

ANÁLISE DOS DADOS COMPOSICIONAIS DE Gener1Mod

Carregando o pacote *compositions*

```
> require(compositions)
```

PREPARANDO OS DADOS PARA ANÁLISE: *Silte* obtido pelo complementar

Lendo os dados

```
> dados <- read.table("DadosGener1Mod.txt", head = TRUE)
> head(dados)
```

	X	Y	AMG	AM	AF	AMF	Silte	Argila	Altimetria	Superficie
1	0	0	2.1	10.4	11.4	5.9	12.8	57	641.37	1
2	50	0	1.1	12.0	9.8	4.6	13.0	58	640.84	1
3	100	0	1.1	7.9	6.9	4.0	15.7	64	640.00	1
4	150	0	1.1	6.6	6.6	4.0	16.5	65	637.50	1
5	200	0	0.5	7.1	7.1	3.5	14.9	66	634.66	1
6	250	0	1.5	3.4	6.6	4.4	14.1	70	631.33	1

Criando uma coluna cujos valores são a soma das colunas AMG,AM,AF e AMF e uma coluna cujos valores correspondem ao complementar da soma dos 4 tipos de Areias e Argila.

```
> dadosC <-with(dados,matrix(c(X,Y,Areia=(AMG+AM+AF+AMF),
+                               SilteC=(100-AMG-AM-AF-AMF-Argila),Silte,Argila,
+                               Altimetria,Superficie),nrow=412))
> #dadosC
> head(dadosC)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	0	0	29.8	13.2	12.8	57	641.37	1
[2,]	50	0	27.5	14.5	13.0	58	640.84	1
[3,]	100	0	19.9	16.1	15.7	64	640.00	1
[4,]	150	0	18.3	16.7	16.5	65	637.50	1
[5,]	200	0	18.2	15.8	14.9	66	634.66	1
[6,]	250	0	15.9	14.1	14.1	70	631.33	1

Transformando dadosC em data.frame.Observa-se que a diferença máxima é igual a 2.

```
> dadosC <- as.data.frame(dadosC)
> names(dadosC)<- c('X','Y','Areia','SilteC','Silte','Argila',
+                  'Altimetria','Superficie')
> #dadosC
> head(dadosC)
```

	X	Y	Areia	SilteC	Silte	Argila	Altimetria	Superficie
1	0	0	29.8	13.2	12.8	57	641.37	1
2	50	0	27.5	14.5	13.0	58	640.84	1
3	100	0	19.9	16.1	15.7	64	640.00	1
4	150	0	18.3	16.7	16.5	65	637.50	1
5	200	0	18.2	15.8	14.9	66	634.66	1
6	250	0	15.9	14.1	14.1	70	631.33	1

```
> # calculando a máxima diferença
> max(dadosC$SilteC-dadosC$Silte)
```

```
[1] 2
```

Acrescentando duas colunas que são a soma dos componentes Areia, Silte e Argila das 412 composições (última coluna) e a soma destes 3 componentes mas considerando-se Silte como resultado do complementar Areia e Argila. A diferença máxima é igual a 2.

```
> dadosC <-with(dadosC,matrix(c(X,Y,Areia,SilteC,Silte,Argila,
+                               Altimetria,Superficie,Areia+SilteC+Argila,
+                               Areia+Silte+Argila),nrow=412))
> #dadosC
> head(dadosC)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0	0	29.8	13.2	12.8	57	641.37	1	100	99.6
[2,]	50	0	27.5	14.5	13.0	58	640.84	1	100	98.5
[3,]	100	0	19.9	16.1	15.7	64	640.00	1	100	99.6
[4,]	150	0	18.3	16.7	16.5	65	637.50	1	100	99.8
[5,]	200	0	18.2	15.8	14.9	66	634.66	1	100	99.1
[6,]	250	0	15.9	14.1	14.1	70	631.33	1	100	100.0

