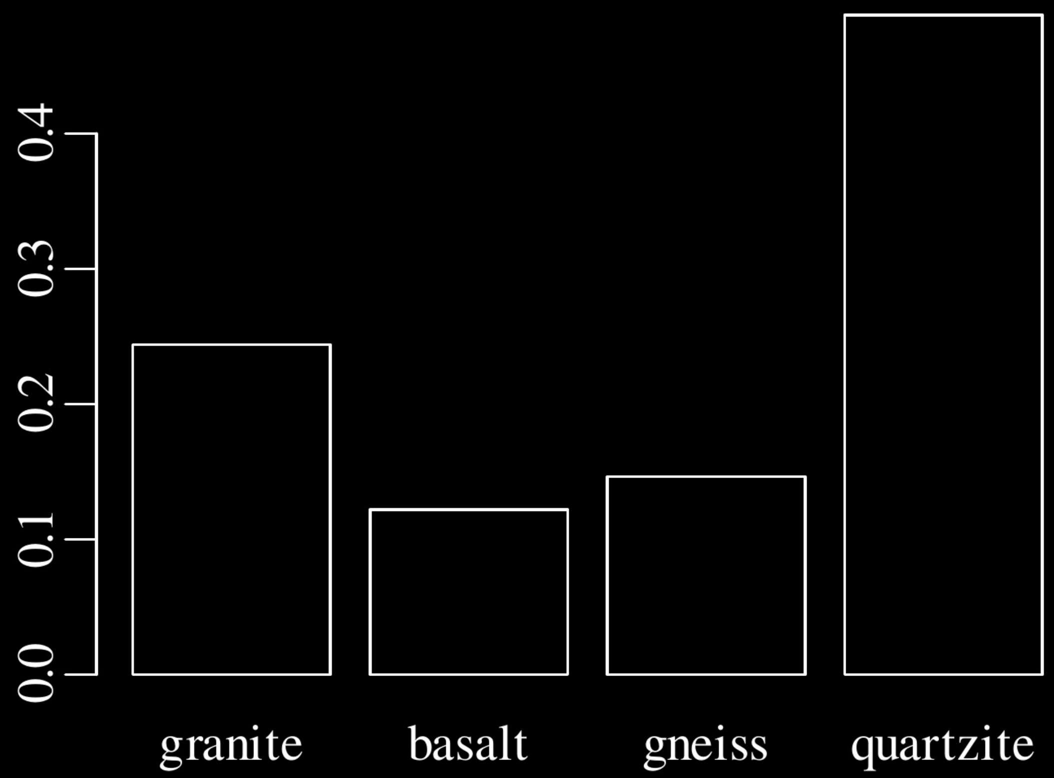
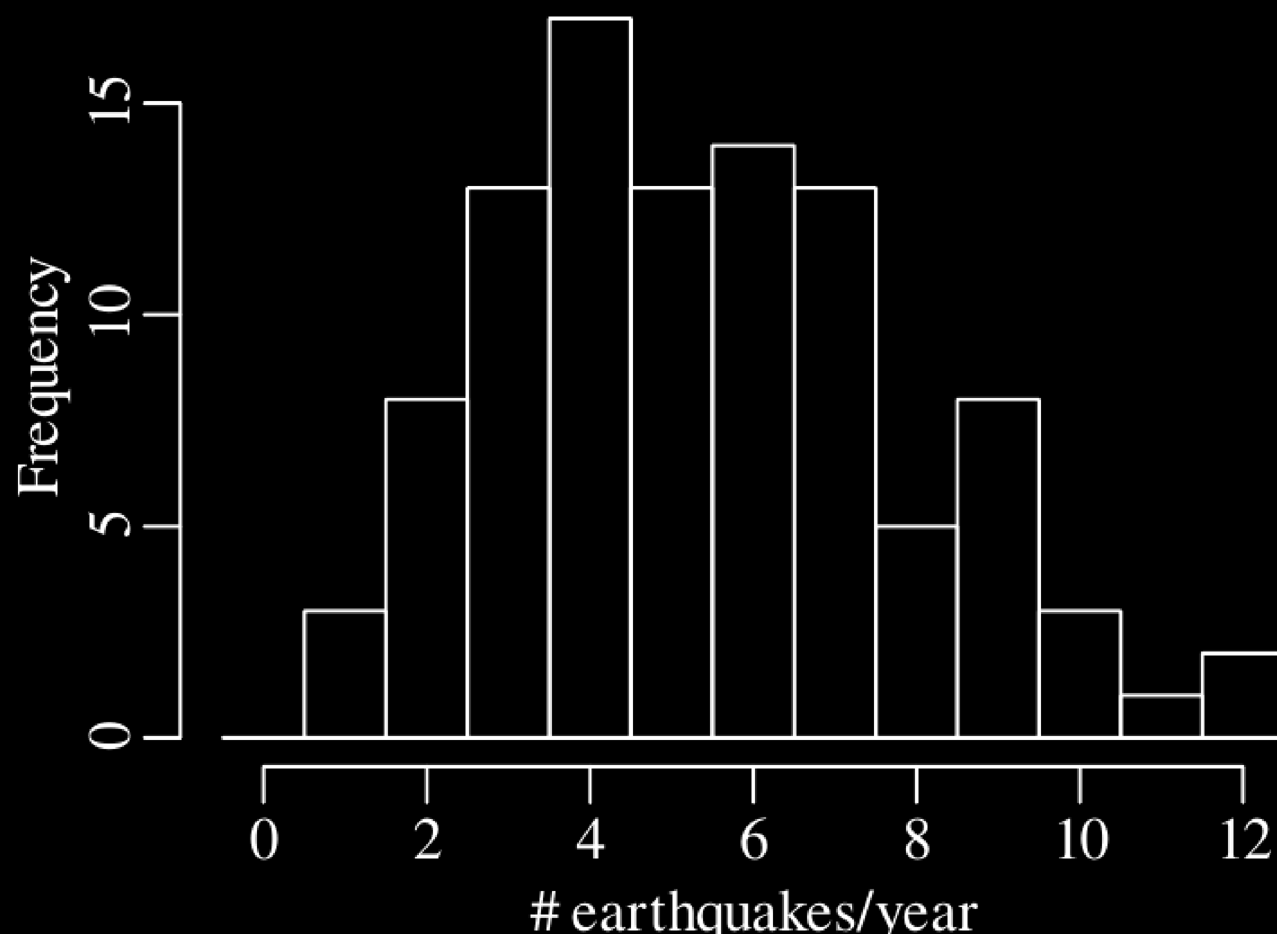
1. What is so ‘normal’ about the Gaussian distribution?

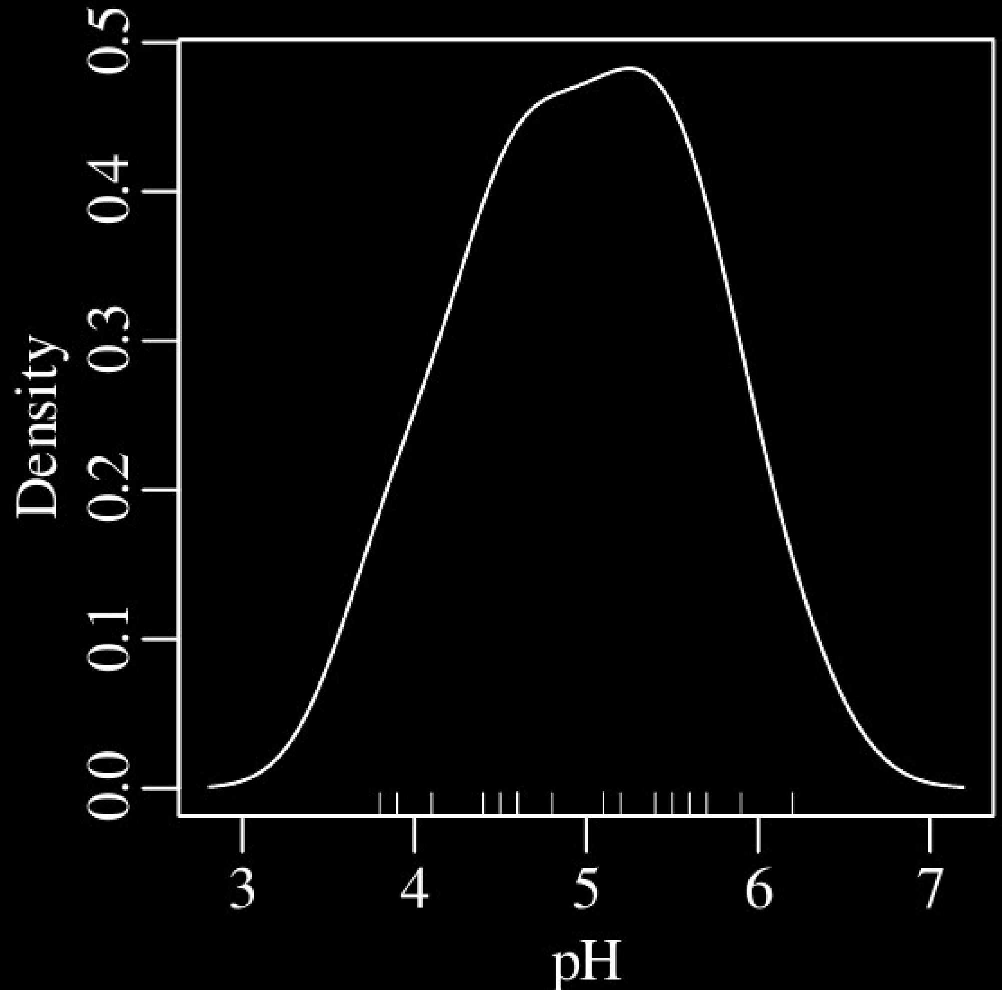
2. Is geology ‘normal’?

