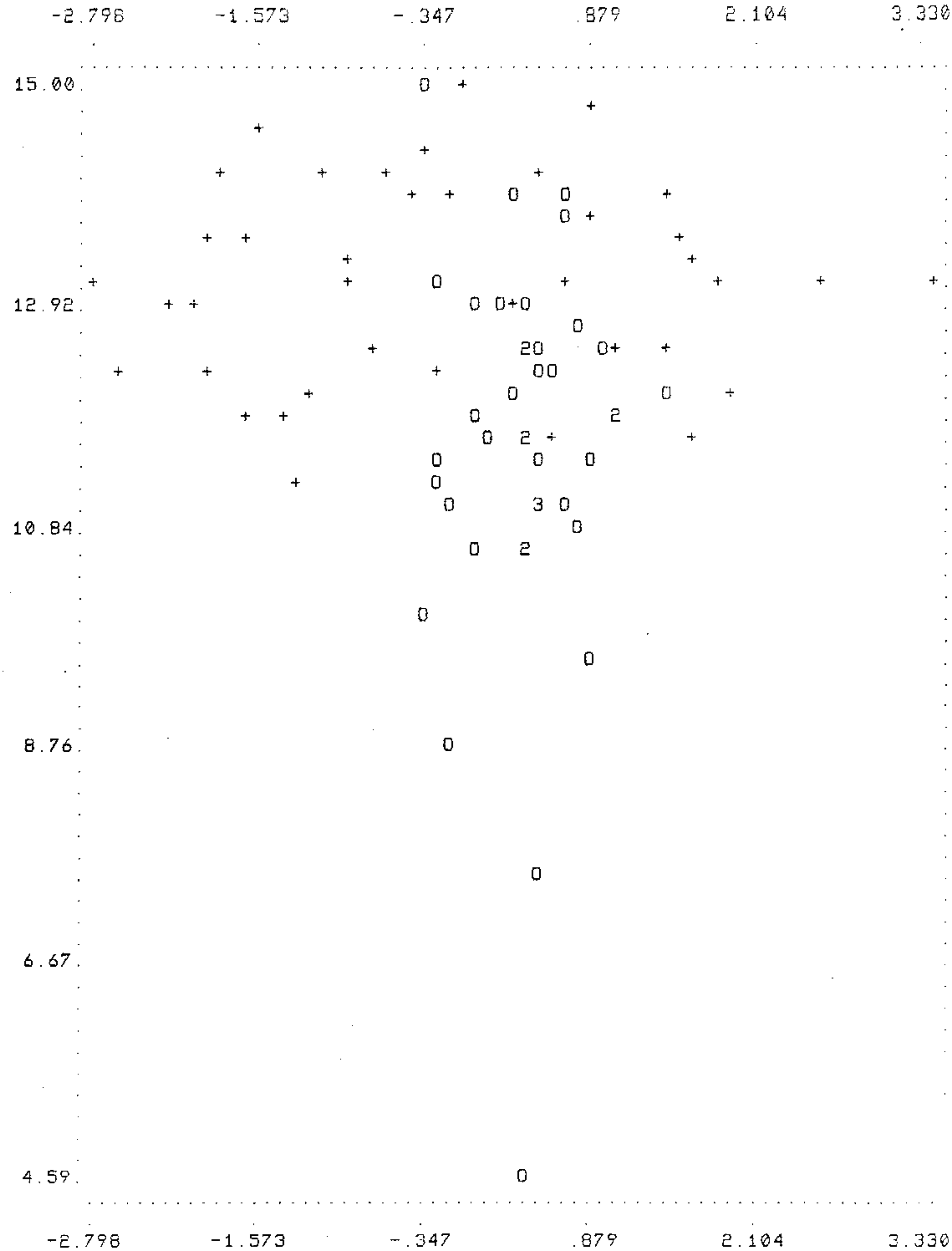
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .117 INTERVALO VERTICAL = .341