Transformando dadosC em data.frame

```
> dadosC <- as.data.frame(dadosC)
> names(dadosC) <- c("X", "Y", "Areia", "SilteC", "Silte", "Argila", "Alti.",
+ "Sup.", "SomaC", "Soma")
> head(dadosC)
```

	X	Y	Areia	SilteC	Silte	Argila	Alti.	Sup.	SomaC	Soma
1	0	0	29.8	13.2	12.8	57	641.37	1	100	99.6
2	50	0	27.5	14.5	13.0	58	640.84	1	100	98.5
3	100	0	19.9	16.1	15.7	64	640.00	1	100	99.6
4	150	0	18.3	16.7	16.5	65	637.50	1	100	99.8
5	200	0	18.2	15.8	14.9	66	634.66	1	100	99.1
6	250	0	15.9	14.1	14.1	70	631.33	1	100	100.0

Montando um novo arquivo somente com os dados composicionais que se referem às colunas Areia, SilteC e Argila.

```
> dadosCom <- with(dadosC, matrix(c(Areia, SilteC, Argila), nrow = 412))
> head(dadosCom)
```

	[,1]	[,2]	[,3]
[1,]	29.8	13.2	57
[2,]	27.5	14.5	58
[3,]	19.9	16.1	64
[4,]	18.3	16.7	65
[5,]	18.2	15.8	66
[6,]	15.9	14.1	70

Novamente transformando dadosCom em data.frame

```
> dados <- as.data.frame(dadosCom)
> names(dados) <- c("Areia", "Silte", "Argila")
> head(dados)
```

	Areia	Silte	Argila
1	29.8	13.2	57
2	27.5	14.5	58
3	19.9	16.1	64
4	18.3	16.7	65
5	18.2	15.8	66
6	15.9	14.1	70

INICIANDO ANÁLISE COMPOSICIONAL

Na classe `acomp` utilizada a seguir as quantidades individuais são parte de um todo e os dados são analisados na geometria relativa.