Pieter Vermeesch
p.vermeesch@ucl.ac.uk

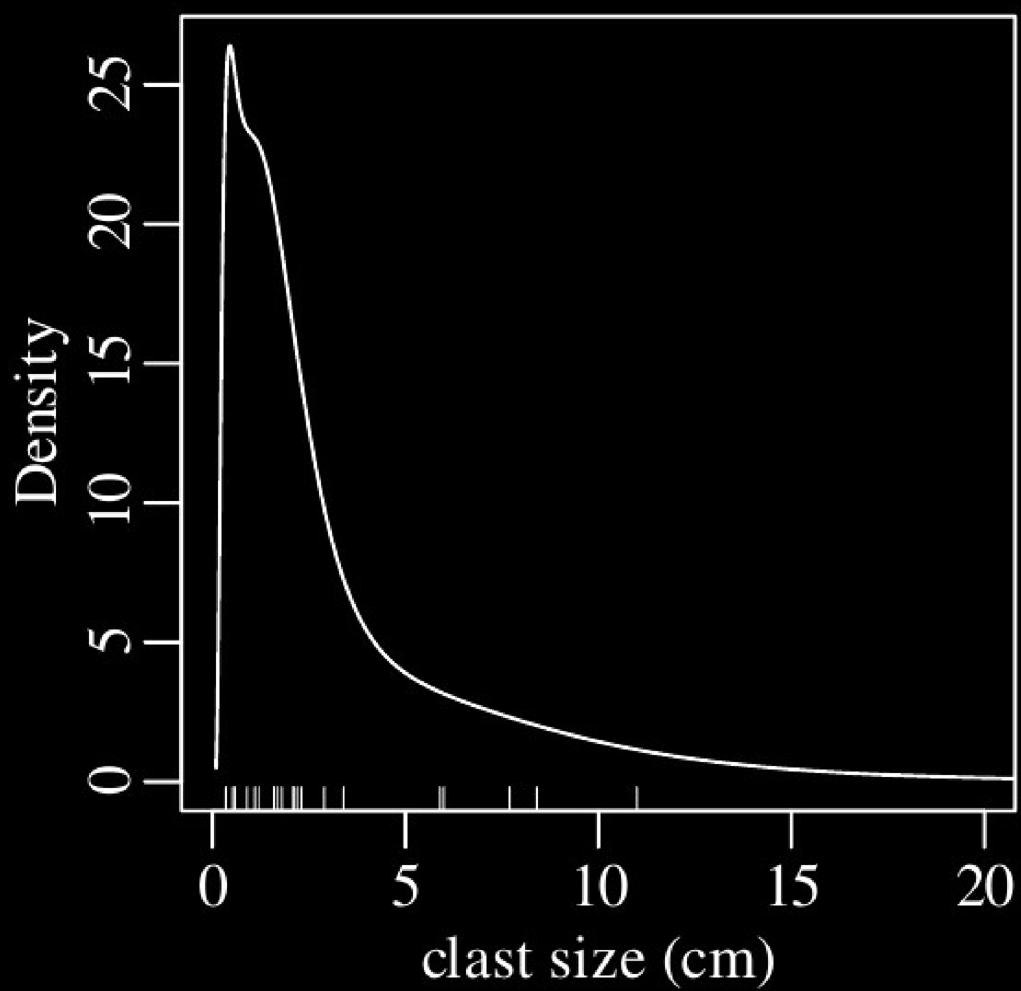


0.0
0.1
0.2
0.3
0.4
granite
basalt
gneiss
quartzite

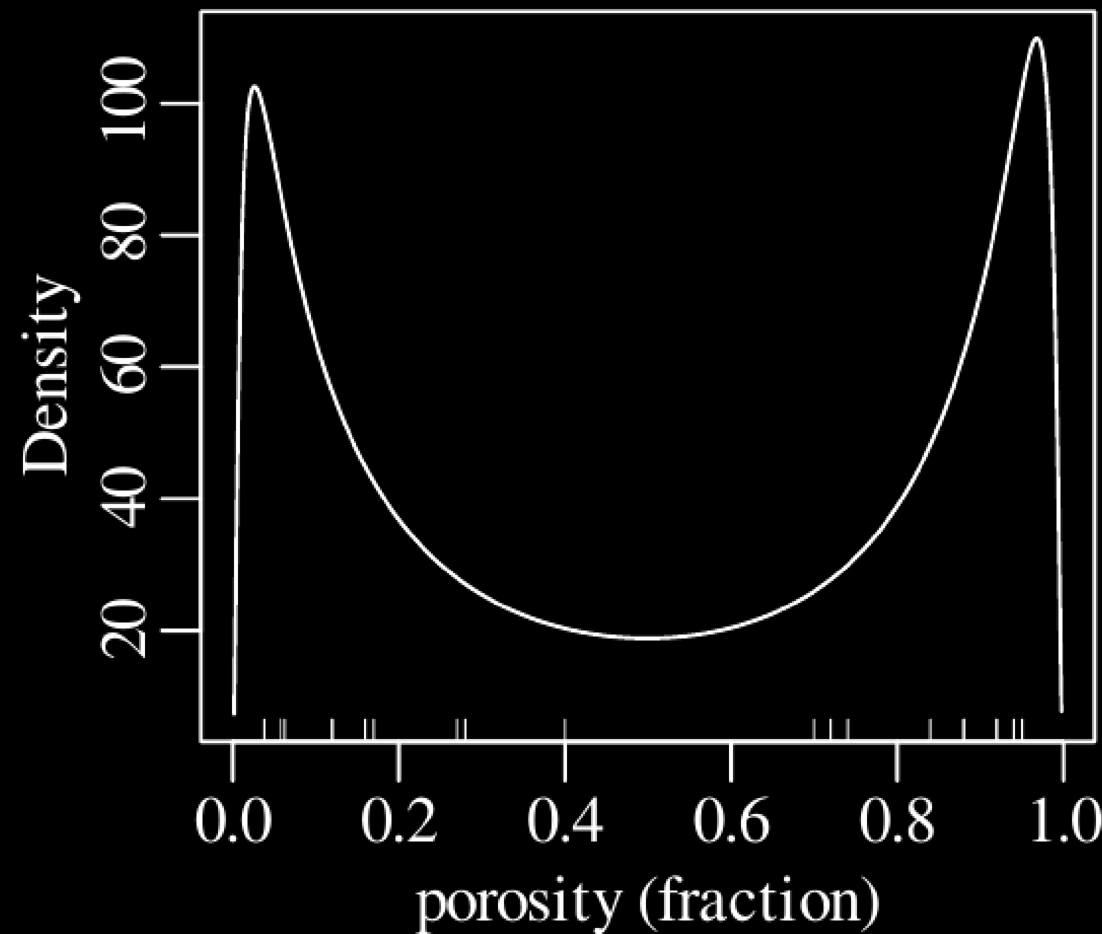

Frequency
0
5
10
15
0
2
4
6
8
10
12
earthquakes/year

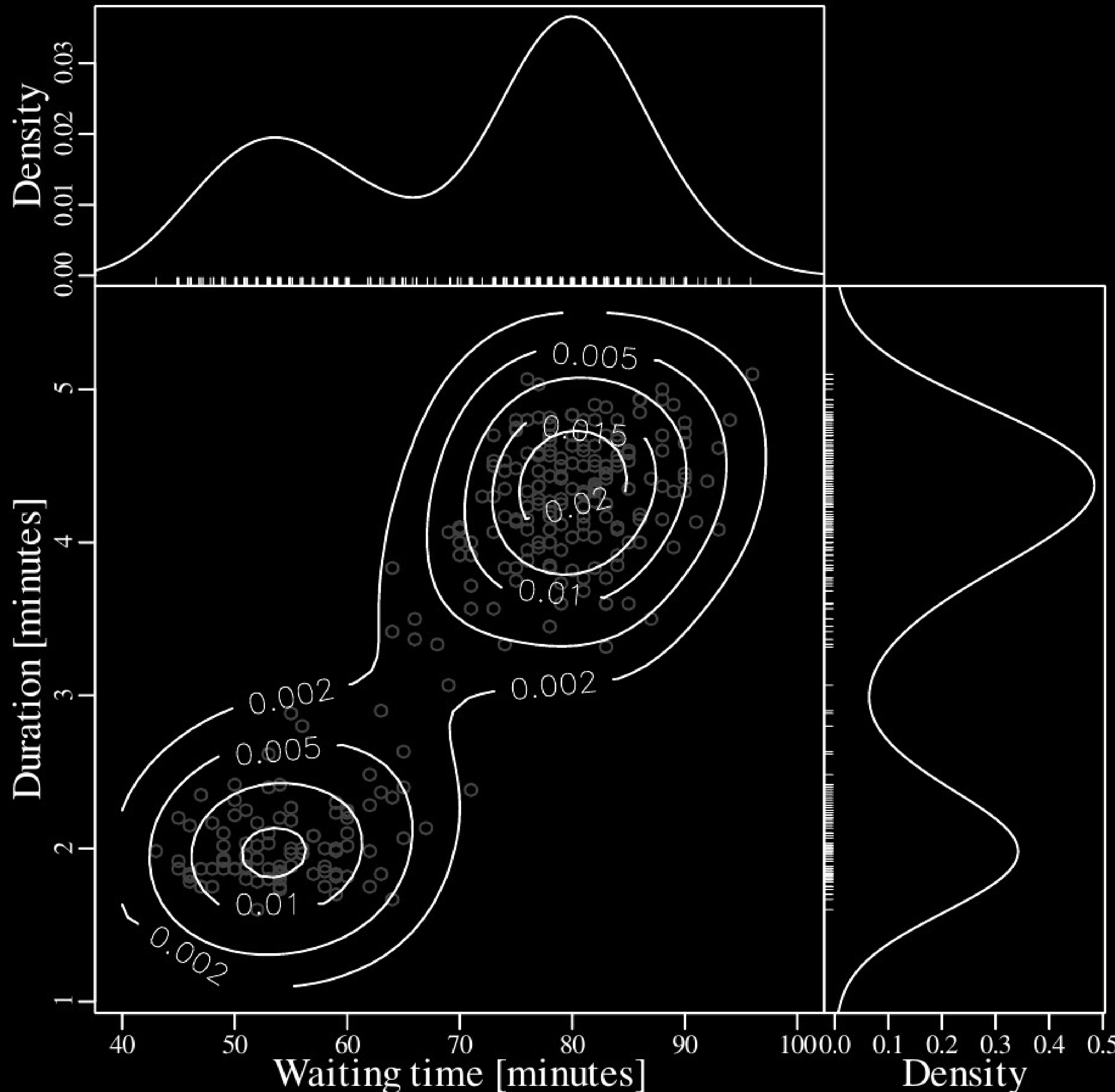


Density
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
7
pH

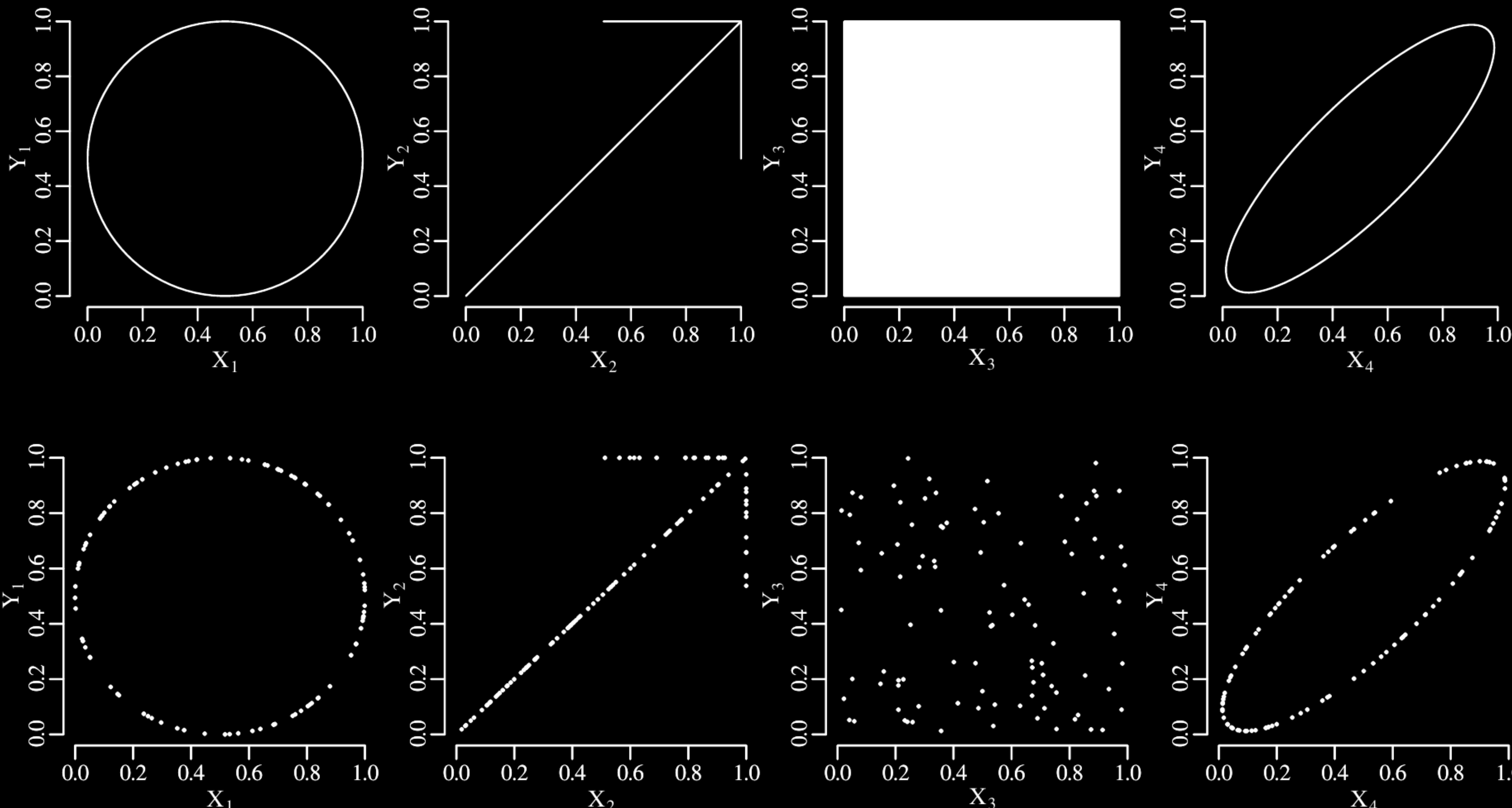


Density
0
5
10
15
20
25
0
5
10
15
20
clast size (cm)


Density
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0
porosity (fraction)



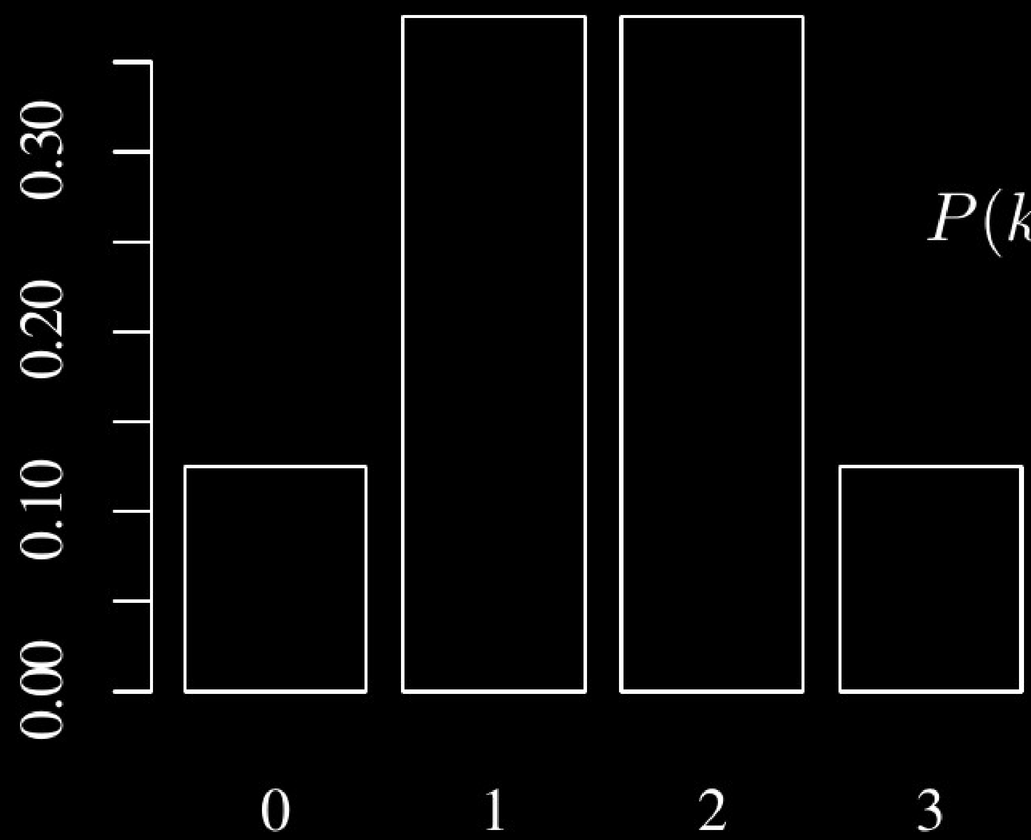
Density
0.00
0.01
0.02
0.03
Duration [minutes]
1
2
3
4
5
0.005
0.015
0.02
0.01
0.002
0.002
0.005
0.01
0.002
40
50
60
70
80
90
100
Waiting time [minutes]
0.0
0.1
0.2
0.3
0.4
0.5
Density