CONCOR - LINHAGENS 1 E 3 METODO DAS COVARIANCIAS

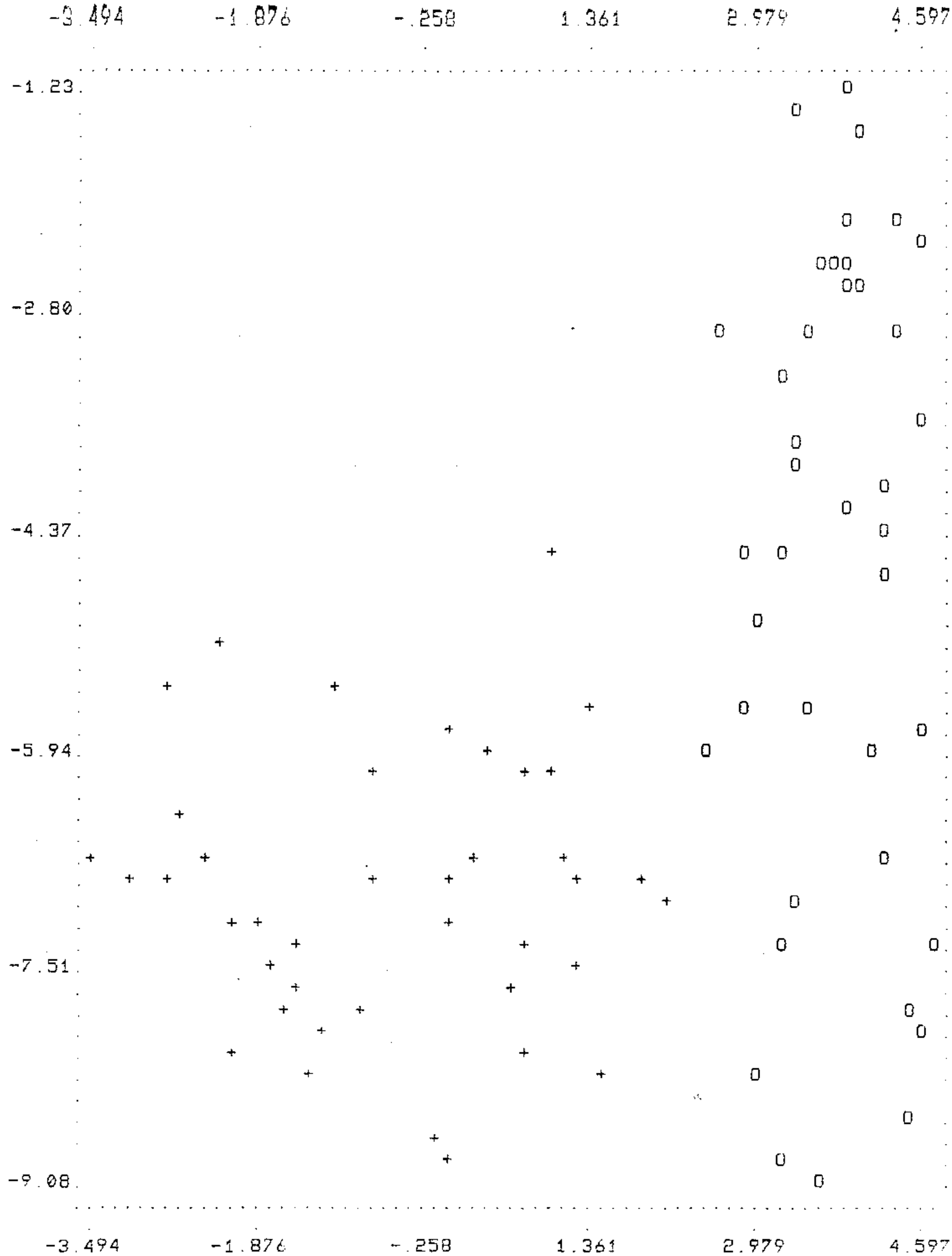
POPULACAO 1 = + POPULACAO 2 = 0



NTERVALO HORIZONTAL = .613E-01 INTERVALO VERTICAL = .208

[illegible] - LINHAGENS 1 E 4 METODO DAS [illegible]