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

[1] 0.298 0.275 0.199 0.183 0.182 0.159

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

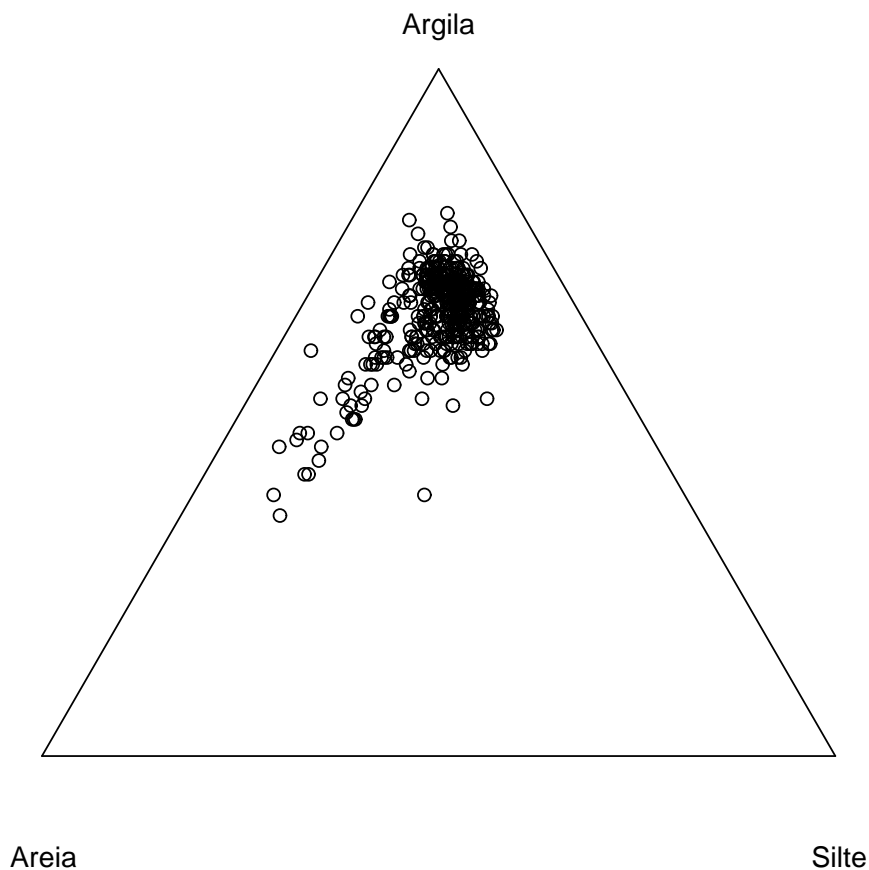


Figura 1: Diagrama Ternário para Areia, Silte e Argila.

Pela Figura 1, observa-se que existe um cluster substancial de composições com alta proporção de Argila e proporções aproximadamente iguais de Silte e Areia. Pela extensão dos pontos ao

longo dos lados Areia/Argila e Areia/Silte tem-se que a razão Argila/Areia é mais variável que a razão Silte/Areia. Isto pode ser confirmado pelos gráficos da Figura 2 a seguir, porém antes é necessário que os dados sejam vistos como quantidades

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

```
[1] 29.8 27.5 19.9 18.3 18.2 15.9
```

Construindo uma matriz de diagramas de dispersão

```
> plot(quant)
```

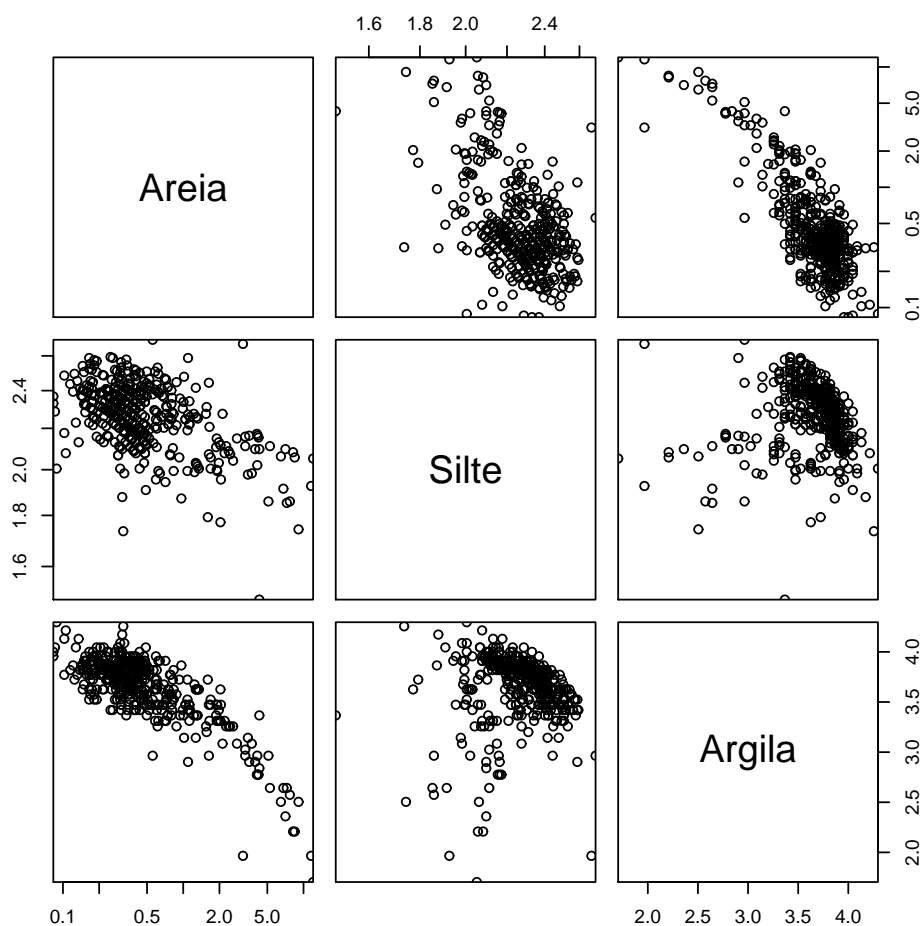


Figura 2: Areia \times Silte, Areia \times Argila e Silte \times Argila.

