ANÁLISE DOS DADOS COMPOSICIONAIS DE UM LAGO ÁRTICO

Conjunto de Dados 5, Aitchison (1986, pg359).

O primeiro passo é carregar o pacote para análise de dados composicionais:

```
> require(compositions)
```

Lendo dados do arquivo de dados ”ArticoLakepont.txt” extraído de Aitchison (1986). Este contém observações de areia, silte e argila em 39 amostras sedimentares num Lago Ártico onde $V1$, $V2$ e $V3$ correspondem a areia (sand), silte (silt) e argila (clay), respectivamente.

```
> dados <- read.table("ArticoLakepont.txt")
> head(dados)
```

```
    V1   V2   V3
1 77.5 19.5  3.0
2 71.9 24.9  3.2
3 50.7 36.1 13.2
4 52.2 40.9  6.6
5 70.0 26.5  3.5
6 66.5 32.2  1.3
```

Vendo os dados como composições, expressas em forma de frações decimais:

```
> comp <- acomp(dados)
> names(comp) <- c("Areia", "Silte", "Argila")
> head(comp)
```

```
[1] 0.7750000 0.7190000 0.5070000 0.5235707 0.7000000 0.6650000
```

Construindo um diagrama ternário:

```
> par(mar=c(5,0.5,0,0.5),mgp=c(2,0.8,0))
> plot(comp)
```

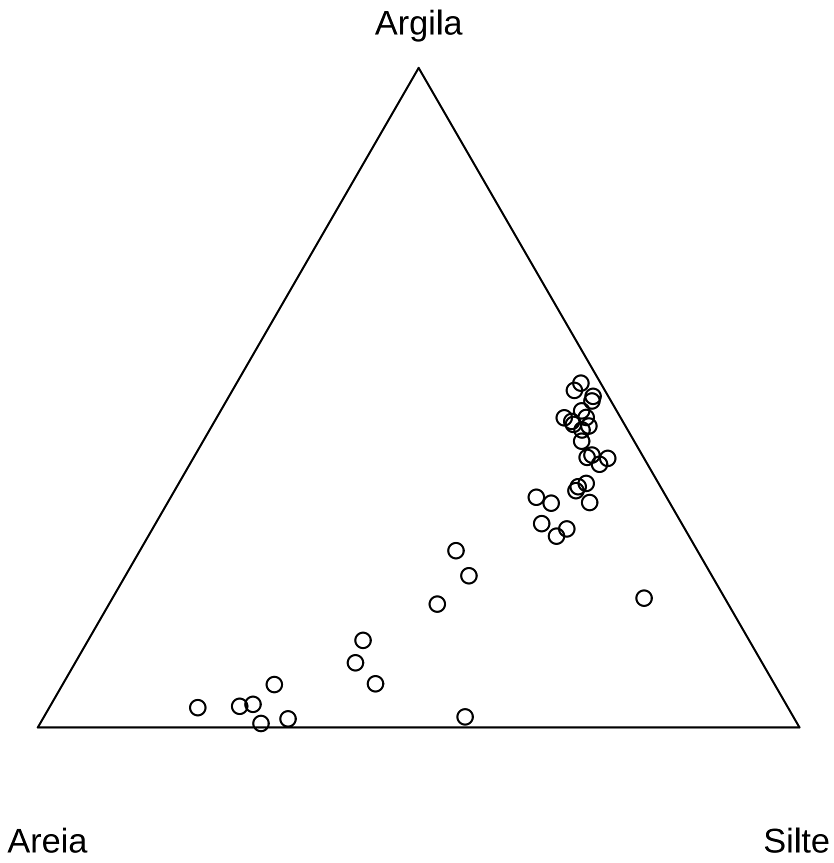

Figura 1: Diagrama Ternário para Dados do Lago Ártico

O diagrama mostra que existe um cluster substancial de composições com baixa proporção de areia, e proporções aproximadamente iguais de silte e argila. Pela extensão dos pontos ao longo do lado areia-argila (V1-V3) e ao longo do lado silte-argila (V2-V3), pode-se afirmar que razão areia/argila é muito mais variável que a razão silte/argila.

Vendo os dados como quantidades:

```
> quant <- aplus(dados)
> names(quant) <- c("Areia", "Silte", "Argila")
> head(quant)
```

```
[1] 77.5 71.9 50.7 52.2 70.0 66.5
```