Y1
X1
Y2
X2
Y3
X3
Y4
X4
0.0
0.2
0.4
0.6
0.8
1.0

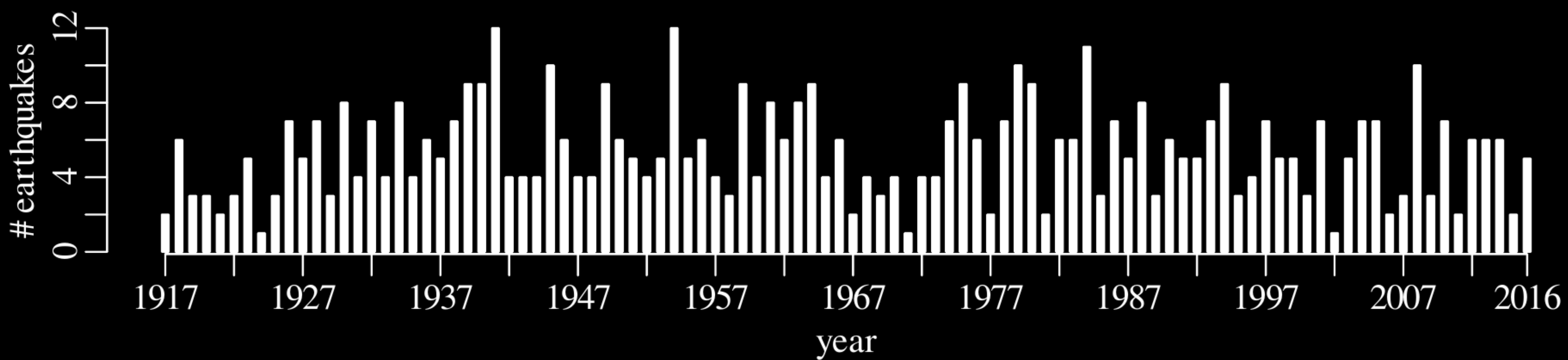
The binomial distribution

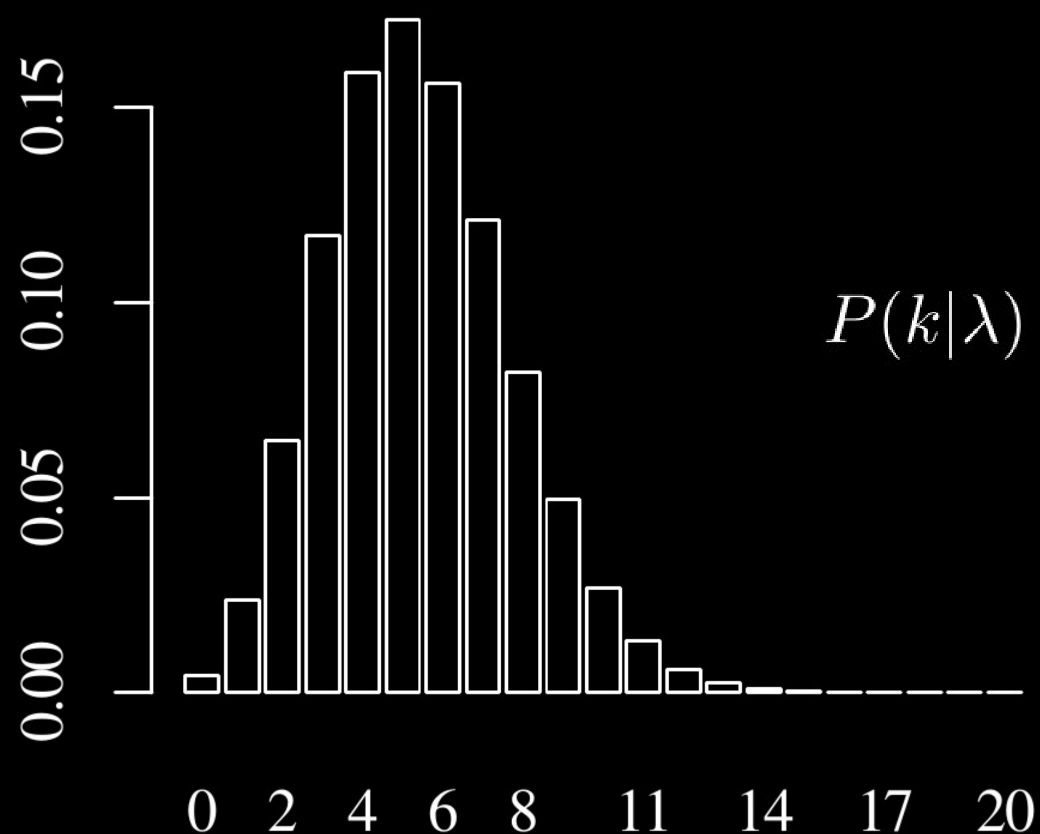
$$P(2 \times H \cap 1 \times T) = \frac{\{THH\}\{HTH\}\{HHT\}}{\{HHH\}\{THH\}\{HTH\}\{HHT\}\{TTH\}\{THT\}\{HTT\}\{TTT\}} = \frac{3}{8}$$