Incluindo a média geométrica e a região 2 e 4-sigma no diagrama ternário

```

> plot(comp)
> plot(mean(comp), add = T, pch = 20, col = "red")
> ellipses(mean(comp), var(comp), col = "red", r = 2)
> ellipses(mean(comp), var(comp), col = "red", r = 4)

```

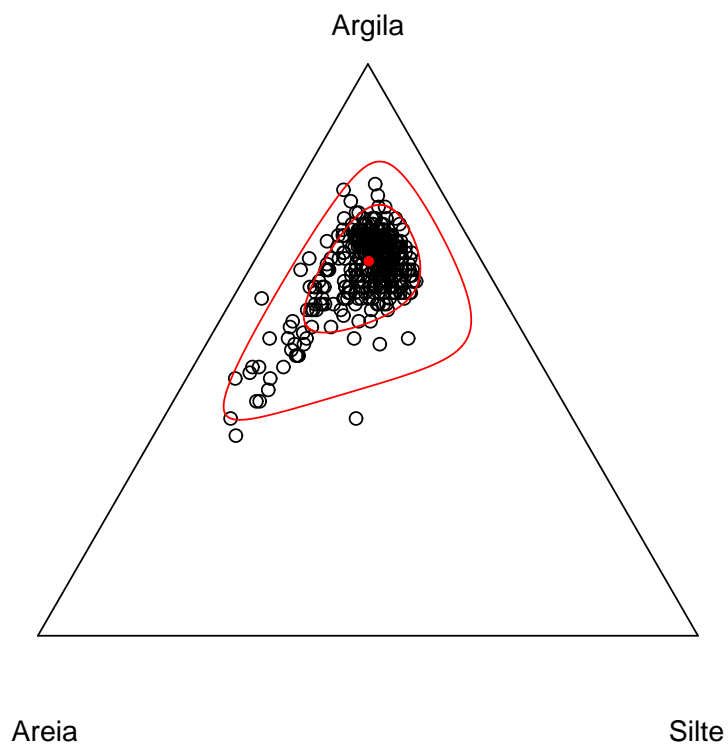


Figura 3: Diagrama Ternário com região de Confiança.

O centro da distribuição é dado por

```

> mean(comp)

      Areia      Silte      Argila
0.1711891 0.1736919 0.6551190
attr(,"class")
[1] "acomp"

```

A variância na estrutura CLR do espaço euclidiano é

```

> var(comp)

```

	Areia	Silte	Argila
Areia	0.08394583	-0.054111774	-0.029834059
Silte	-0.05411177	0.051160782	0.002950992
Argila	-0.02983406	0.002950992	0.026883067

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

```
> mvar(comp)
```

```
[1] 0.1619897
```

O desvio padrão métrico - $\sqrt{\text{mvar}/(D-1)}$ é dado por

```
> msd(comp)
```

```
[1] 0.2845959
```

O desvio padrão clássico é obtido através de

```
> sd(comp)
```

	Areia	Silte	Argila
	0.07122083	0.03972980	0.06312555

E a matriz variação, estrutura LR, cujos elementos são dados por $\text{var}(\log(x_i/x_j))$ é:

```
> variation(comp)
```

	Areia	Silte	Argila
Areia	0.0000000	0.24333016	0.17049702
Silte	0.2433302	0.00000000	0.07214187
Argila	0.1704970	0.07214187	0.00000000

Por fim, a variância total, a partir da variância na estrutura CLR é obtida. Para isto cria-se um vetor com 3 elementos iguais a 1 e o transposto deste