Matriz de diagramas de dispersão:

```
> plot(quant)
```

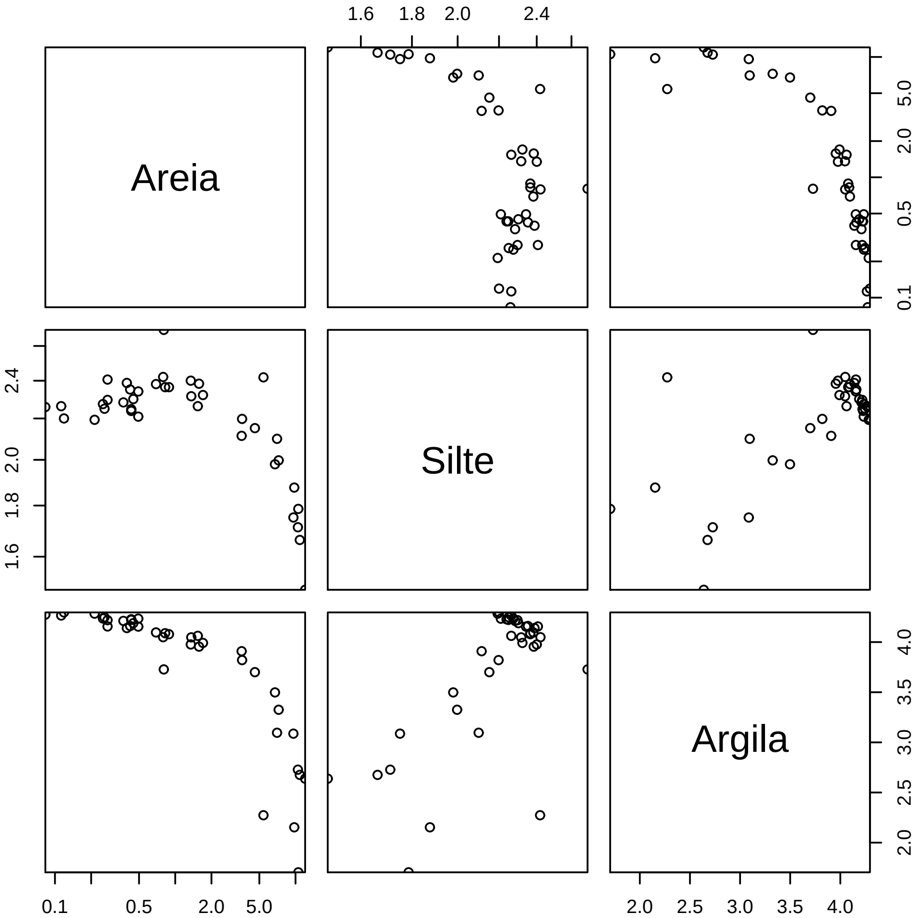

Figura 2: areia × silte, areia × argila e silte × argila

Para incluir no diagrama ternário a média geométrica, representada por um ponto vermelho e também a região 2-sigma, executa-se os comandos:

```
> plot(comp)
> plot(mean(comp),add=T,pch=20,col="red")
> # exibindo a região 2-sigma
> ellipses(mean(comp),var(comp),col="red", r=2)
```

A seguir apresenta-se os comandos para a obtenção das principais estatísticas descritivas.

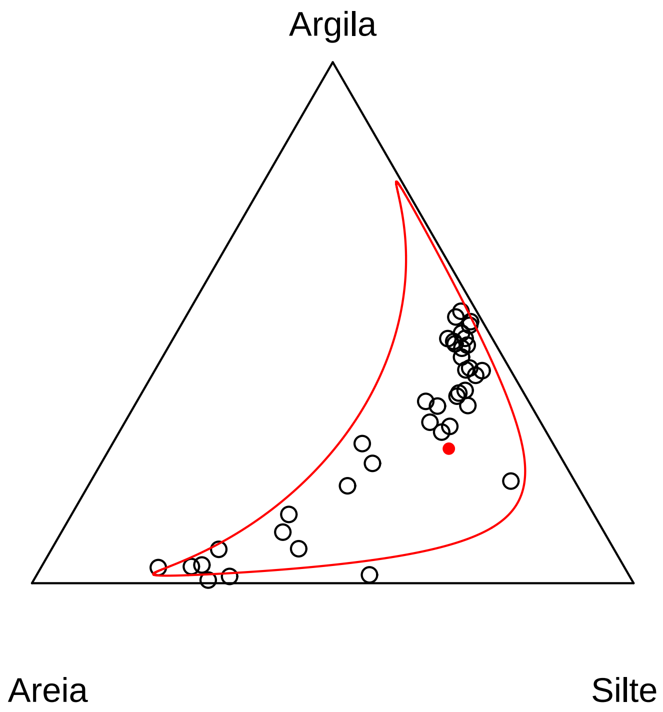

Figura 3: Diagrama Ternário com a Região de Confiança

O centro da distribuição, definido como a média aritmética das médias geométricas de cada componentes é obtido através de

```
> mean(comp)

    Areia     Silte    Argila
0.1779980 0.5637489 0.2582530
attr(,"class")
[1] "acomp"
```

A variância na CLR estrutura do espaço euclidiano é

```
> var(comp)
```

```
           Areia      Silte     Argila
Areia   1.3011517 -0.1355277 -1.1656240
Silte  -0.1355277  0.0689969  0.0665308
Argila -1.1656240  0.0665308  1.0990932
```

Obtém-se a variância métrica que nada mais é do que o traço da matriz CLR fazendo-se

```
> mvar(comp)

[1] 2.469242
```

O desvio padrão métrico - sqrt(mvar/(D-1)) é dado por

```
> msd(comp)

[1] 1.111135
```

O desvio padrão clássico de cada componente é obtido fazendo

```
> sd(comp)

    Areia     Silte    Argila
0.2454293 0.1011509 0.1715132
```

A matriz variação, de LR autocovariâncias, com elementos $var(log(x_i/x_j))$ é

```
> variation(comp)

          Areia    Silte   Argila
Areia  0.000000 1.641204 4.731493
Silte  1.641204 0.000000 1.035028
Argila 4.731493 1.035028 0.000000
```

Ainda, é possível calcular a variância total, a partir da variância na estrutura CLR. Para isto cria-se um vetor de uns com 3 elementos.

```
> x<-c(1,1,1)
> x

[1] 1 1 1

> # calculando o transposto
> xt<-t(x)
> xt

     [,1] [,2] [,3]
[1,]    1    1    1
```

e, então, obtém-se a variância total:

```
> varTot <- mvar(comp)-(1/3)*(xt%*%var(comp)%*%x)
> varTot

         [,1]
[1,] 2.469242
```