

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP
RELATÓRIO DE CONSULTA

TÍTULO: “Parâmetros para a descrição da distribuição de peças arqueológicas na superfície de terrenos arados”

PESQUISADOR: Astolfo Gomes de Mello Araujo

INSTITUIÇÃO: Museu de Arqueologia e Etnologia – USP

FINALIDADE: Publicação

PARTICIPANTES DA ENTREVISTA: Carlos Alberto de Bragança Pereira
Júlia Maria Pavan Soler
Bianca de Carla Hernandez Matos
Astolfo Gomes de Mello Araujo

DATA: 16/05/2000

FINALIDADE DA CONSULTA: Assessoria na análise de dados

RELATÓRIO ELABORADO POR: Bianca de Carla Hernandez Matos

1. INTRODUÇÃO

A Arqueologia tem se deparado cada vez mais com situações em que o material arqueológico aflora à superfície, principalmente por conta da ação de implementos agrícolas (arado, subsolador, grade). Apesar de tradicionalmente encarados como destruídos, tais sítios arqueológicos são fontes valiosas de informação, na medida em que é possível observar padrões de arranjo de peças em áreas bastantes extensas, ao contrário do que ocorre nas áreas restritas expostas pelas escavações.

A descrição matemática desses arranjos espaciais é de extrema importância para que seja possível fazer comparações entre sítios arqueológicos, desenvolver métodos mais eficazes de prospecção (detecção de material enterrado) e desenvolver modelos de dispersão de peças pela ação de implementos agrícolas.

Esse trabalho tem como objetivos determinar os parâmetros relevantes na descrição de uma distribuição de peças arqueológicas no espaço, avaliar quais funções matemáticas melhor descrevem as distribuições empiricamente observadas, simular amostragens com base nos parâmetros definidos e determinar quais desenhos amostrais e frações amostrais são necessárias para se obter parâmetros confiáveis a respeito da forma e conteúdo de um sítio arqueológico, caso o mesmo esteja coberto de vegetação.

2. DESCRIÇÃO DO ESTUDO E DAS VARIÁVEIS

Os dados deste estudo foram coletados da seguinte forma: após os sítios arqueológicos terem sido arados, uma equipe de arqueólogos marcou com bandeirolas todas as peças arqueológicas visíveis à superfície. Em seguida, foram registradas as posições de cada uma das peças marcadas, com o auxílio de um aparelho de topografia (teodolito eletrônico). Foram tomados os ângulos e as distâncias das peças, tendo-se transformado posteriormente essas medidas em coordenadas cartesianas, por meio de senos e cossenos. As medidas de distância

foram feitas por taqueometria, estando sujeitas a um erro que aumenta com a distância. Porém, tomou-se o cuidado de deslocar o aparelho sempre que a distância deste à peça se aproximasse de 30 m, fazendo com que o erro máximo nas medidas de distância esteja em torno de 10 cm. Além disso, as medidas são bastante confiáveis com relação às leituras de ângulos.

Até o momento, dois sítios arqueológicos foram sujeitos a esse tipo de coleta. Em um deles, foram registradas 2731 peças (Sítio 1), e no outro, 1880 peças (Sítio 2).

As variáveis consideradas nesse estudo, medidas em cada uma das peças encontradas, são:

- X: coordenada do eixo das abscissas da peça, em metros.
- Y: coordenada do eixo das ordenadas da peça, em metros.
- Material: material da peça (c: cerâmica; p: pedra).