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

```
[1] 1 1 1
```

```
> xt <- t(x)
```

```
> xt
```

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

e, então:

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

```
> varTot
```

```
      [,1]
[1,] 0.1619897
```

PREPARANDO OS DADOS PARA ANÁLISE: *Areia* obtida pelo complementar

Lendo os dados

```
> dados <- read.table("DadosGener1Mod.txt", head = TRUE)
```

```
> head(dados)
```

	X	Y	AMG	AM	AF	AMF	Silte	Argila	Altimetria	Superficie
1	0	0	2.1	10.4	11.4	5.9	12.8	57	641.37	1
2	50	0	1.1	12.0	9.8	4.6	13.0	58	640.84	1
3	100	0	1.1	7.9	6.9	4.0	15.7	64	640.00	1
4	150	0	1.1	6.6	6.6	4.0	16.5	65	637.50	1
5	200	0	0.5	7.1	7.1	3.5	14.9	66	634.66	1
6	250	0	1.5	3.4	6.6	4.4	14.1	70	631.33	1

Criando uma coluna cujos valores são a soma das colunas AMG,AM,AF e AMF e uma coluna cujos valores correspondem ao complementar da soma Silte mais Argila. Observa-se que a diferença máxima é de 0.4.

```
> dadosC <-with(dados,matrix(c(X,Y,Sareia=(AMG+AM+AF+AMF),
+                               Dareia=(100-Silte-Argila),Silte,Argila,
+                               Altimetria,Superficie),nrow=412))
> head(dadosC)
```


	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	0	0	29.8	30.2	12.8	57	641.37	1
[2,]	50	0	27.5	29.0	13.0	58	640.84	1
[3,]	100	0	19.9	20.3	15.7	64	640.00	1
[4,]	150	0	18.3	18.5	16.5	65	637.50	1
[5,]	200	0	18.2	19.1	14.9	66	634.66	1
[6,]	250	0	15.9	15.9	14.1	70	631.33	1

Transformando dadosC em data.frame

```
> dadosC <- as.data.frame(dadosC)
> names(dadosC) <- c('X','Y','Sareia','Dareia','Silte','Argila',
+                    'Altimetria','Superficie')
> head(dadosC)
```

	X	Y	Sareia	Dareia	Silte	Argila	Altimetria	Superficie
1	0	0	29.8	30.2	12.8	57	641.37	1
2	50	0	27.5	29.0	13.0	58	640.84	1
3	100	0	19.9	20.3	15.7	64	640.00	1
4	150	0	18.3	18.5	16.5	65	637.50	1
5	200	0	18.2	19.1	14.9	66	634.66	1
6	250	0	15.9	15.9	14.1	70	631.33	1

Acrescentando duas colunas que são a soma dos componentes das 412 composições considerando-se a Areia como resultado da soma dos 4 tipos de Areia e considerando-se a Areia como resultado do complementar de Silte e Argila. A diferença máxima é igual a 2.

```
> dadosC <-with(dadosC,matrix(c(X,Y,Sareia,Dareia,Silte,Argila,Altimetria,
+                               Superficie,Sareia+Silte+Argila,Dareia+Silte+Argila),nrow=412))
> head(dadosC)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0	0	29.8	30.2	12.8	57	641.37	1	99.6	100
[2,]	50	0	27.5	29.0	13.0	58	640.84	1	98.5	100
[3,]	100	0	19.9	20.3	15.7	64	640.00	1	99.6	100
[4,]	150	0	18.3	18.5	16.5	65	637.50	1	99.8	100
[5,]	200	0	18.2	19.1	14.9	66	634.66	1	99.1	100
[6,]	250	0	15.9	15.9	14.1	70	631.33	1	100.0	100

Transformando dadosC em data.frame