$$P(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

The Poisson distribution



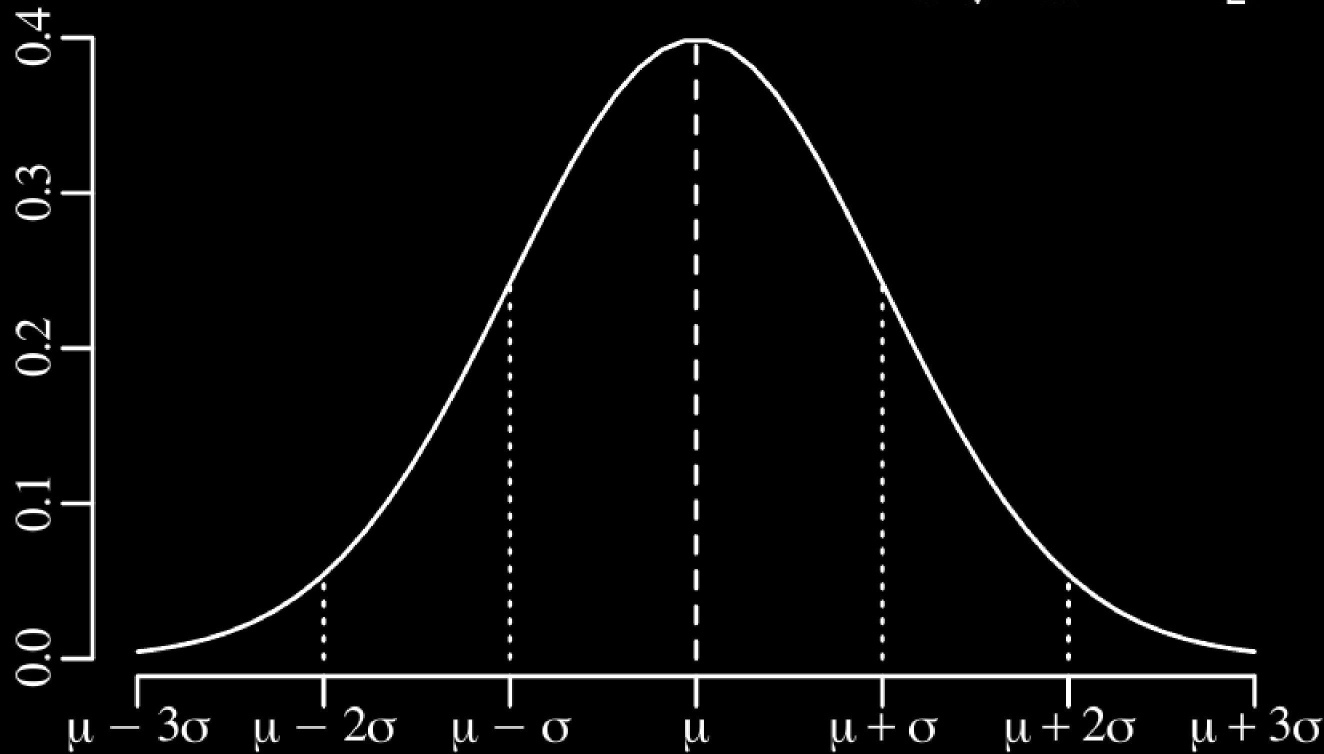


$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

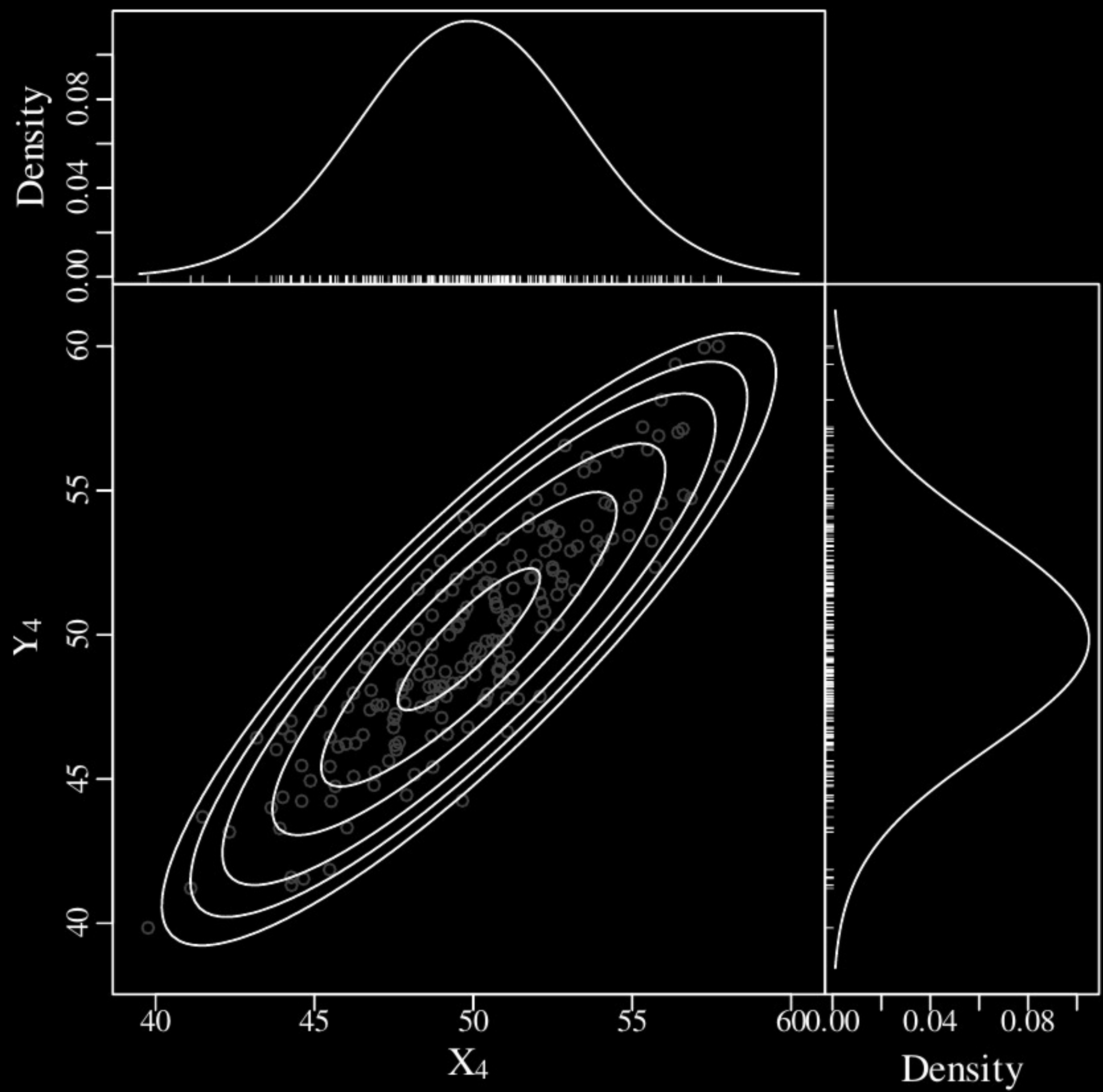
the Gaussian distribution

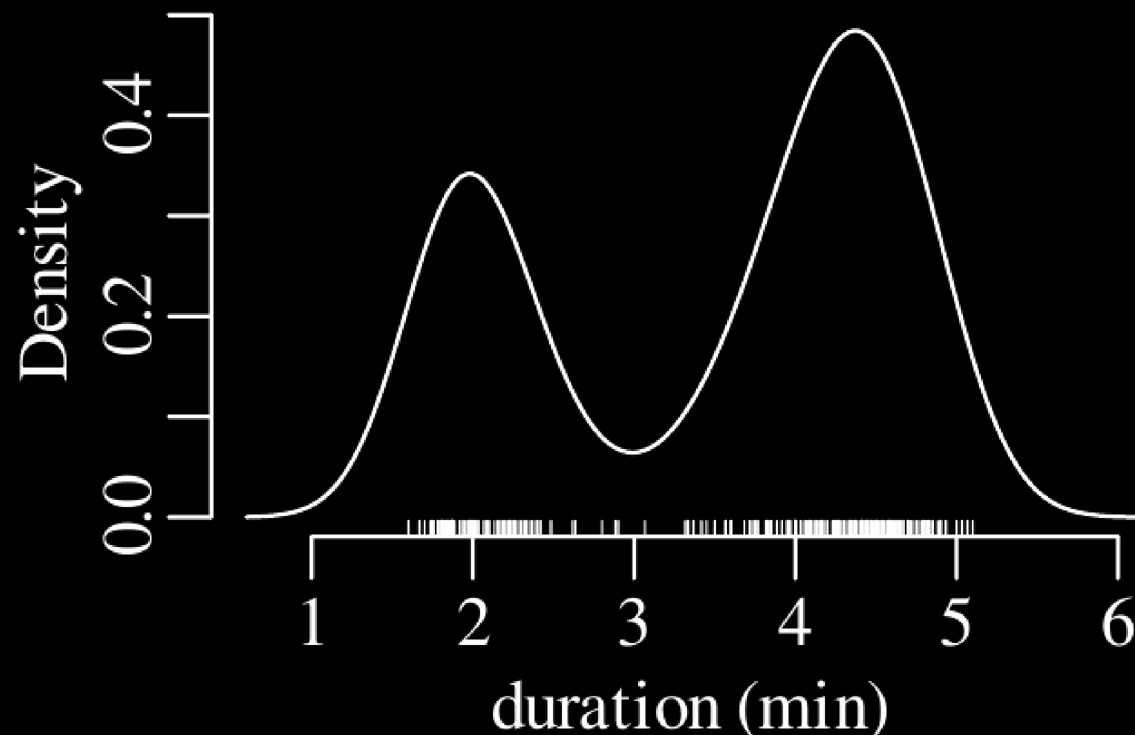
the **normal** distribution

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

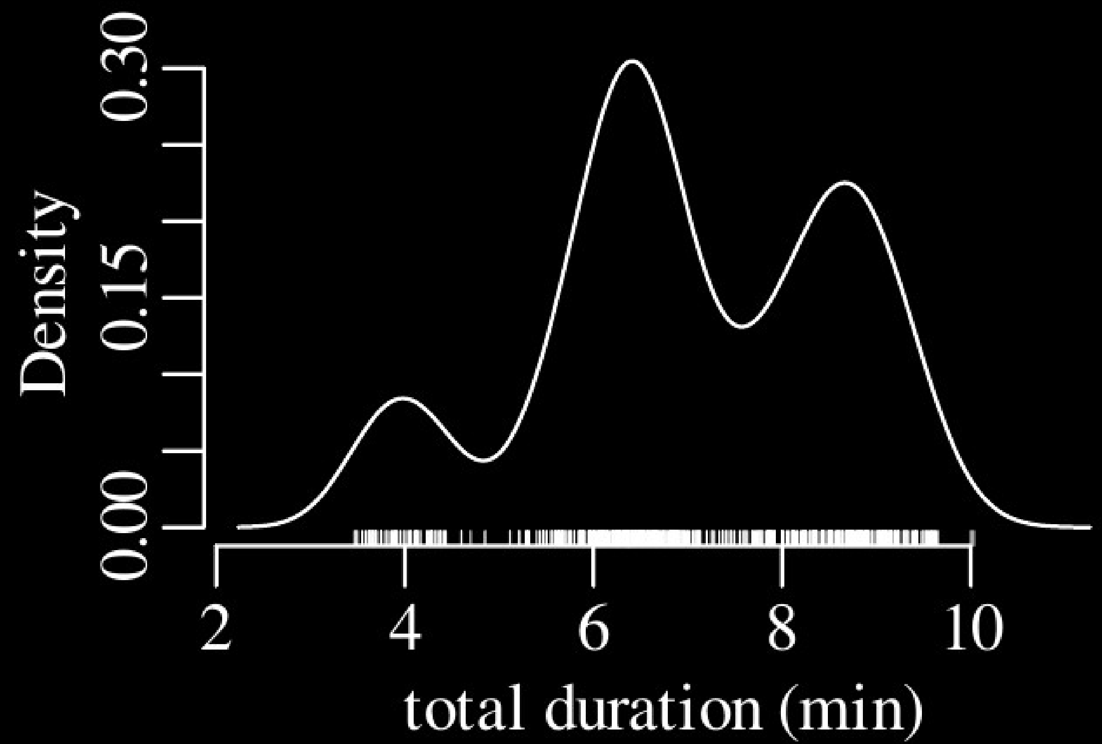


$$f(x, y|\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{x,y}) = \frac{\exp\left(-\begin{bmatrix}(x-\mu_x) & (y-\mu_y)\end{bmatrix}\begin{bmatrix}\sigma_x^2 & \sigma_{x,y} \\ \sigma_{x,y} & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}x-\mu_x \\ y-\mu_y\end{bmatrix}\Big/2\right)}{2\pi\sqrt{\begin{vmatrix}\sigma_x^2 & \sigma_{x,y} \\ \sigma_{x,y} & \sigma_y^2\end{vmatrix}}}$$

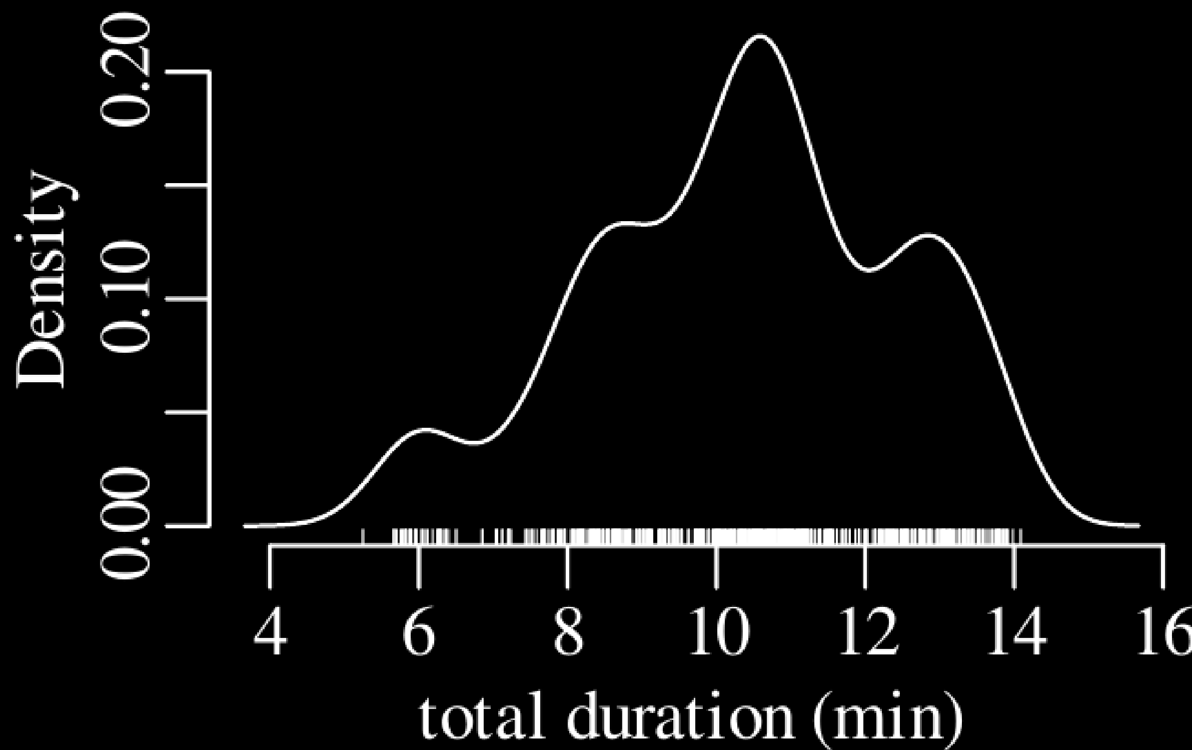




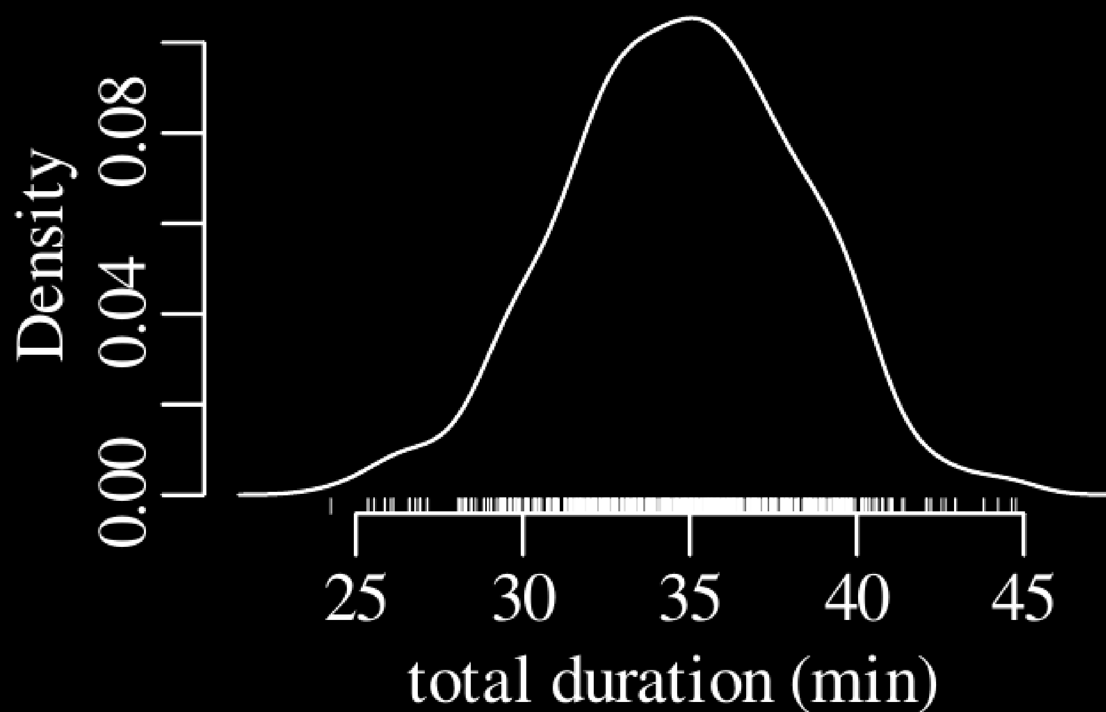
Density
0.0
0.2
0.4
1
2
3
4
5
6
duration (min)