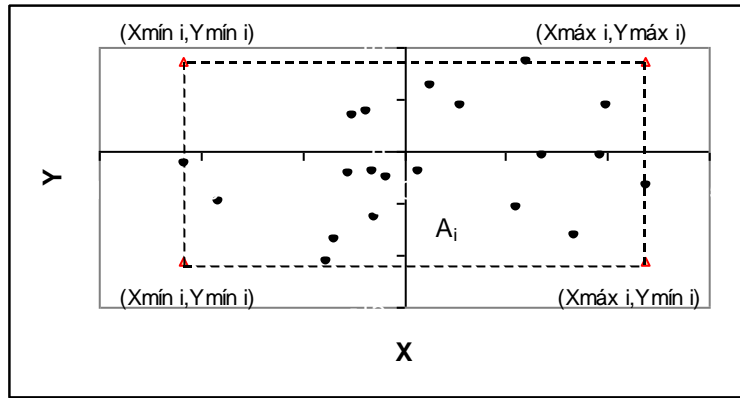
3. SUGESTÕES DO CEA

Para descrever a distribuição das peças arqueológicas no espaço, sugerimos inicialmente agrupá-las através do método de agrupamento das k-médias (ver Bussab et al, 1990, p. 58 ou Johnson e Wishern, 1992). Para utilizar esse método, devemos fixar o número k de partições desejadas. O critério de homogeneidade dentro dos grupos consiste em uma soma de quadrados residual, denotada por $SQRes(k)$, no caso de k grupos. Para decidirmos qual número de grupos é razoável, podemos considerar vários valores de k e, para cada um deles, calcular $SQRes(k)$, construindo, em seguida, um gráfico dos valores de $SQRes(k)$ contra k . Visualiza-se nesse gráfico que, a partir de um determinado número de grupos, que denotaremos por k_0 , a soma de quadrados residual tende a se estabilizar. Logo, é adequado dividir as peças arqueológicas em k_0 grupos.

Feito isso, o próximo passo é tentar descrever a distribuição das peças dentro de cada um dos k_0 grupos formados. Consideremos a área retangular A_i ocupada pelas peças do grupo i , isto é, a área compreendida no retângulo de vértices ($X_{\min i}$,

$Y_{\min i}$, $(X_{\max i}, Y_{\min i})$, $(X_{\min i}, Y_{\max i})$, $(X_{\max i}, Y_{\max i})$, onde $X_{\min i}$ e $X_{\max i}$ denotam, respectivamente, a menor e a maior coordenadas das abscissas observadas entre as peças do grupo i , e $Y_{\min i}$ e $Y_{\max i}$ denotam a menor e a maior ordenadas observadas nesse grupo, sendo i um número inteiro entre 1 e k_0 . A Figura 1 permite-nos visualizar melhor essa área.

Figura 1. Área retangular ocupada pelas peças do grupo i (A_i)



Seja N_i a variável aleatória correspondente ao número de peças arqueológicas encontradas em uma área de $t \text{ m}^2$ contida em A_i . Caso as peças se distribuam aleatoriamente dentro das áreas retangulares de cada um dos grupos formados, N_i terá distribuição de Poisson com parâmetro $t\lambda_i$ (ver Feller, 1976, p. 137), ou seja, a probabilidade de encontrar exatamente c peças em uma área de $t \text{ m}^2$ contida em A_i será dada por

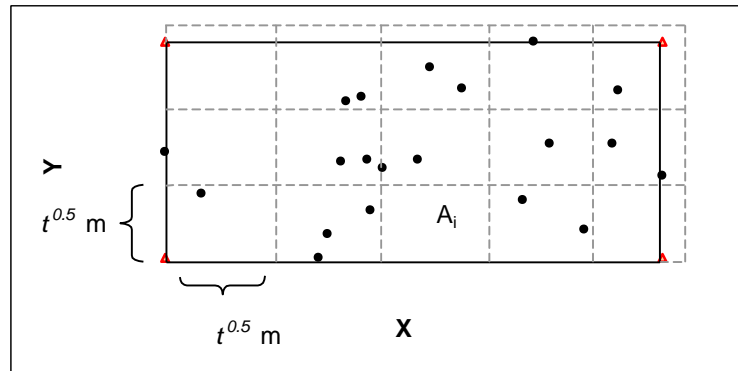
$$P(N_i = c) = \frac{e^{-t\lambda_i} (t\lambda_i)^c}{c!}, \quad c = 0, 1, 2, \dots$$

Nesse caso, o número esperado de peças encontradas em uma área de $t \text{ m}^2$ é igual a $t\lambda_i$. Além disso, $t\lambda_i$ também é o valor da variância de N_i (a variância é uma medida da dispersão dos números de peças encontrados, em relação ao número esperado).

Para estimar $t\lambda_i$, devemos dividir a área A_i em quadrados de área $t \text{ m}^2$ (lado $\sqrt{t} \text{ m}$). Caso a área A_i não seja divisível em um número exato de quadrados de

área $t \text{ m}^2$ (o que ocorre quando $X_{\text{máx}} - X_{\text{mín}}$ e/ou $Y_{\text{máx}} - Y_{\text{mín}}$ não são divisíveis por \sqrt{t}), será utilizado o menor número de quadrados de área $t \text{ m}^2$ necessários para conter todos os pontos do grupo i , sendo que o quadrado do canto inferior esquerdo tem um vértice em $(X_{\text{mín } i}, Y_{\text{mín } i})$. Essa divisão é ilustrada na Figura 2.