```
> dadosC <- as.data.frame(dadosC)
> names(dadosC) <- c("X", "Y", "Sareia", "Dareia", "Silte", "Argila", "Alti.",
+ "Sup.", "SomaS", "SomaD")
> head(dadosC)
```

	X	Y	Sareia	Dareia	Silte	Argila	Alti.	Sup.	SomaS	SomaD
1	0	0	29.8	30.2	12.8	57	641.37	1	99.6	100
2	50	0	27.5	29.0	13.0	58	640.84	1	98.5	100
3	100	0	19.9	20.3	15.7	64	640.00	1	99.6	100
4	150	0	18.3	18.5	16.5	65	637.50	1	99.8	100
5	200	0	18.2	19.1	14.9	66	634.66	1	99.1	100
6	250	0	15.9	15.9	14.1	70	631.33	1	100.0	100

Montando um novo arquivo somente com os dados composicionais que se referem às colunas Dareia, Silte e Argila.

```
> dadosCom <- with(dadosC, matrix(c(Dareia, Silte, Argila), nrow = 412))
> head(dadosCom)
```

	[,1]	[,2]	[,3]
[1,]	30.2	12.8	57
[2,]	29.0	13.0	58
[3,]	20.3	15.7	64
[4,]	18.5	16.5	65
[5,]	19.1	14.9	66
[6,]	15.9	14.1	70

Novamente transformando dadosCom em data.frame

```
> dados <- as.data.frame(dadosCom)
> names(dados) <- c("Areia", "Silte", "Argila")
> head(dados)
```

	Areia	Silte	Argila
1	30.2	12.8	57
2	29.0	13.0	58

3	20.3	15.7	64
4	18.5	16.5	65
5	19.1	14.9	66
6	15.9	14.1	70

INICIANDO ANÁLISE COMPOSICIONAL

Na classe `acomp` utilizada a seguir as quantidades individuais são parte de um todo e os dados são analisados na geometria relativa.

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

```
[1] 0.302 0.290 0.203 0.185 0.191 0.159
```

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

Pela Figura 4, observa-se que existe um cluster substancial de composições com alta proporção de Argila e proporções aproximadamente iguais de Silte e Areia. Pela extensão dos pontos ao longo dos lados Areia/Argila e Areia/Silte tem-se que a razão Argila/Areia é mais variável que a razão Silte/Areia. Isto pode ser confirmado pelos gráficos da Figura 5 a seguir, porém antes é necessário que os dados sejam vistos como quantidades

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

```
[1] 30.2 29.0 20.3 18.5 19.1 15.9
```

Construindo uma matriz de diagramas de dispersão

```
> plot(quant)
```

Incluindo a média geométrica e a região 2 e 4-sigma no diagrama ternário

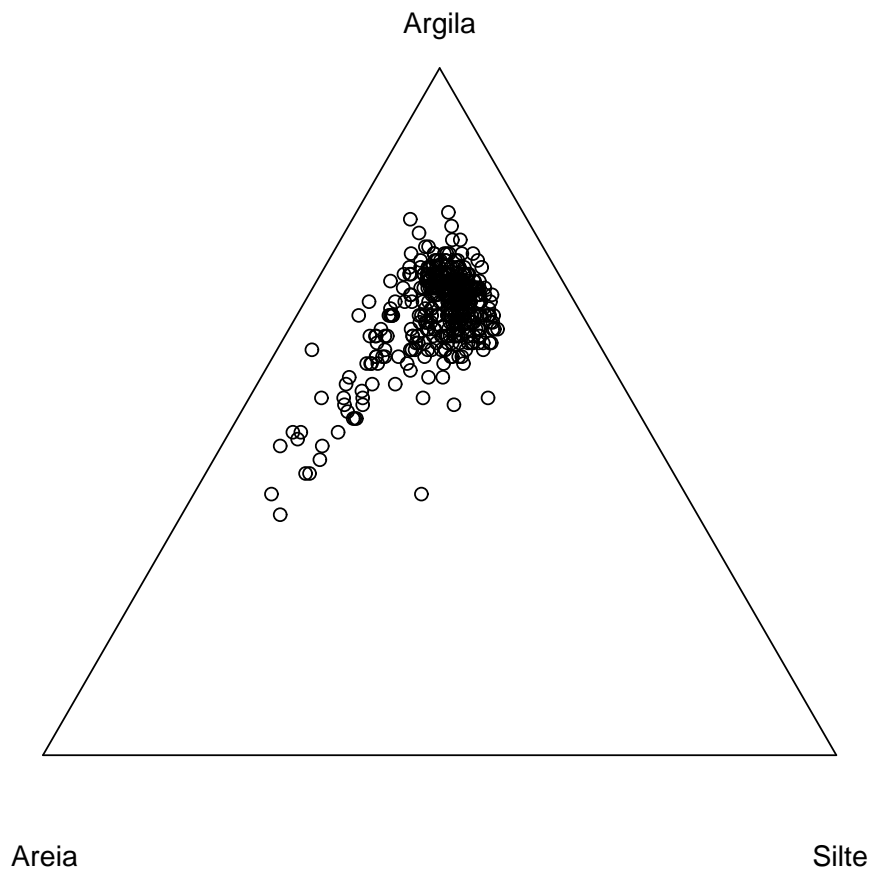


Figura 4: Diagrama Ternário para Areia, Silte e Argila.