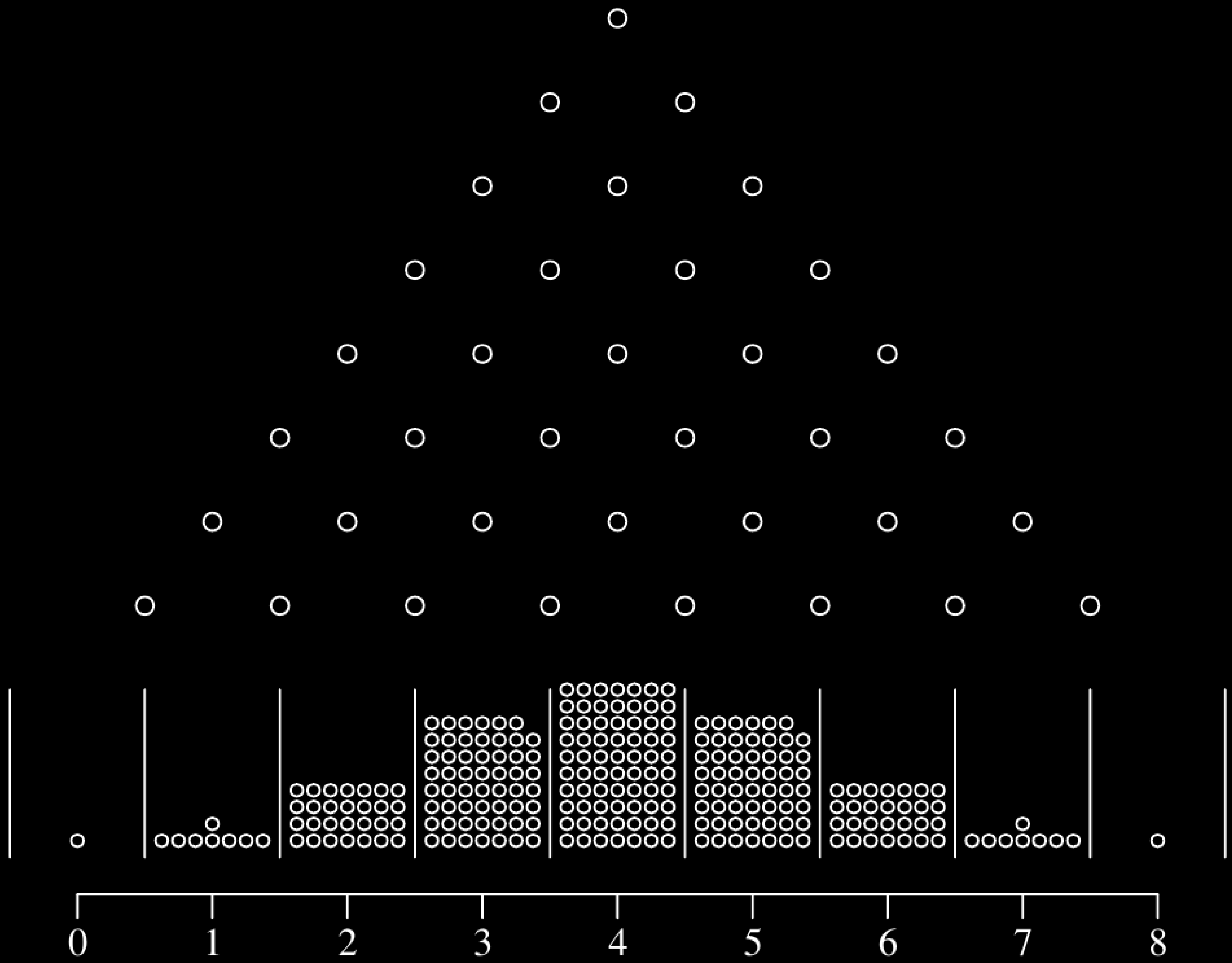


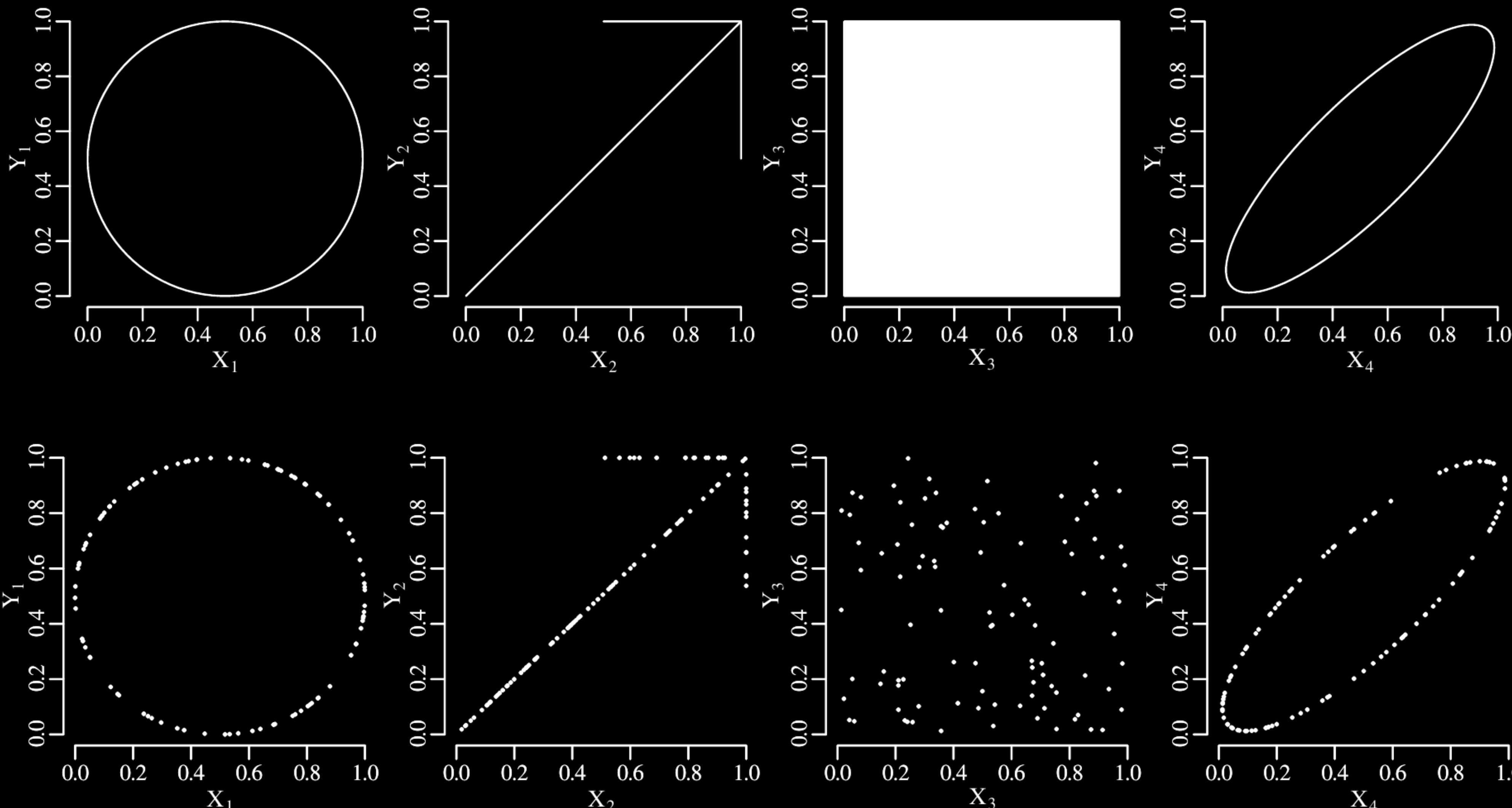
0.00
0.15
0.30
Density
2
4
6
8
10
total duration (min)



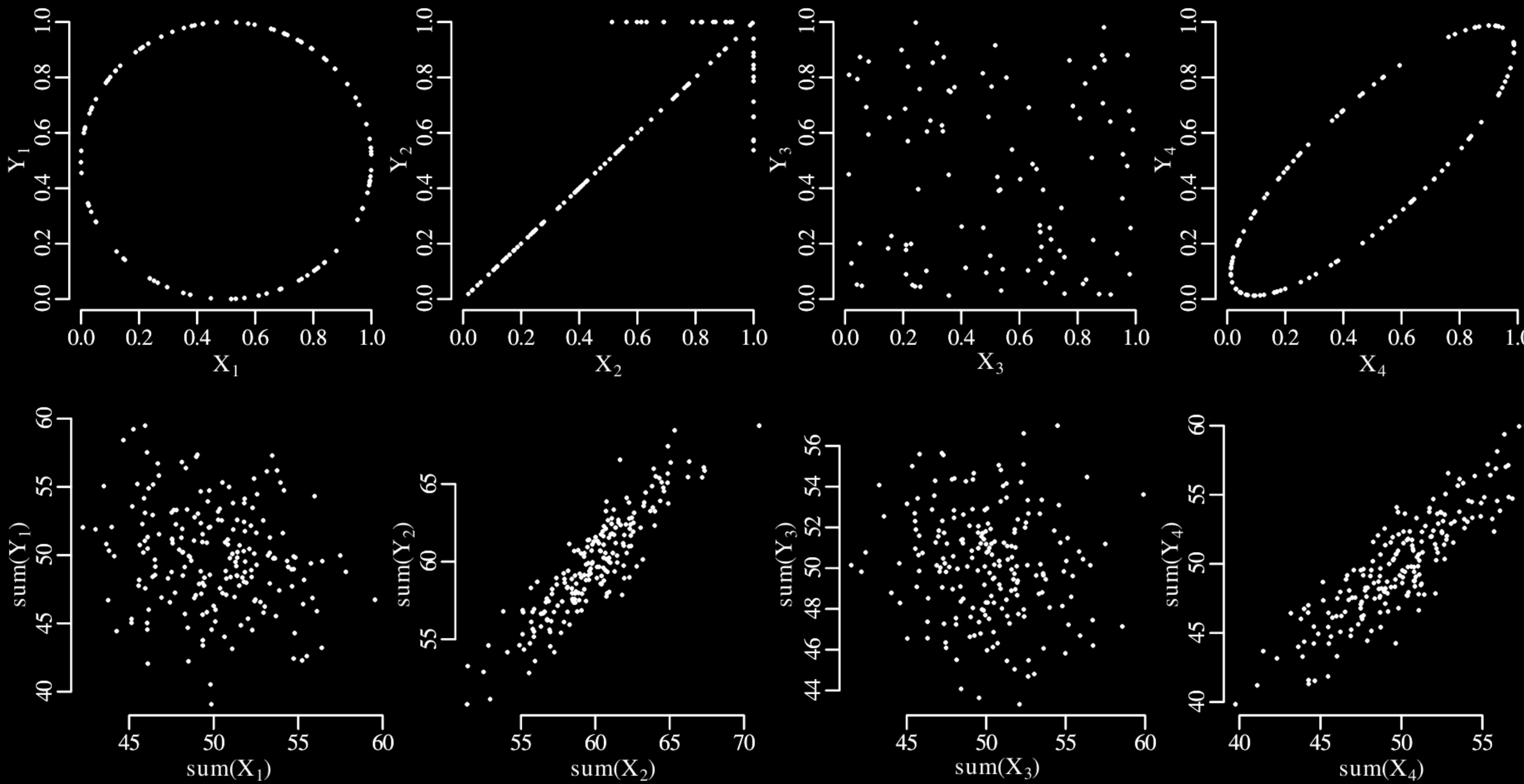
0.00
0.10
0.20
Density
4
6
8
10
12
14
16
total duration (min)


Density
0.00
0.04
0.08
25
30
35
40
45
total duration (min)


0
1
2
3
4
5
6
7
8



Y1
X1
Y2
X2
Y3
X3
Y4
X4
0.0
0.2
0.4
0.6
0.8
1.0
Y1
X1
Y2
X2
Y3
X3
Y4
X4
0.0
0.2
0.4
0.6
0.8
1.0



X1
Y1
X2
Y2
X3
Y3
X4
Y4
sum(X1)
sum(Y1)
sum(X2)
sum(Y2)
sum(X3)
sum(Y3)
sum(X4)
sum(Y4)

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\mathcal{L}(\mu,\sigma|x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} f(x_i|\mu,\sigma)$$

$$\begin{aligned}
\mathcal{LL}(\mu,\sigma|x_1,x_2,\ldots,x_n) &= \sum_{i=1}^{n} \ln\left[f(x_i|\mu,\sigma)\right] \\
&= \sum_{i=1}^{n} -\ln[\sigma] - \frac{1}{2}\ln[2\pi] - \frac{(x_i-\mu)^2}{2\sigma^2}
\end{aligned}$$

$$\frac{\partial \mathcal{LL}}{\partial \mu} = -\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} = 0$$

$$\Rightarrow n\mu - \sum_{i=1}^{n} x_i = 0$$

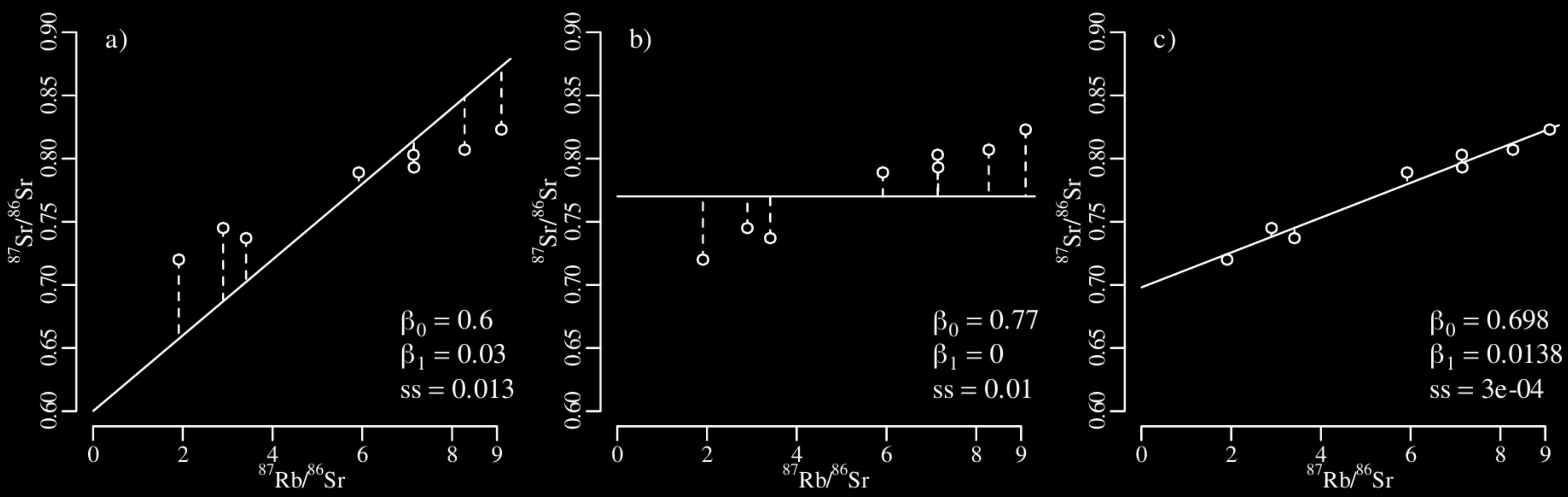
$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \mathcal{LL}}{\partial \sigma} = \sum_{i=1}^{n} -\frac{1}{\sigma} + \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = \frac{n}{\sigma}$$