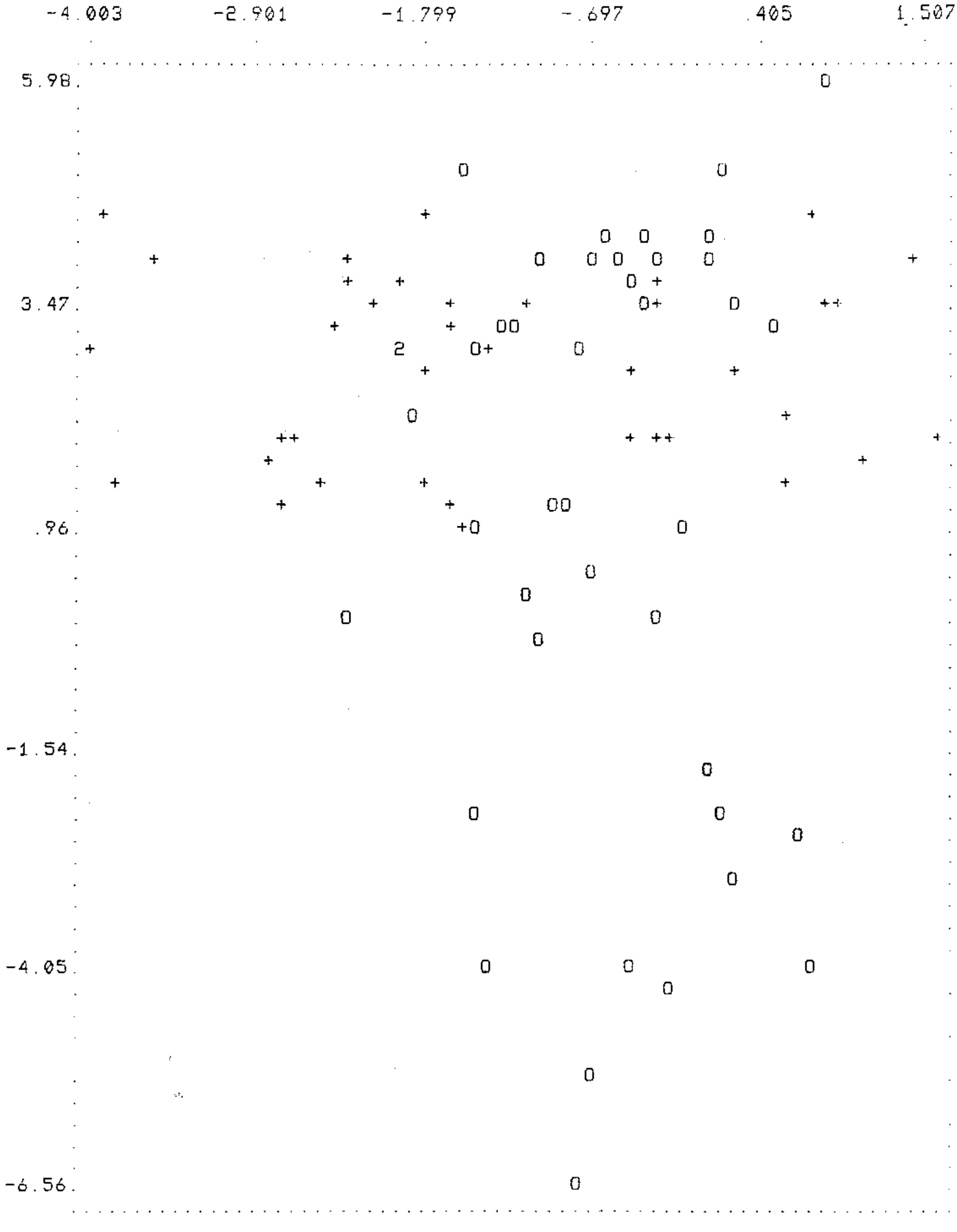
POPULACAO 1 = +   POPULACAO 2 = 0



NTERVALO HORIZONTAL = .809E-01   INTERVALO VERTICAL = .157

CONCOR - LINHAGENS 1 E 4 METODO DAS COVARIANCIAS

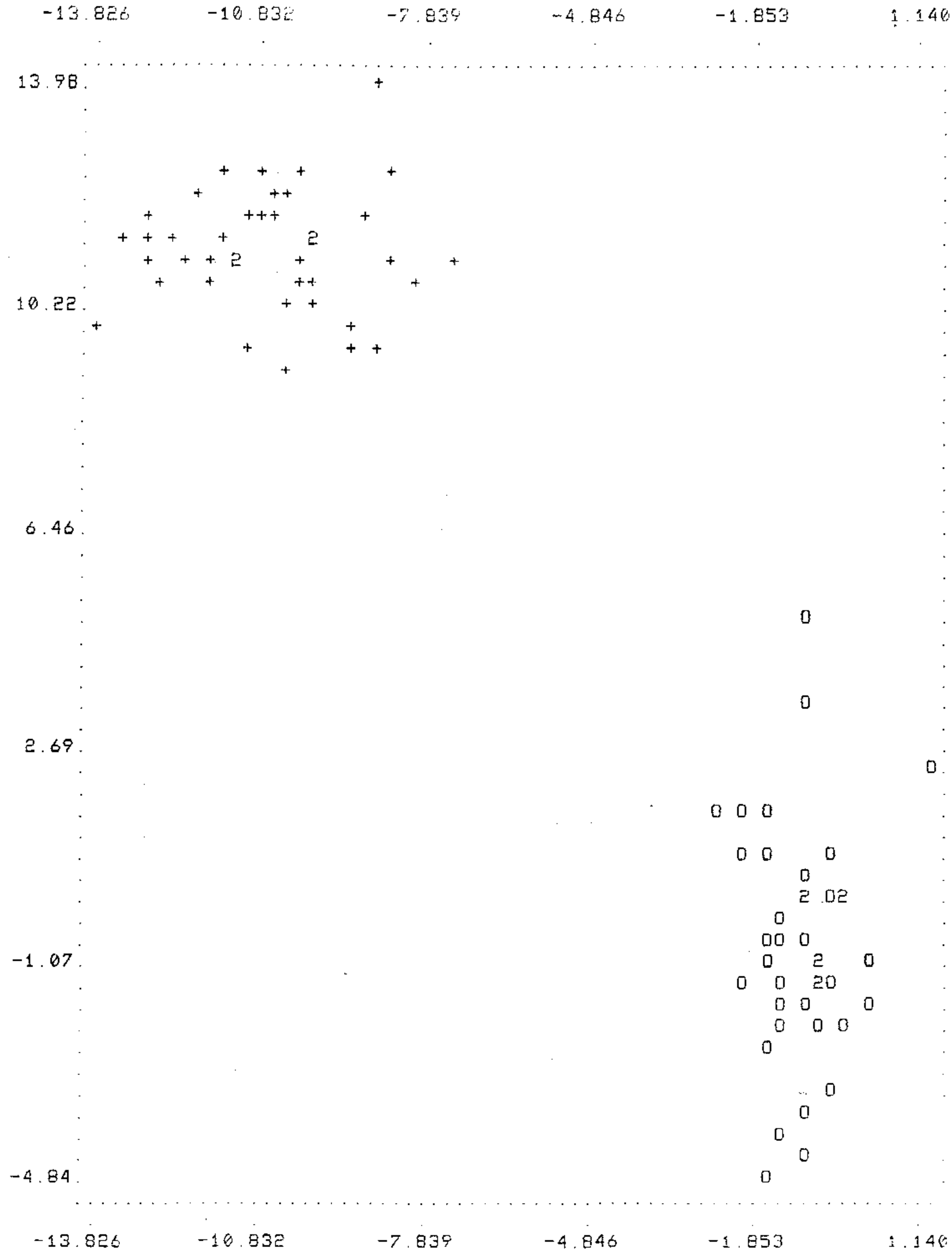
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .551E-01 INTERVALO VERTICAL = .251