```
> plot(comp)
> plot(mean(comp), add = T, pch = 20, col = "red")
> ellipses(mean(comp), var(comp), col = "red", r = 2)
> ellipses(mean(comp), var(comp), col = "red", r = 4)
```

O centro da distribuição é dado por

```
> mean(comp)

      Areia      Silte      Argila
0.1712660 0.1734996 0.6552344
attr(,"class")
[1] "acomp"
```

A variância na estrutura CLR do espaço euclidiano é

```
> var(comp)
```

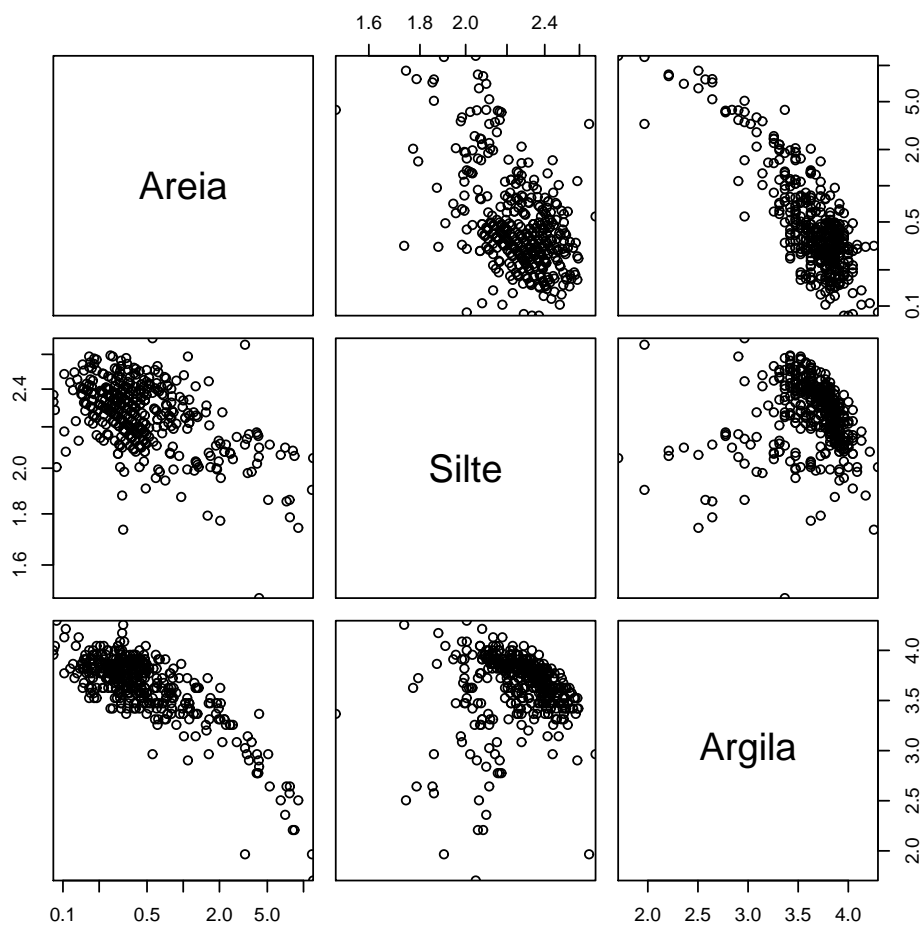


Figura 5: Areia \times Silte, Areia \times Argila e Silte \times Argila.

	Areia	Silte	Argila
Areia	0.08513246	-0.055311674	-0.029820790
Silte	-0.05531167	0.052191066	0.003120607
Argila	-0.02982079	0.003120607	0.026700182

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