$$\Rightarrow \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

$$ss \equiv \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i - y_i)^2$$

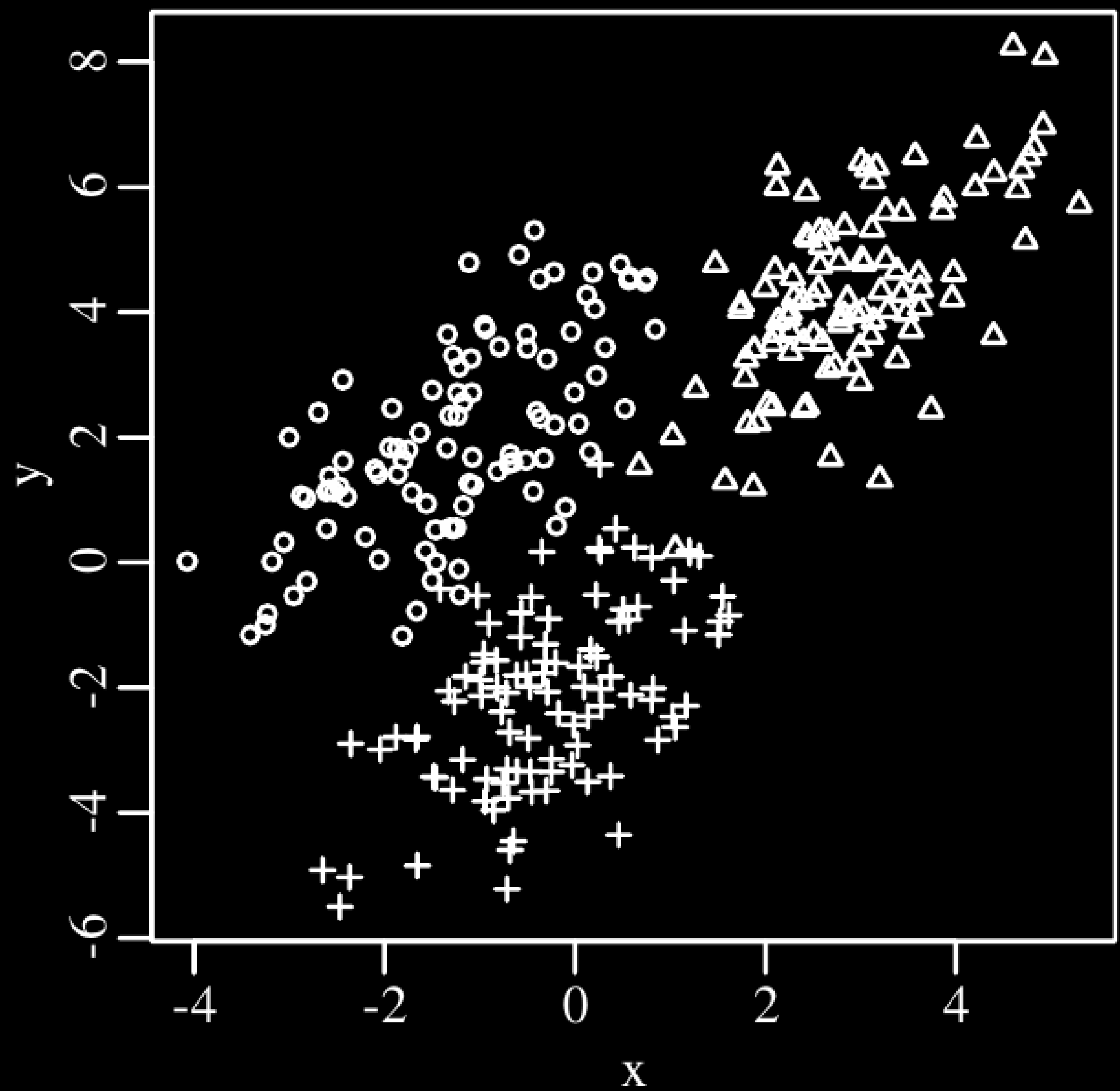


$$y_i = \beta_0 + \beta_1 \ x_i + \epsilon_i$$

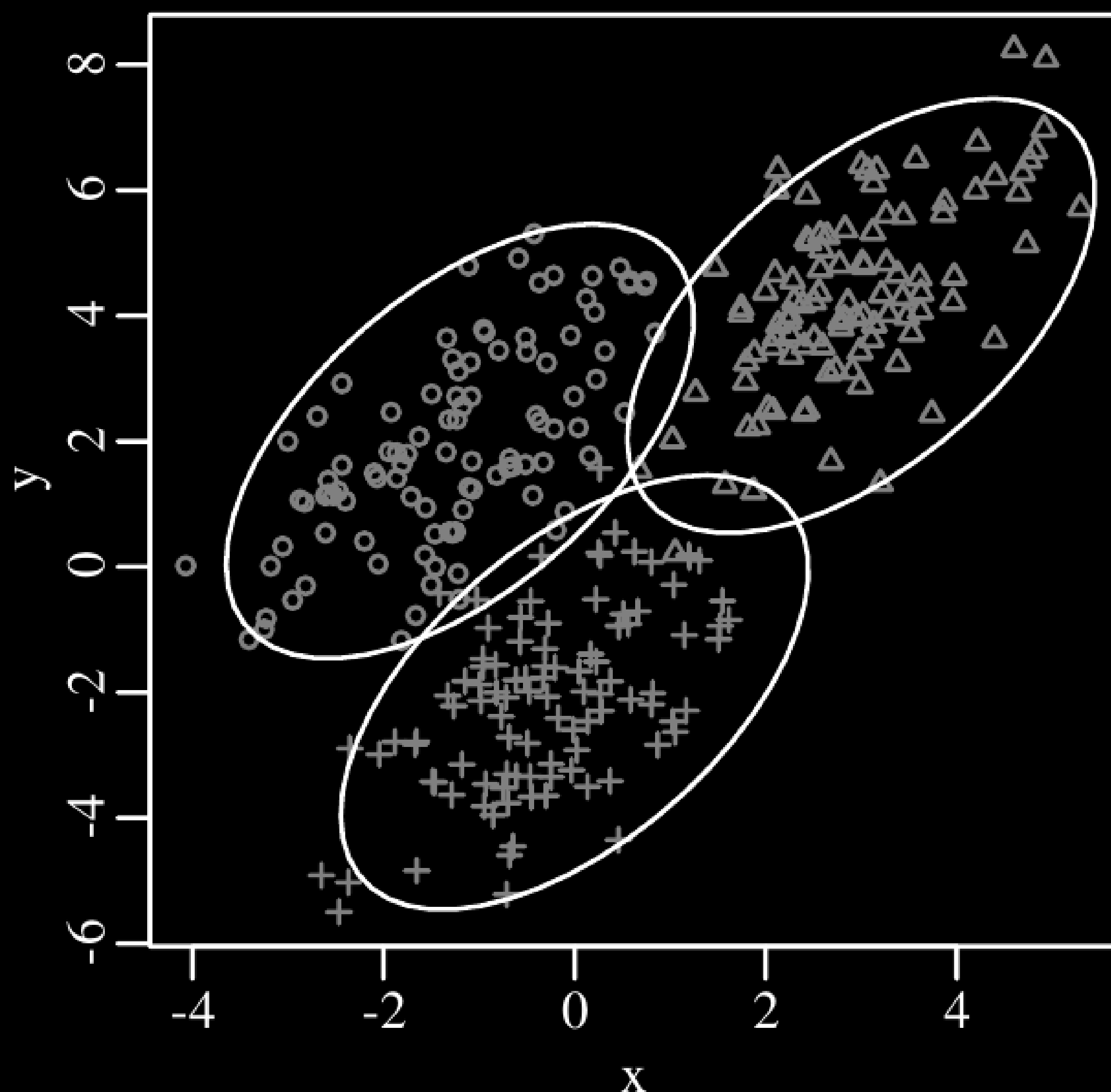
$$f(\epsilon_i|\beta_0, \beta_1, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\epsilon_i^2}{2\sigma^2}\right], \text{ where } \epsilon_i = \beta_0 + \beta_1 x_i - y_i$$

$$\mathcal{L}(\beta_0, \beta_1, \sigma|\epsilon_1, \ldots, \epsilon_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\epsilon_i^2}{2\sigma^2}\right], \text{ where } \epsilon_i = \beta_0 + \beta_1 x_i - y_i$$

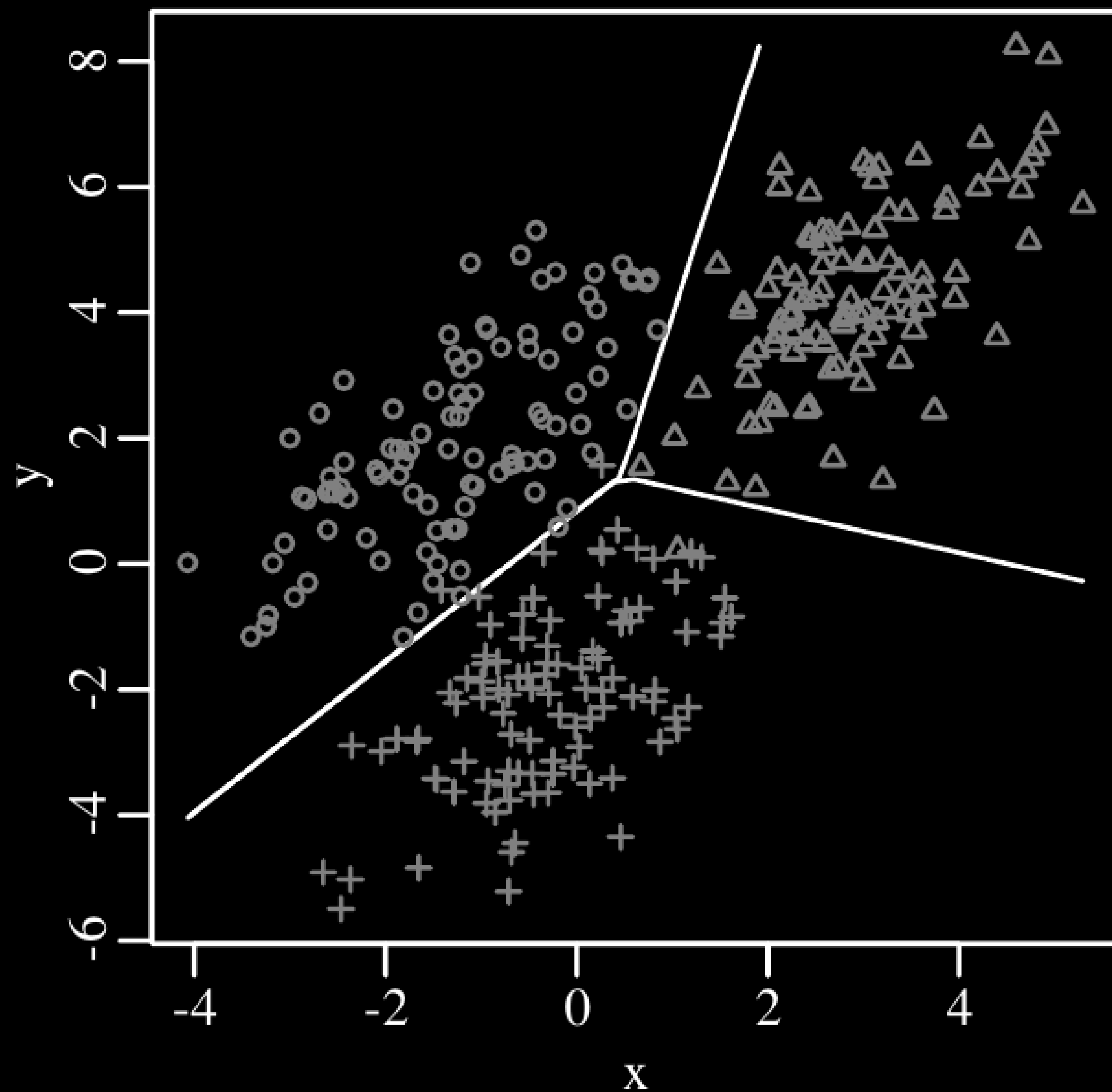
$$\begin{aligned}\max_{\beta_0,\beta_1}\left[\prod_{i=1}^{n}\mathcal{L}\right] &= \max_{\beta_0,\beta_1}\left[\sum_{i=1}^{n}\ln\mathcal{L}\right] = \max_{\beta_0,\beta_1}\left[\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n}\left(\frac{\epsilon_i^2}{2\sigma^2}\right)\right] \\ &= \max_{\beta_0,\beta_1}\left[-\sum_{i=1}^{n}\epsilon_i^2\right] = \min_{\beta_0,\beta_1}\left[\sum_{i=1}^{n}\epsilon_i^2\right] = \min_{\beta_0,\beta_1}(ss)\end{aligned}$$



y
8
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
x

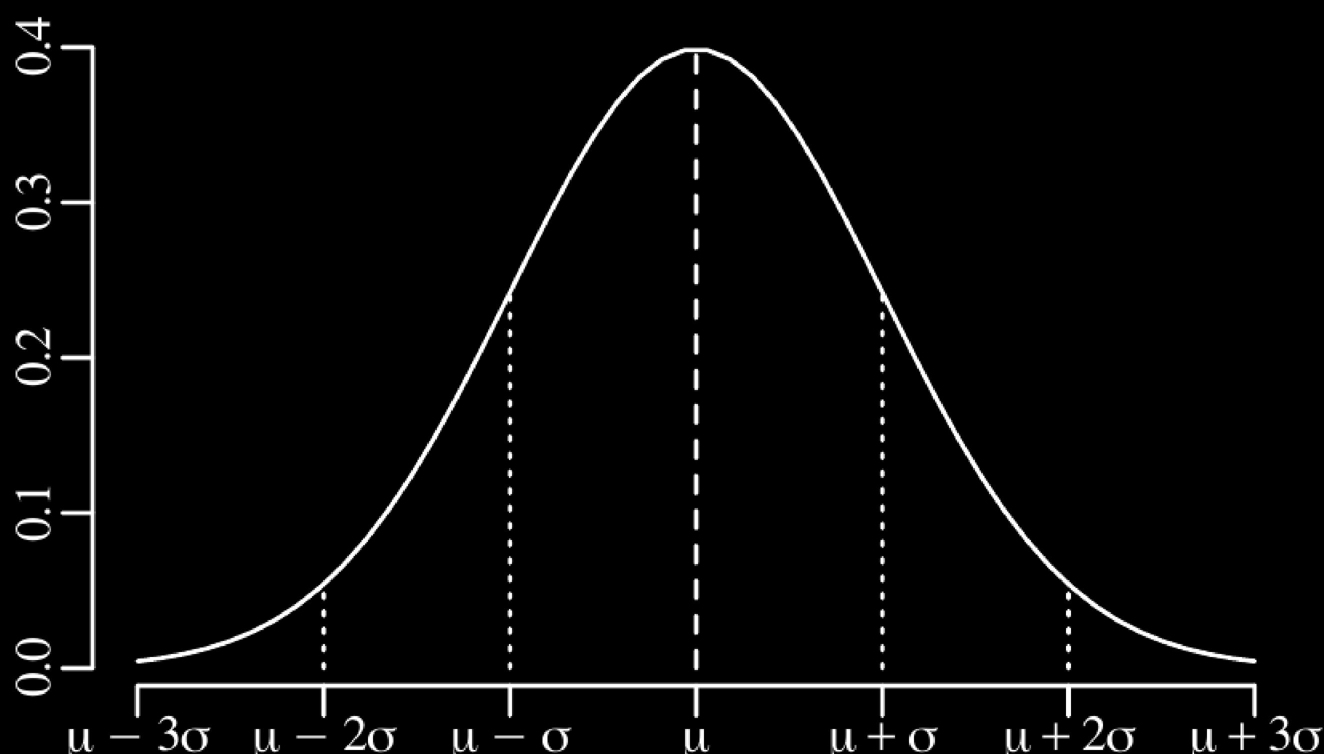


8
6
4
2
0
-2
-4
-6
y
-4
-2
0
2
4
x



8
6
4
2
0
-2
-4
-6
y
-4
-2
0
2
4
x

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$



	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10
A/B	1.30	2.10	0.83	2.40	0.26	1.80	1.30	0.67	1.70	0.079
B/A	0.78	0.47	1.20	0.42	3.80	0.55	0.76	1.50	0.60	13.0