Figura 2. Divisão da área A_i em quadrados de lado $\sqrt{t} \text{ m}$



Denotemos por Q_i o número de quadrados obtidos no i -ésimo grupo e por T_i o número total de peças do i -ésimo grupo. A estimativa do parâmetro $t\lambda_i$ será dada pelo número médio de peças observadas por quadrado, isto é

$$t\hat{\lambda}_i = \frac{T_i}{Q_i}.$$

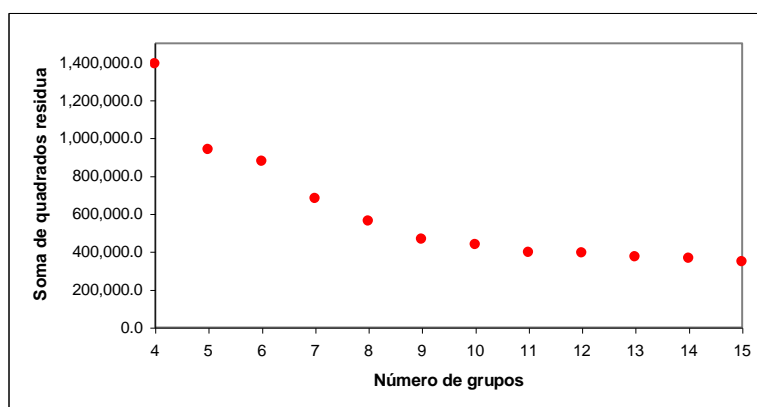
Para verificar a qualidade do ajuste da distribuição de Poisson, podemos comparar as proporções observadas de quadrados com c peças com as probabilidades previstas pela distribuição de Poisson com parâmetro $t\hat{\lambda}_i$. Quanto mais próximos estes valores, melhor as contagens de peças por quadrado são descritas pela distribuição de Poisson. Uma técnica formal para avaliar se a distribuição de Poisson se ajusta bem aos valores observados é o teste qui-quadrado para “bondade” de ajuste (ver Neter et al, 1990, p. 593).

A análise acima descrita foi feita para o Sítio 1, e é apresentada a seguir.

3.1. ANÁLISE DOS DADOS DO SÍTIO 1

Aplicou-se o método de agrupamento das k-Médias considerando números de grupo entre 4 e 15, e para cada um, calculou-se a soma de quadrados residual. Os resultados são apresentados na Figura 3.

Figura 3. Somas de quadrados residuais obtidas pelo método das k-Médias



Podemos observar que, a partir de 9 grupos, a diminuição na soma de quadrados residual é pequena ao incluir mais grupos. Logo, é razoável dividir as 2731 peças arqueológicas encontradas no Sítio 1 em 9 grupos.

A Figura 4 mostra as peças do sítio 1 divididas por grupo.

A Tabela 1 traz os números de peças, coordenadas X e Y mínimas e máximas e as coordenadas dos centróides de cada um dos 9 grupos, sendo que o centróide de determinado grupo é o ponto (X_c, Y_c) formado pelas médias das coordenadas X e Y de suas peças, podendo ser visto como o ponto de maior concentração de peças.

Figura 4. Grupos formados no Sítio 1 pelo método das k-Médias

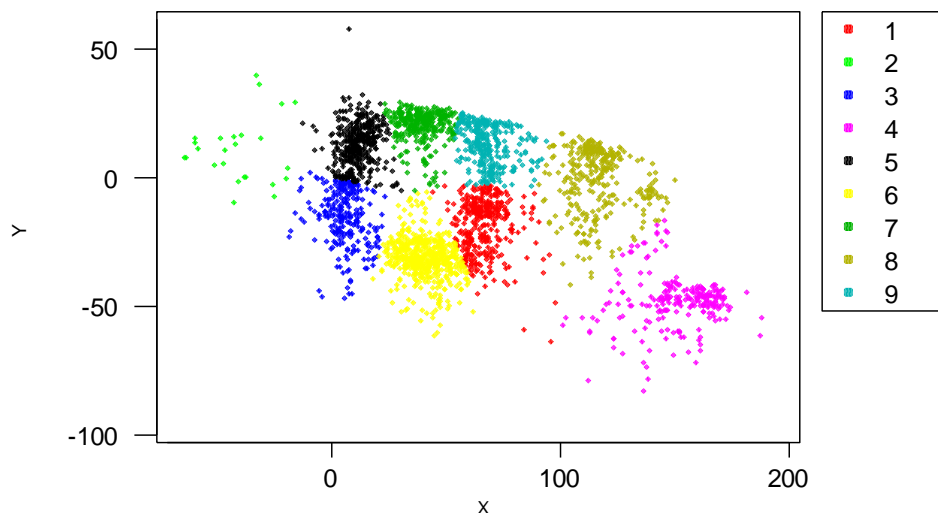
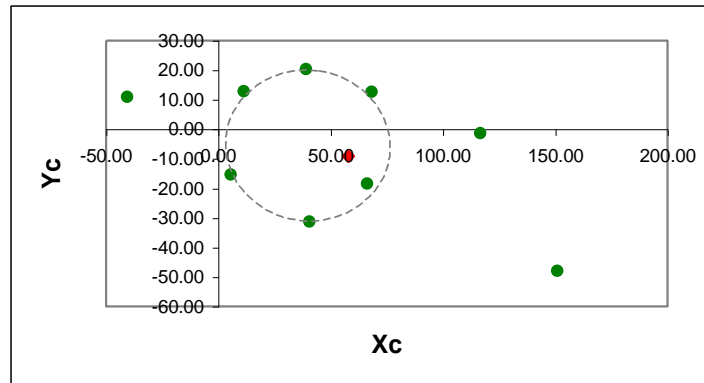