- LINHAGENS METODO DAS CORRELACOES

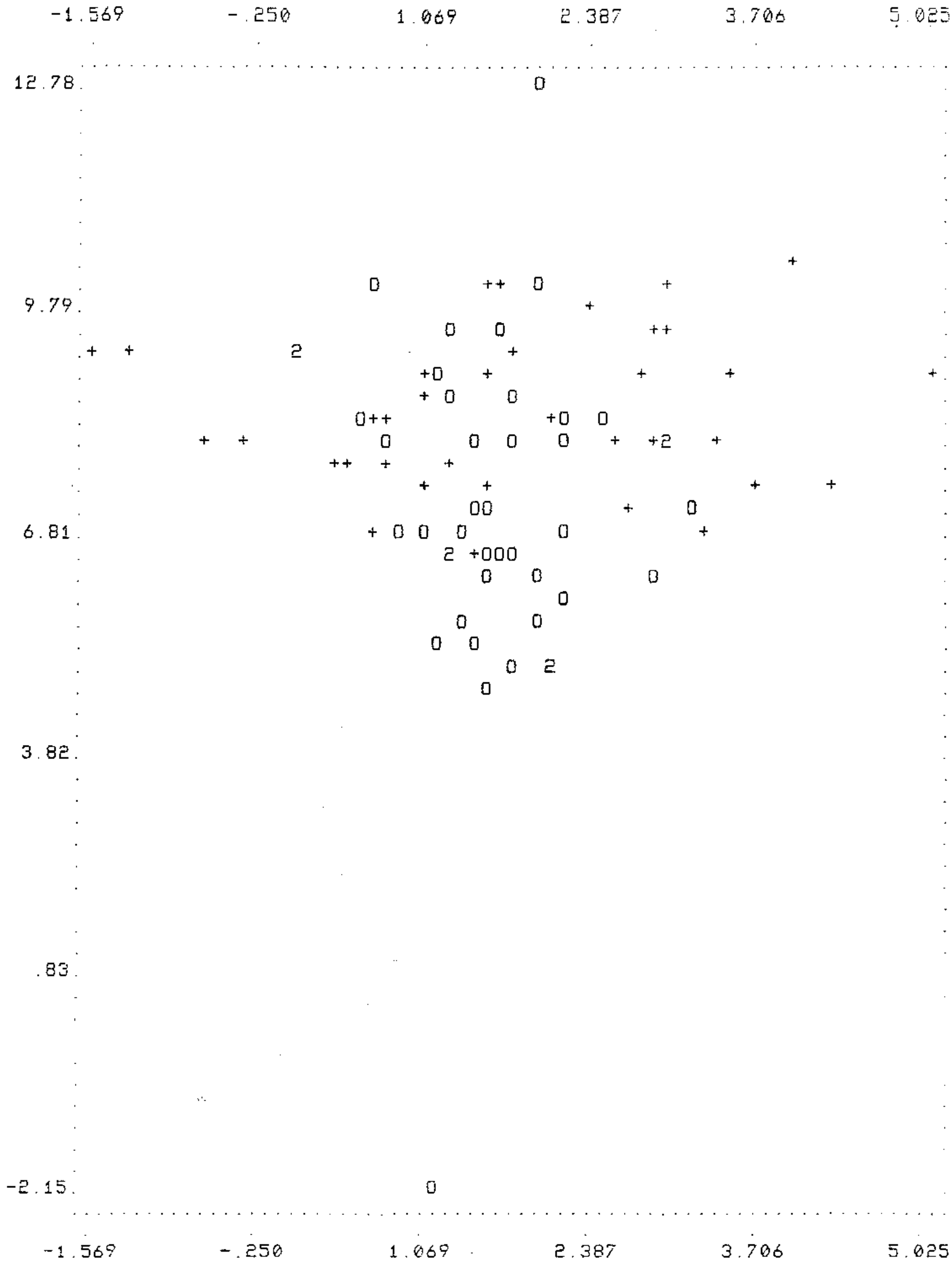
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .150 INTERVALO VERTICAL = .376