```
> mvar(comp)
```

```
[1] 0.1640237
```

O desvio padrão métrico - $\sqrt{\text{mvar}/(D-1)}$ é dado por

```
> msd(comp)
```

```
[1] 0.2863771
```

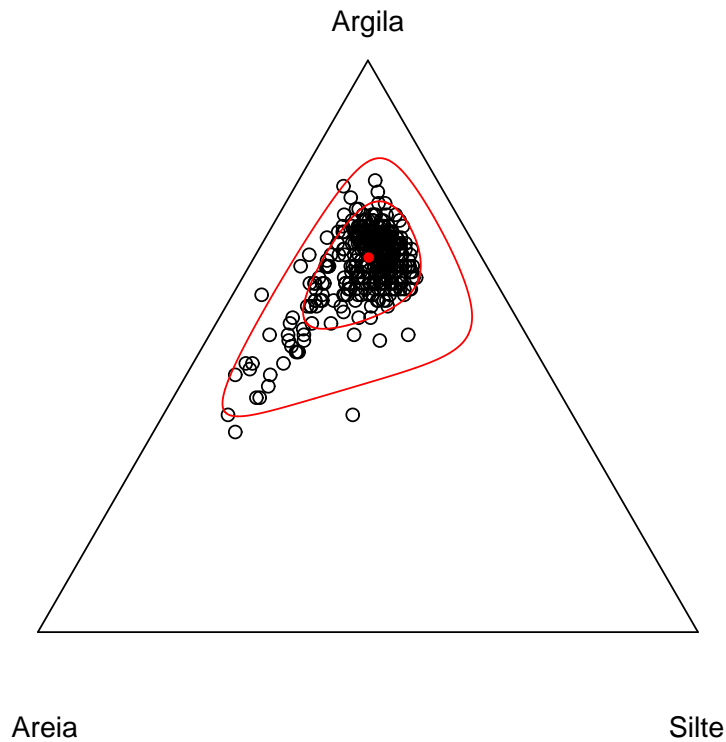


Figura 6: Diagrama Ternário com região de Confiança.

O desvio padrão clássico é obtido através de

```
> sd(comp)
```

```
      Areia      Silte      Argila
0.07166059 0.03996679 0.06312555
```

E a matriz variação, estrutura LR, cujos elementos são dados por $\text{var}(\log(x_i/x_j))$ é:

```
> variation(comp)
```

```
      Areia      Silte      Argila
Areia  0.0000000 0.24794688 0.17147422
Silte  0.2479469 0.00000000 0.07265003
Argila 0.1714742 0.07265003 0.00000000
```

Por fim, a variância total, a partir da variância na estrutura CLR é obtida. Para isto cria-se um vetor com 3 elementos iguais a 1 e o transposto deste

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

```
[1] 1 1 1
```

```
> xt <- t(x)
> xt
```

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

e, então:

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

```
      [,1]
[1,] 0.1640237
```