	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10
A/B	1.30	2.10	0.83	2.40	0.26	1.80	1.30	0.67	1.70	0.079
B/A	0.78	0.47	1.20	0.42	3.80	0.55	0.76	1.50	0.60	13.0
$1/(A/B)$	0.78	0.47	1.20	0.42	3.80	0.55	0.76	1.50	0.60	13.0

	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10	mean
A/B	1.30	2.10	0.83	2.40	0.26	1.80	1.30	0.67	1.70	0.079	1.20
B/A	0.78	0.47	1.20	0.42	3.80	0.55	0.76	1.50	0.60	13.0	2.30
$1/(A/B)$	0.78	0.47	1.20	0.42	3.80	0.55	0.76	1.50	0.60	13.0	

$$\frac{1}{\overline{A/B}} = \frac{1}{1.20} = 0.81 \neq 2.30 = \overline{B/A}$$

$$\text{and } \frac{1}{\overline{B/A}} = \frac{1}{2.30} = 0.44 \neq 1.20 = \overline{A/B}$$

	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10	mean
$\ln[A/B]$	0.25	0.75	-0.18	0.86	-1.30	0.59	0.27	-0.41	0.50	-2.50	-0.12
$\ln[B/A]$	-0.25	-0.75	0.18	-0.86	1.30	-0.59	-0.27	0.41	-0.50	2.50	0.12

	mean	exp[mean]
$\ln[A/B]$	-0.12	0.88
$\ln[B/A]$	0.12	1.13

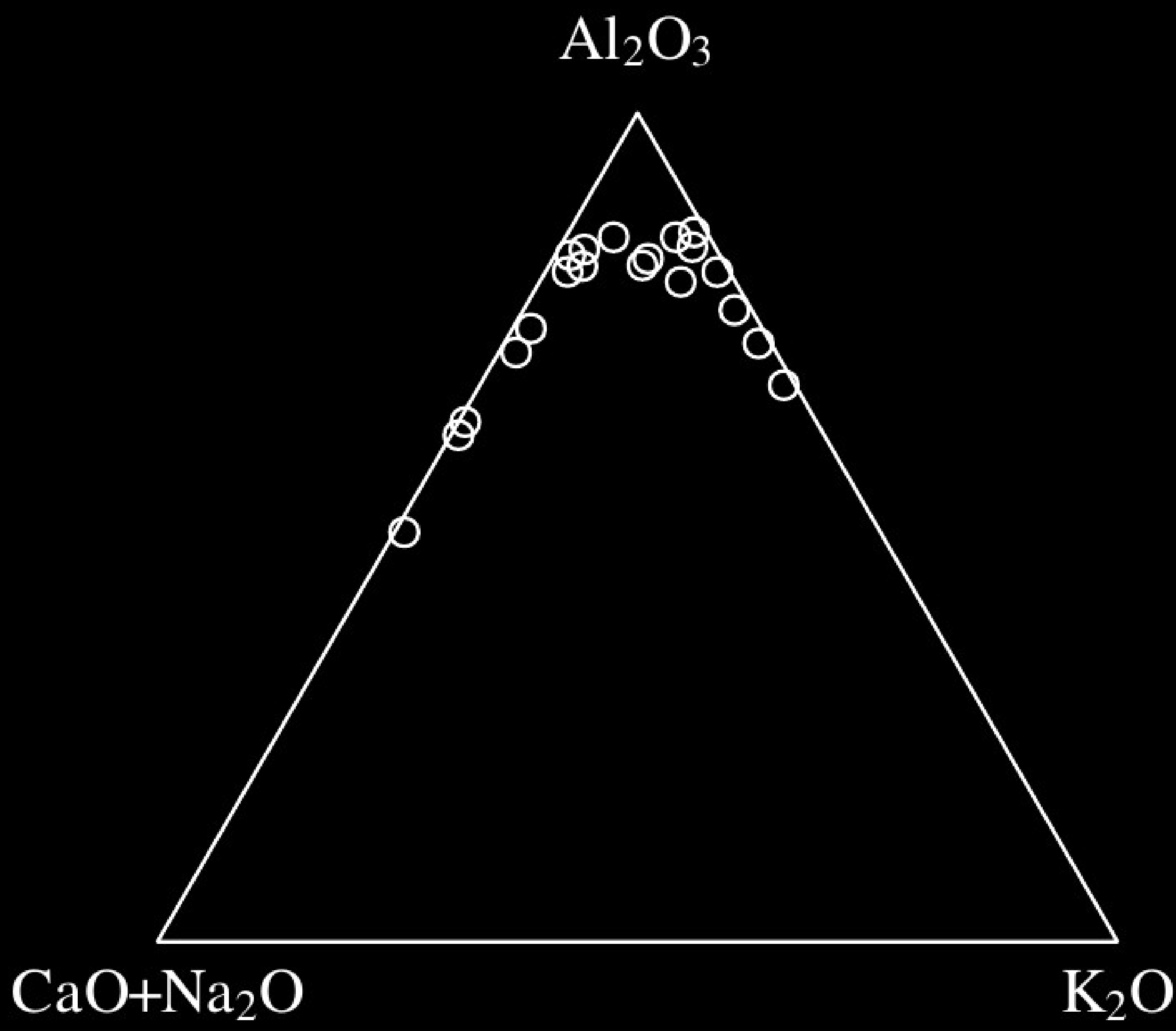
	1	2	3	4	5	6	7	8	9	10
A	0.27	0.37	0.57	0.91	0.20	0.90	0.94	0.66	0.63	0.062
B	0.21	0.18	0.69	0.38	0.77	0.50	0.72	0.99	0.38	0.780

	1	2	3	4	5	6	7	8	9	10	mean
$\ln[A/B]$	0.25	0.75	-0.18	0.86	-1.30	0.59	0.27	-0.41	0.50	-2.50	-0.12
$\ln[B/A]$	-0.25	-0.75	0.18	-0.86	1.30	-0.59	-0.27	0.41	-0.50	2.50	0.12

	mean	exp[mean]
$\ln[A/B]$	-0.12	0.88
$\ln[B/A]$	0.12	1.13

$$\frac{1}{g(A/B)} = \frac{1}{0.88} = 1.13 = g(B/A)$$

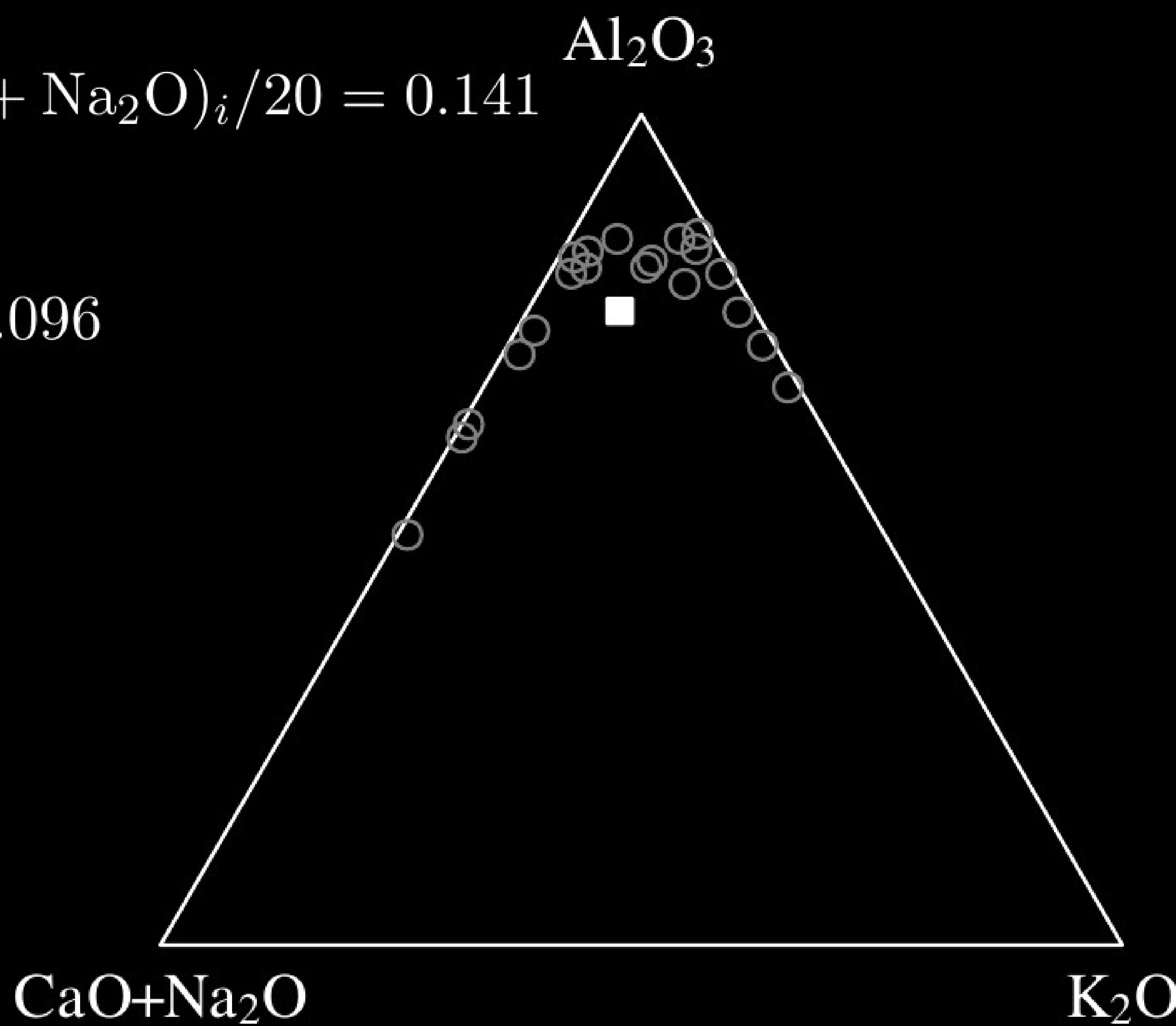
$$\text{and } \frac{1}{g(B/A)} = \frac{1}{1.13} = 0.88 = g(A/B)$$


Al_2O_3
$CaO+Na_2O$
K_2O

$$\overline{\mathrm{Al_2O_3}} = \sum_{i=1}^{20} (\mathrm{Al_2O_3})_i/20 = 0.763$$

$$\overline{\mathrm{CaO + Na_2O}} = \sum_{i=1}^{20} (\mathrm{CaO_2 + Na_2O})_i/20 = 0.141$$

$$\overline{\mathrm{K_2O}} = \sum_{i=1}^{20} (\mathrm{K_2O})_i/20 = 0.096$$



$$s[\mathrm{Al_2O_3}] = \sqrt{\sum_{i=1}^{20} \frac{((\mathrm{Al_2O_3})_i - 0.763)^2}{19}} = 0.0975$$

$$s[\mathrm{CaO} + \mathrm{Na_2O}] = \sqrt{\sum_{i=1}^{20} \frac{((\mathrm{CaO_2} + \mathrm{Na_2O})_i - 0.141)^2}{19}} = 0.142$$

$$s[\mathrm{K_2O}] = \sqrt{\sum_{i=1}^{20} \frac{((\mathrm{K_2O})_i - 0.096)^2}{19}} = 0.0926$$