CONCOR - LINHAGENS 2 E 3 METODO DAS [illegible]

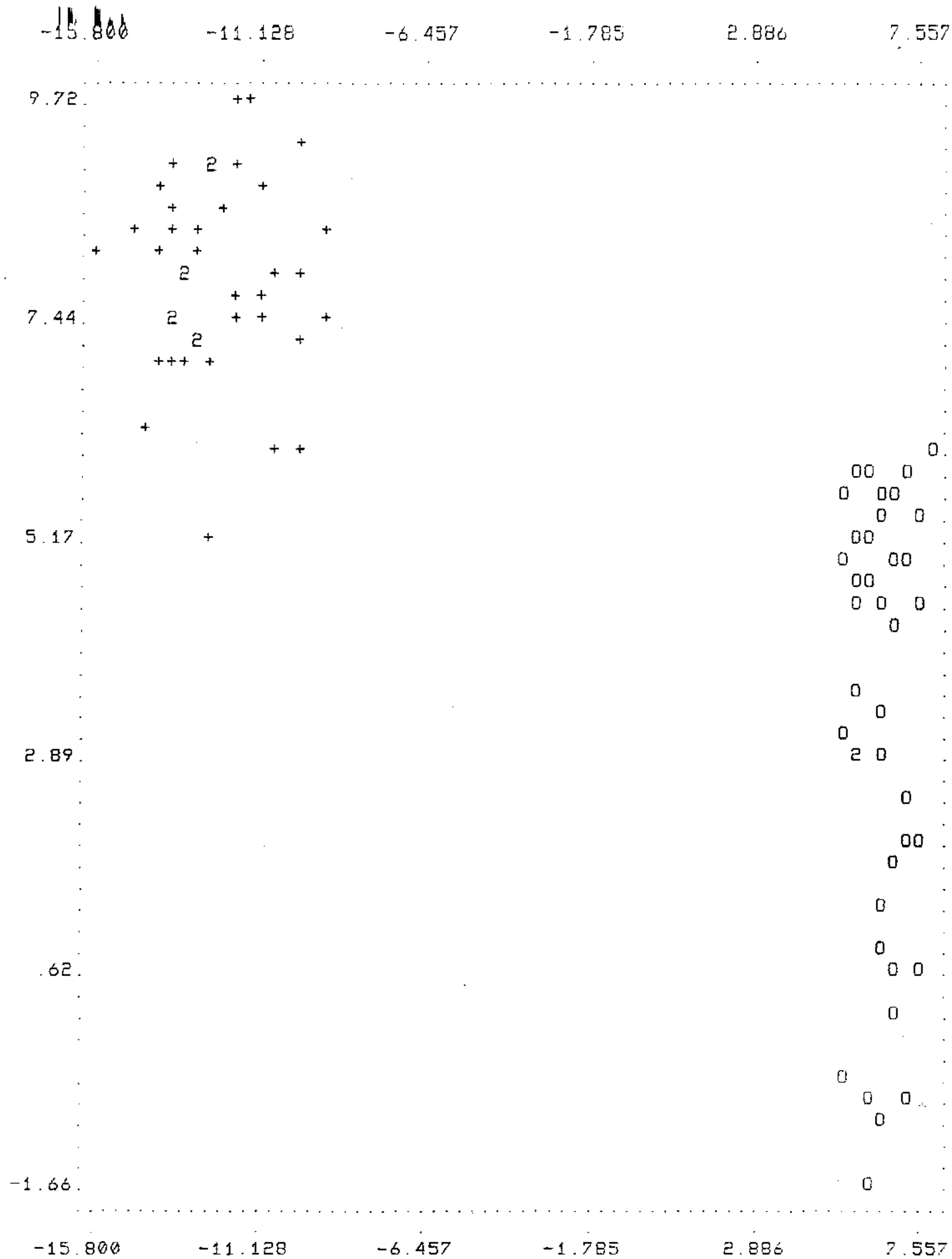
POPULACAO 1 = + POPULACAO 2 = 0



INTERVALO HORIZONTAL = .659E-01 INTERVALO VERTICAL = .299

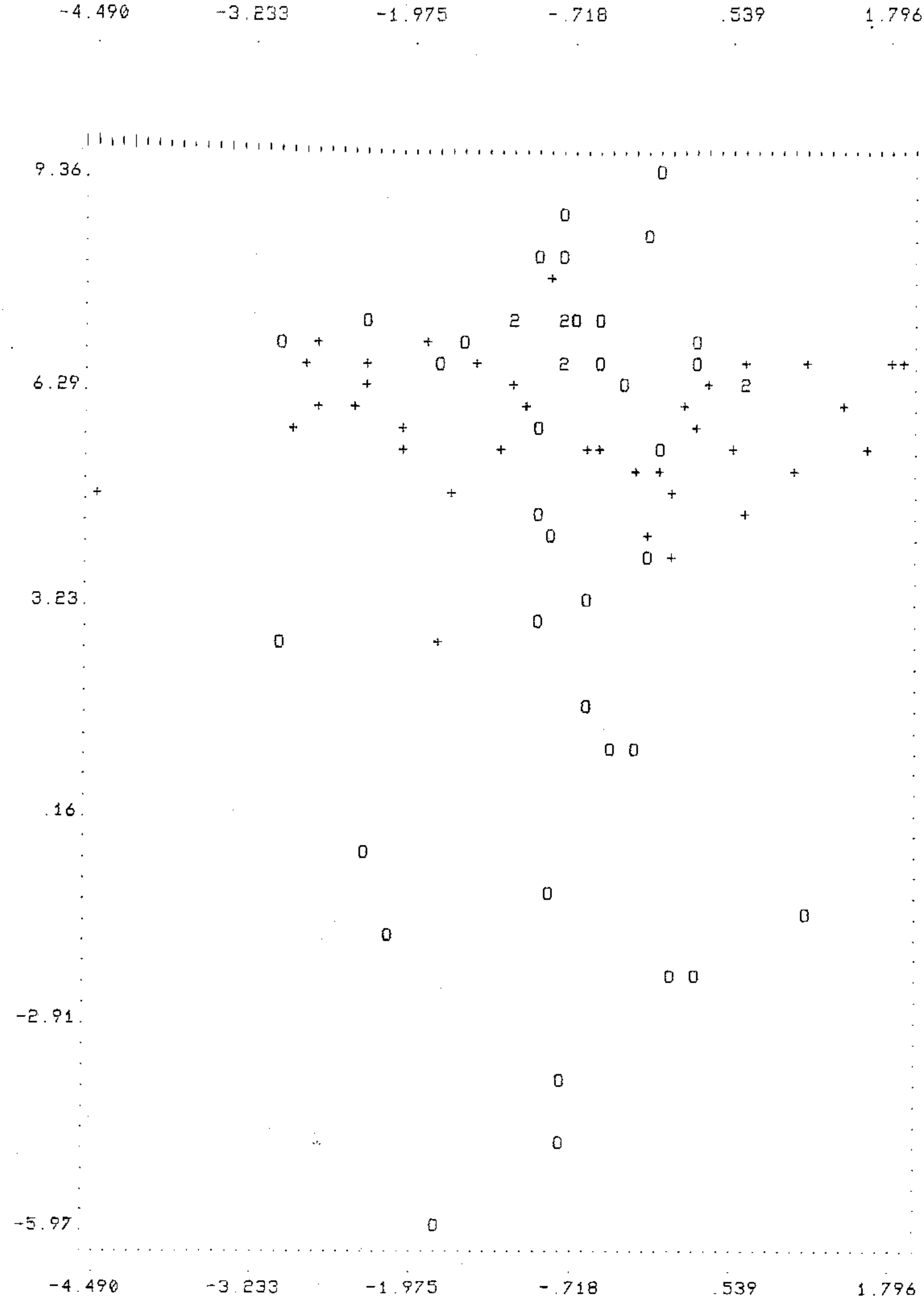
CONCOR - LINHAGENS [illegible] METODO DAS CORRELACOES