Tabela 1. Dados dos 9 grupos do Sítio 1

Grupo (i)	# peças (Ti)	$X_{\min i}$	$X_{\max i}$	$Y_{\min i}$	$Y_{\max i}$	Centróide	
						X_c	Y_c
1	359	44.36	97.72	-63.95	-3.12	66.54	-18.62
2	25	-64.66	-15.90	-9.85	39.49	-40.36	10.74
3	242	-18.61	23.51	-47.06	1.93	5.75	-15.52
4	241	101.46	188.42	-83.12	-16.48	151.11	-48.17
5	378	-12.74	30.46	-5.00	57.67	11.60	12.80
6	553	17.98	61.74	-61.38	-5.89	40.79	-31.41
7	309	22.98	55.21	-5.29	29.36	39.30	20.17
8	320	91.02	150.49	-42.04	17.29	116.73	-1.43
9	304	53.76	93.76	-3.31	24.98	68.46	12.44

Na Figura 5 são plotados os centróides dos 9 grupos. O ponto destacado (em vermelho) é o centróide geral, formado pelas médias de X e Y de todas as peças arqueológicas. É interessante notar que os centróides dos grupos 1, 3, 5, 6, 7 e 9 situam-se aproximadamente sobre uma circunferência, exibida no gráfico. Vale ressaltar que outras medidas, além dos centróides, podem ser avaliadas para resumir a informação de cada grupo, como por exemplo, as medianas das coordenadas.

Figura 5. Centróides dos grupos



Em cada um dos 9 grupos formados, dividiu-se a área A_i em quadrados de 1m^2 ($t=1\text{m}$), conforme explicado anteriormente. A Tabela 2 apresenta os números de peças, números de quadrados de 1m^2 , parâmetros da distribuição de Poisson estimados e seus intervalos de confiança com coeficiente de confiança de 95%, obtidos nos 9 grupos. No Anexo, apresentamos uma função construída para calcular essas quantidades.

Tabela 2. Dados dos 9 grupos do Sítio 1

Grupo (i)	# peças (T_i)	# quadrados (Q_i)	$\hat{\lambda}_i$	Erro Padrão de $\hat{\lambda}_i$	Limite inferior	Limite superior
1	359	3294	0.1090	0.0066	0.0960	0.1220
2	25	2450	0.0102	0.0020	0.0062	0.0142
3	242	2107	0.1149	0.0081	0.0991	0.1307
4	241	5829	0.0413	0.0031	0.0353	0.0474
5	378	2772	0.1364	0.0090	0.1187	0.1540
6	553	2464	0.2244	0.0117	0.2015	0.2474
7	309	1155	0.2675	0.0187	0.2309	0.3042
8	320	3600	0.0889	0.0054	0.0784	0.0994
9	304	1189	0.2557	0.0164	0.2236	0.2878

Observamos que os grupos 2 e 4 são os que apresentam os menores valores $\hat{\lambda}_i$, isto é, o número esperado de peças por quadrado de 1m^2 . Logo, a densidade de peças nesses grupos é pequena. Os grupos 6, 7 e 9 apresentam valores esperados altos, ou seja, a densidade de peças é alta. Vemos ainda que os intervalos de confiança para o parâmetro da distribuição de Poisson dos grupos 2 e 4 não têm pontos em comum com os intervalos dos grupos 6, 7 e 9.