POPULACAO 1 = + POPULACAO 2 = 0



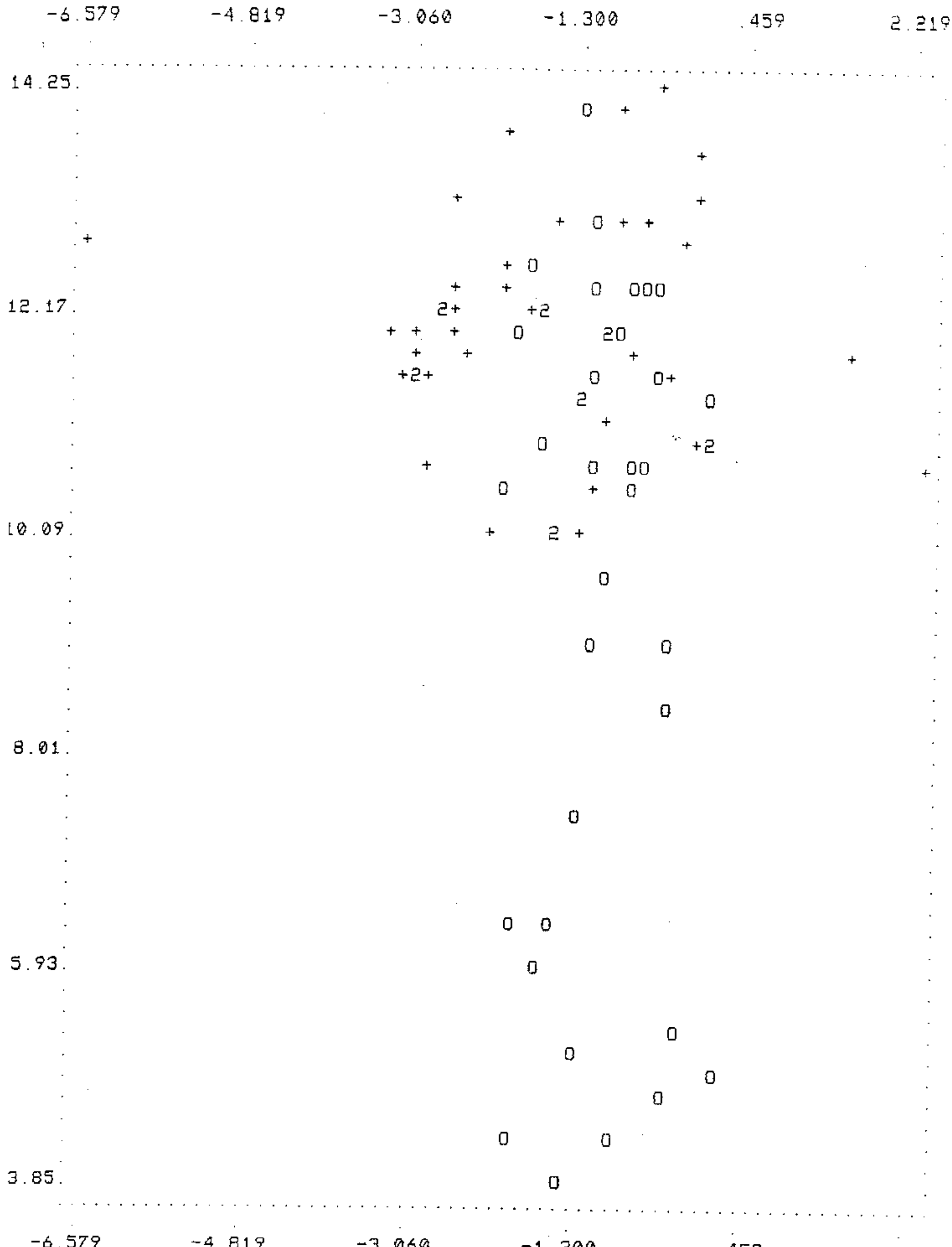
INTERVALO HORIZONTAL = .234 INTERVALO VERTICAL = .228

[illegible] - LINHAGENS 2 E 4 METODO DAS COVARIANCIAS
POPULACAO 1 = + POPULACAO 2 = 0



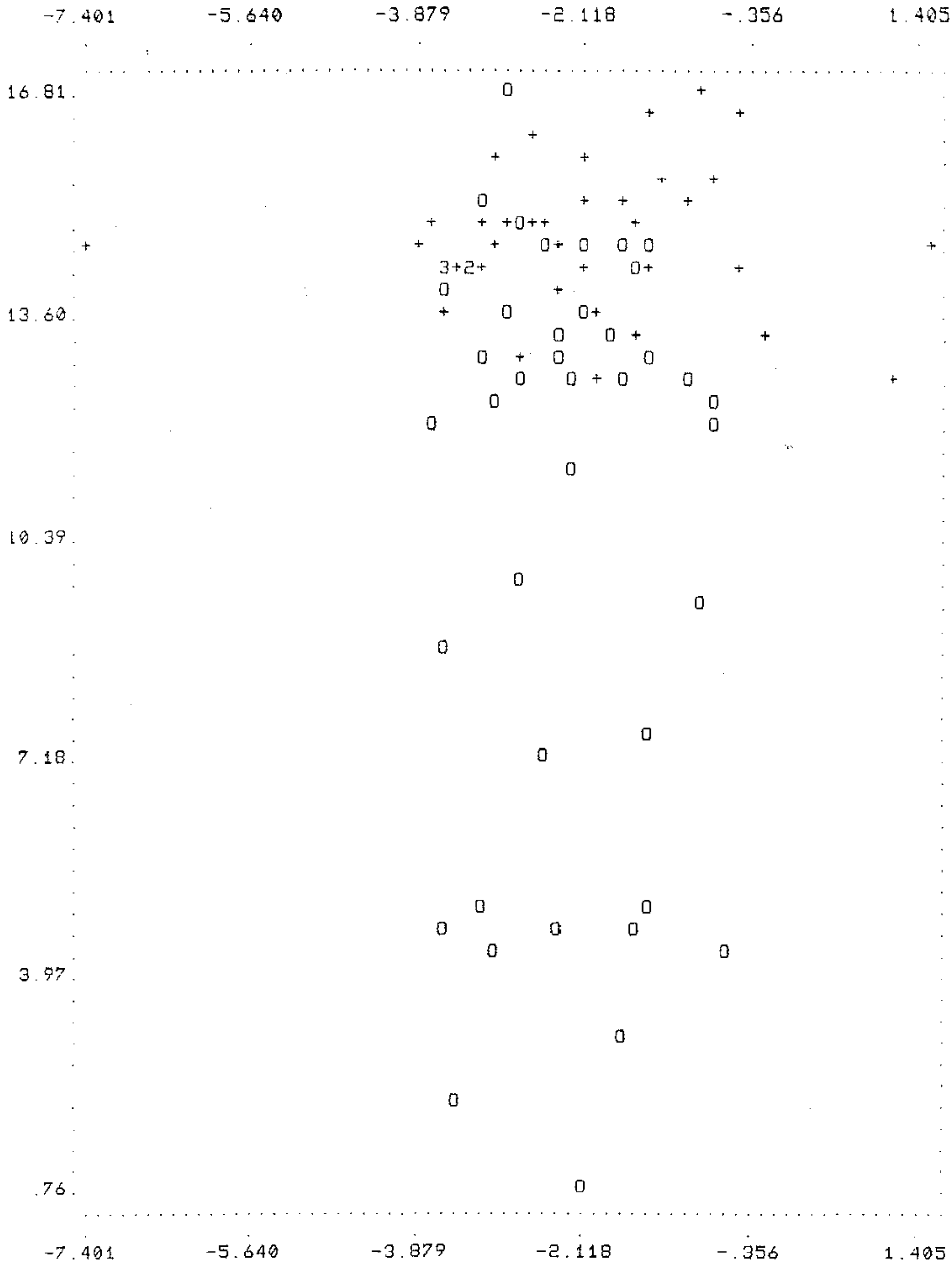
NTERVALO HORIZONTAL = .629E-01 INTERVALO VERTICAL = .307