Na Tabela 3, são apresentadas as freqüências e proporções observadas de quadrados com c peças, e as probabilidades previstas pela distribuição de Poisson com parâmetro $\hat{\lambda}_i$, para os 9 grupos.

Vemos que em todos os grupos as proporções observadas de quadrados com c peças ($c = 0,1,2, \dots$) são razoavelmente próximas das probabilidades previstas pela distribuição de Poisson. Ou seja, a distribuição de Poisson parece adequada para descrever a distribuição espacial das peças dentro de cada um dos grupos formados pelo método das k-Médias.

Tabela 2. Freqüências e proporções observadas de quadrados com exatamente c peças, e probabilidades de encontrar exatamente c peças previstas pela distribuição de Poisson de parâmetro $\hat{\lambda}_i$

Grupo		Número observado de peças nos quadrados							Total
		0	1	2	3	4	5	6	
1	Freq. obs.	3002	236	46	9	1	0	0	3294
	Prop. obs.	0.911	0.072	0.014	0.003	0.000	0.000	0.000	
	Prob. Prevista	0.897	0.098	0.005	0.000	0.000	0.000	0.000	
2	Freq. obs.	2425	25	0	0	0	0	0	2450
	Prop. obs.	0.990	0.010	0.000	0.000	0.000	0.000	0.000	
	Prob. prevista	0.990	0.010	0.000	0.000	0.000	0.000	0.000	
3	Freq. obs.	1894	190	19	2	2	0	0	2107
	Prop. obs.	0.899	0.090	0.009	0.001	0.001	0.000	0.000	
	Prob. prevista	0.891	0.102	0.006	0.000	0.000	0.000	0.000	
4	Freq. obs.	5623	178	23	4	0	1	0	5829
	Prop. obs.	0.965	0.031	0.004	0.001	0.000	0.000	0.000	
	Prob. prevista	0.959	0.040	0.001	0.000	0.000	0.000	0.000	
5	Freq. obs.	2500	195	57	14	3	3	0	2772
	Prop. obs.	0.902	0.070	0.021	0.005	0.001	0.001	0.000	
	Prob. prevista	0.873	0.119	0.008	0.000	0.000	0.000	0.000	
6	Freq. obs.	2060	295	78	25	4	1	1	2464
	Prop. obs.	0.836	0.120	0.032	0.010	0.002	0.000	0.000	
	Prob. prevista	0.799	0.179	0.020	0.002	0.000	0.000	0.000	
7	Freq. obs.	931	165	41	11	6	1	0	1155
	Prop. obs.	0.806	0.143	0.035	0.010	0.005	0.001	0.000	
	Prob. prevista	0.765	0.205	0.027	0.002	0.000	0.000	0.000	
8	Freq. obs.	3317	250	29	4	0	0	0	3600
	Prop. obs.	0.921	0.069	0.008	0.001	0.000	0.000	0.000	
	Prob. prevista	0.915	0.081	0.004	0.000	0.000	0.000	0.000	
9	Freq. obs.	944	199	36	8	1	1	0	1189
	Prop. obs.	0.794	0.167	0.030	0.007	0.001	0.001	0.000	
	Prob. prevista	0.774	0.198	0.025	0.002	0.000	0.000	0.000	

4. CONCLUSÃO

Sugere-se ao pesquisador entrar em contato com o CEA para submeter este trabalho à triagem de projetos que serão selecionados no segundo semestre de 2000, para que sejam analisados os demais sítios arqueológicos coletados, assim como para estudar mais detalhadamente a aplicação do método desenvolvido no aperfeiçoamento dos métodos de prospecção.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BUSSAB, W.O., MIAZAKI, É.S. e ANDRADE, D.F. (1990). **Introdução à Análise de Agrupamentos**. 9º Simpósio Brasileiro de Probabilidade e Estatística, São Paulo. 105p.
- JOHNSON, R.A. e WICHERN, D.W. (1992). **Applied Multivariate Statistical Analysis**. Englewood Cliffs: Prentice Hall. 642p.
- FELLER, Willian (1976). **Introdução à teoria das probabilidades e suas aplicações**. São Paulo, Blücher. 236p.
- NETER, J. e WASSERMAN, W. (1990). **Applied linear statistical models**. Homewood, Illinois: Richard D. Irwin.

Anexo: Função para calcular as estimativas dos parâmetros $\hat{\lambda}_i$'s

A seguir, é apresentada uma função desenvolvida no software S-Plus 4.5 que fornece como saída uma tabela contendo os valores mínimos e máximos de X e Y, número de peças, número de quadrados de área $t \text{ m}^2$, estimativa de $\hat{\lambda}_i$ e erro padrão estimado de $\hat{\lambda}_i$, para cada um dos grupos. Os argumentos de entrada são:

- **mat.dados:** matriz com 3 colunas e número de linhas igual ao número total de peças do sítio arqueológico. A primeira coluna contém as coordenadas X das peças, a segunda contém as coordenadas Y e a terceira contém o grupo ao qual a peça pertence, definido previamente pelo método das k-médias;
- **num.grupos:** número de grupos formados a partir do método das k-médias;
- **t:** área dos quadrados contidos nos grupos.

```
arqueologia<-function(mat.dados, num.grupos,t)

# Função para calcular as frequências de quadrados com c peças, em cada
um dos grupos

{
  mat.freq<-matrix(0,nrow=num.grupos,      ncol=30)      #matriz      das
distribuições observadas
  l<-sqrt(t) # lado dos quadrados
  #
  for(j in 1:num.grupos){ # preencher a 1a coluna de mat.freq com os
números dos grupos
    mat.freq[j,1]<-j
  }
  #
  mat.resumo<-matrix(0,nrow=num.grupos, ncol=9) # matriz de resumo
  #
  for(j in 1:num.grupos){ # preencher a 1a coluna de mat.resumo com os
números dos grupos
    mat.resumo[j,1]<-j
  }
  #
  for (i in 1:num.grupos){
    # obter pecas do grupo i
    pecas.grupo <- mat.dados[mat.dados[,3]==i,]
    xmin <- min(pecas.grupo[,1])
    xmax <- max(pecas.grupo[,1])
    ymin <- min(pecas.grupo[,2])
    ymax <- max(pecas.grupo[,2])
    mat.resumo[i, 3:6] <- c(xmin,xmax,ymin,ymax)
    # obter sub-grupos do grupo i
    int.x <- ceiling((xmax-xmin)/l) # numero de intervalos em x
```

```

int.y <- ceiling((ymax-ymin)/1)
num.subgrupos <- int.x * int.y # número de quadrados
sub.grupos <- matrix(0,nrow=num.subgrupos,ncol=3)
# coluna 1: n. do subgrupo, coluna 2: xmin, coluna 3: ymin
sub.ymin <- ymin
for(j in 1:int.y) {
  sub.xmin <- xmin
  for (k in 1:int.x) {
    sub.grupos[(j-1)*int.x+k,] <- c((j-
1)*int.x+k,sub.xmin, sub.ymin)
    sub.xmin <- sub.xmin +1
  }
  sub.ymin <- sub.ymin+1
}
#
# contar peças por subgrupo
cont.pecas <- matrix(0, nrow=num.subgrupos, ncol=2)
# numero de pecas em cada subgrupo
#
for (j in 1:num.subgrupos) {
  cont.pecas[j,1] <- j
}
#
# encontrar subgrupo de cada peca
#
for (j in 1:dim(pecas.grupo)[1]){ # para cada peça do grupo
  grupo.pecas <- acha.grupo(pecas.grupo[j,1],
pecas.grupo[j,2],
+ sub.grupos, int.x, 1)
  cont.pecas[grupo.pecas,2]<-cont.pecas[grupo.pecas,2]+1
}
#
mat.resumo[i,2]<- dim(pecas.grupo)[1]
mat.resumo[i,7]<-dim(sub.grupos)[1]
mat.resumo[i,8]<-mat.resumo[i,2]/mat.resumo[i,7] # estimativa
de lambda
mat.resumo[i,9]<-sqrt(var(cont.pecas[,2])/num.subgrupos) #
erro padrao de lambda
for (j in 1:num.subgrupos){
  mat.freq[i,
cont.pecas[j,2]+2]+1
  cont.pecas[j,2]+2]<-mat.freq[i,
}
}
print(mat.resumo)
print(mat.freq)
}

```

```

acha.grupo <- function(X,Y, sub.grupos, int.x, 1) { # encontrar
sub.grupo da peça
  achou.grupo <- F
  pos.x <- trunc((X-sub.grupos[1,2])/1)+1 # posicao x
  k<-pos.x
  while (!achou.grupo){
    if (Y>=sub.grupos[k,3] && Y< sub.grupos[k,3]+1){

```

```
        achou.grupo<-T
        grupo.pecas <- k
    }
    k<-k+int.x
}
grupo.pecas
}
```