CONCOR - LINHAGENS 3 E 4 METODO DAS CORRELACOES
POPULACAO 1 = + POPULACAO 2 = 0



ERVALO HORIZONTAL = .880E-01 INTERVALO VERTICAL = .208

CONCOR - LINHAGENS 3 E 4 METODO DAS COVARIANCIAS
POPULACAO 1 = + POPULACAO 2 = 0



ERVALO HORIZONTAL = .881E-01 INTERVALO VERTICAL = .321

Referências Bibliográficas

1. Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons. New York

2. Beasley, J.D. e Springer, S.G. (1977) The Percentage Points of the Normal Distribution. **Applied Statistics,26,** 118-120

3. Box, G.E.P. (1949) A general distribution theory for a class of likelihood criteria. **Biometrika,36, 317-346**

4. Dachs, J.N.W. (1988) *Estatística Computacional.* Livros Técnicos e Cientificos Editora Ltda. Rio de Janeiro

5. Fisher, R.A. (1936) The use of multiple measurements in taxonomic problems. **Annals of Eugenics,7,** 179-188

6. Flury, B. e Riedwyl H. (1988) *Multivariate Statistics - A pratical approach.* Chapman and Hall. New York

7. Flury, B. (1983) Some relations between the comparison of covariance matrices and principal component analysis.

Computacional Statistics & Data Analysis,93, 97-109

8. Flury, B. (1985) Analysis of linear combinations with extreme ratios of variance. Journal of the American Statistical Association,80, 915-922

9. Hehl, M.E. (1986) *Linguagem de Programação Estruturada FORTRAN 77.* Mc-Graw-Hill, Ltda. São Paulo

10. Johnson, R. A. e Wichern, D.W. (1982) *Applied multivariate Statistical Analysis.* Prentice-Hall. London

11. Larntz, K. e Perlman (1985) A simple test for the equality of correlation matrices.

12. Maindonald, J.H. (1984) *Statistical Computacional.* John Wiley & Sons. New York, NY.

13. Mardia, K.V., Kent, J.T. e Bibby, J.M. (1979) *Multivariate Analysis.* Academic Press. London

14 Morrison, D.F. (1967) *Multivariate Statistical Methods.* McGraw-Hill. New York

15. Mujander, K.L. e Bahttacharjee, G.P. (1973) Algorithm AS

63 The Incomplete Beta Integral. Applied Statistics,22, 409-411

16. Peizer, D.B. e Pratt, J.W. (1968) A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. American Statistical Association Journal,63, 1416-1455

17. Roveratti, E. (1979) *Estudo do efeito da castração em bovinos da raça Nelore, sobre o desenvolvimento do animal por ocasião de sua preparação em carcaças e sub-produtos*, Relatório Final de Estágio. IMECC, UNICAMP

18. Seber, G.A.F. (1984) *Multivariate Observations*. John Wiley & Sons. New York

19. Smith, C.A.B. (1947) Some examples of discrimination. Annals of Eugenics,13, 272-282

20. Ticeran, D. A. G. (1988) *Discriminação de duas populações multivariadas com base em desigualdade de matrizes de covariâncias*. Tese de Mestrado. UNICAMP Campinas