$$\mathrm{Al_2O_3} : 0.763 \pm 0.195$$

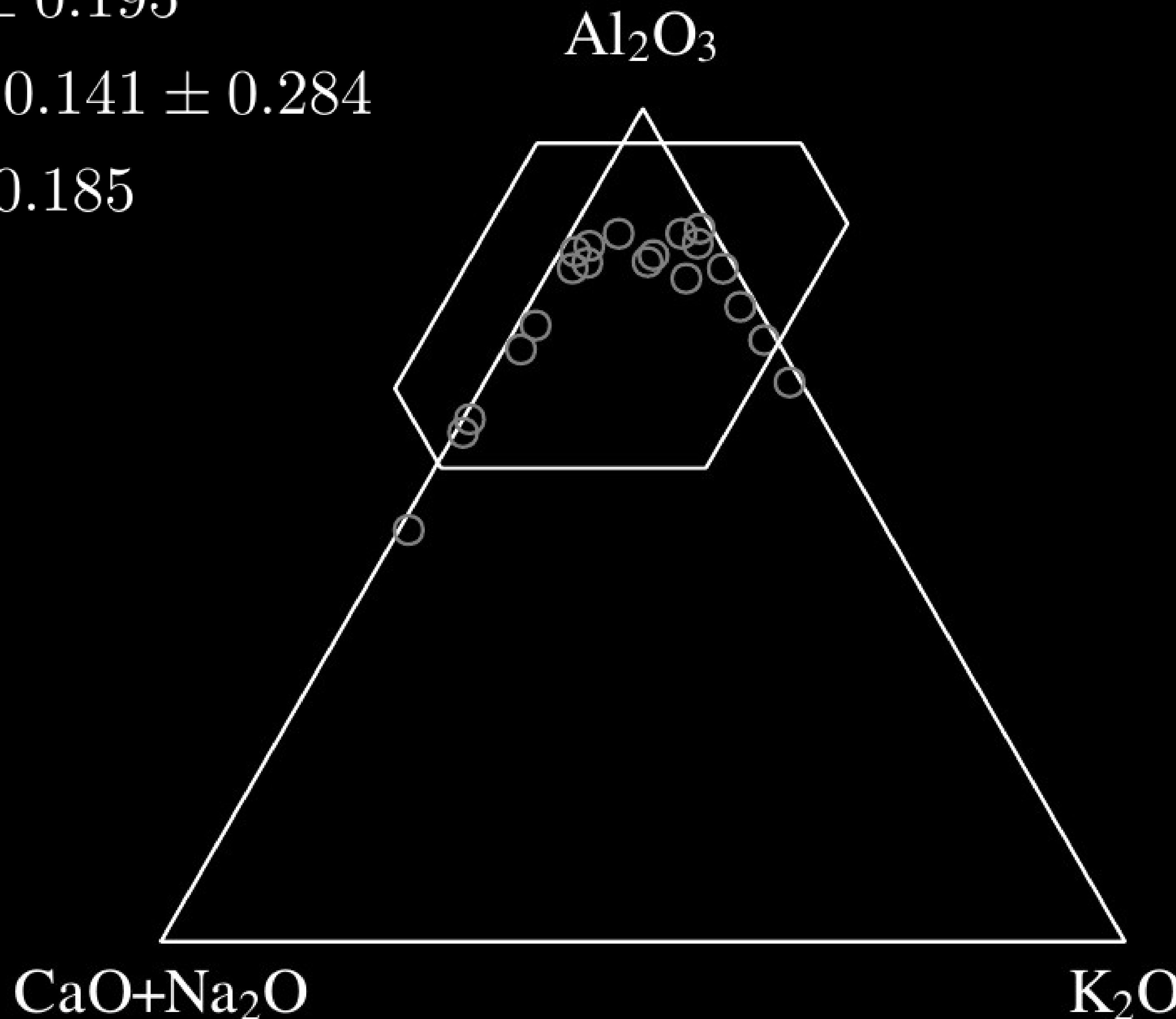
$$\mathrm{CaO} + \mathrm{Na_2O} : 0.141 \pm 0.284$$

$$\mathrm{K_2O} : 0.096 \pm 0.185$$

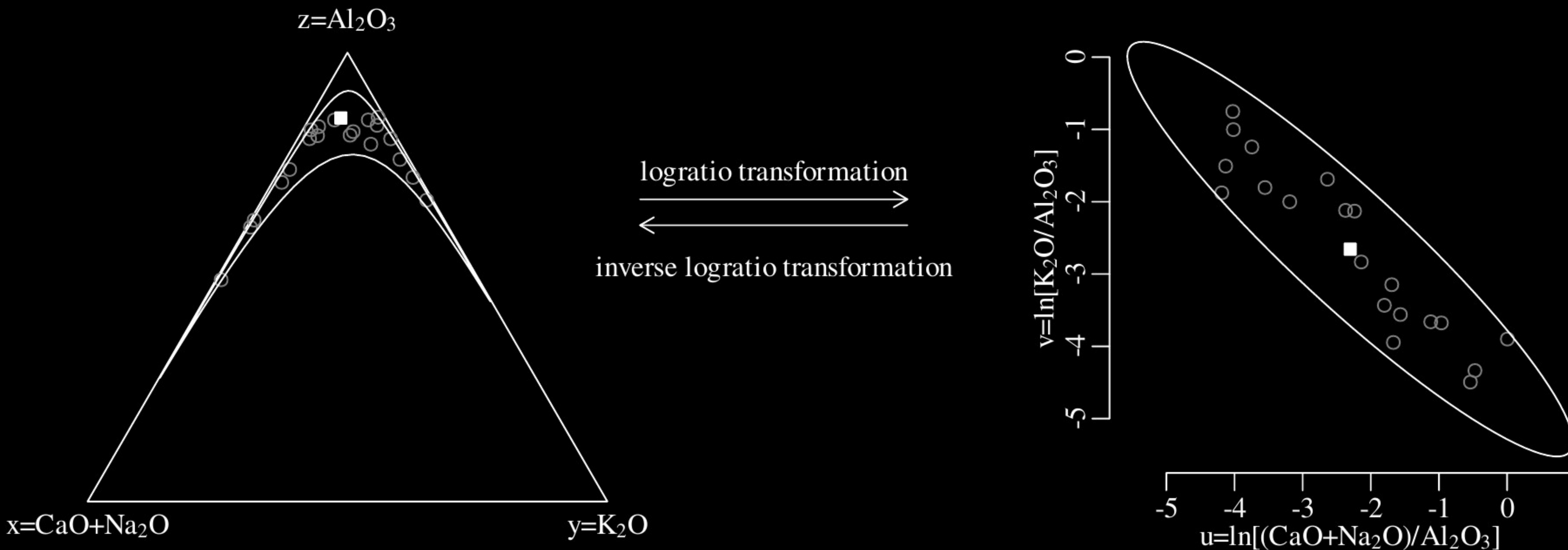
$Al_2O_3 : 0.763 \pm 0.195$

$CaO + Na_2O : 0.141 \pm 0.284$

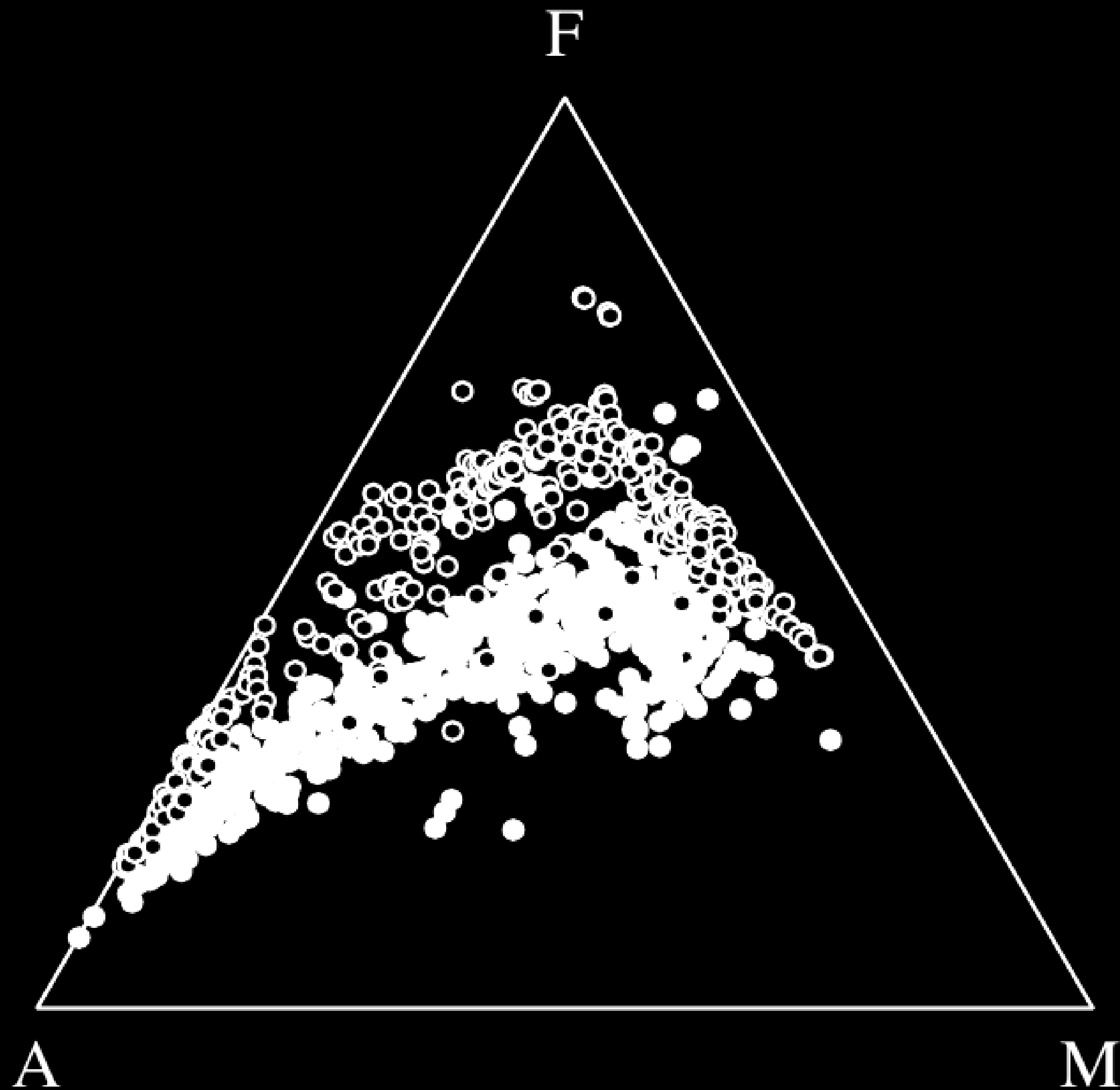
$K_2O : 0.096 \pm 0.185$

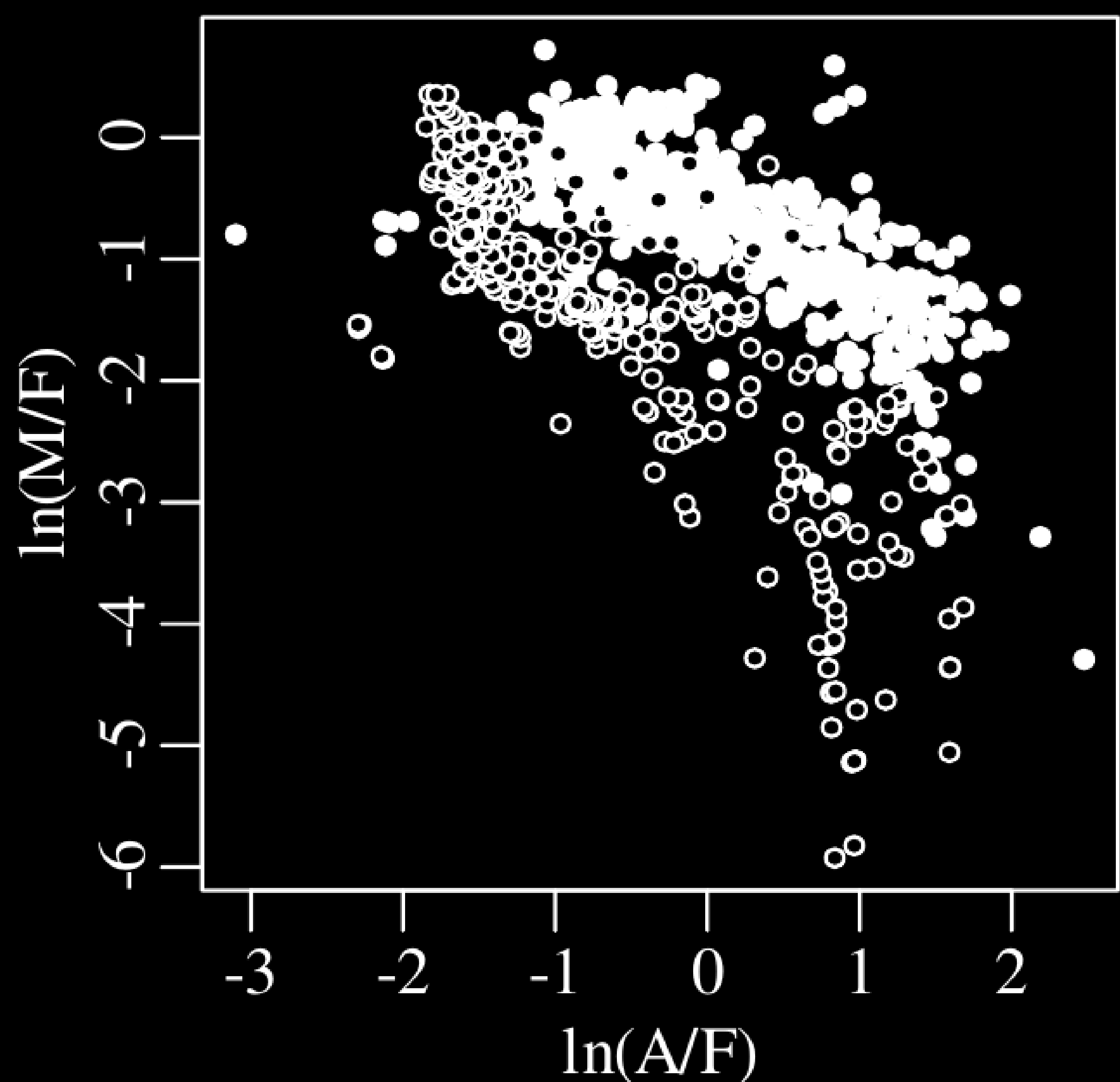


$$v = \ln\left[\frac{x}{z}\right], \; w = \ln\left[\frac{y}{z}\right]$$

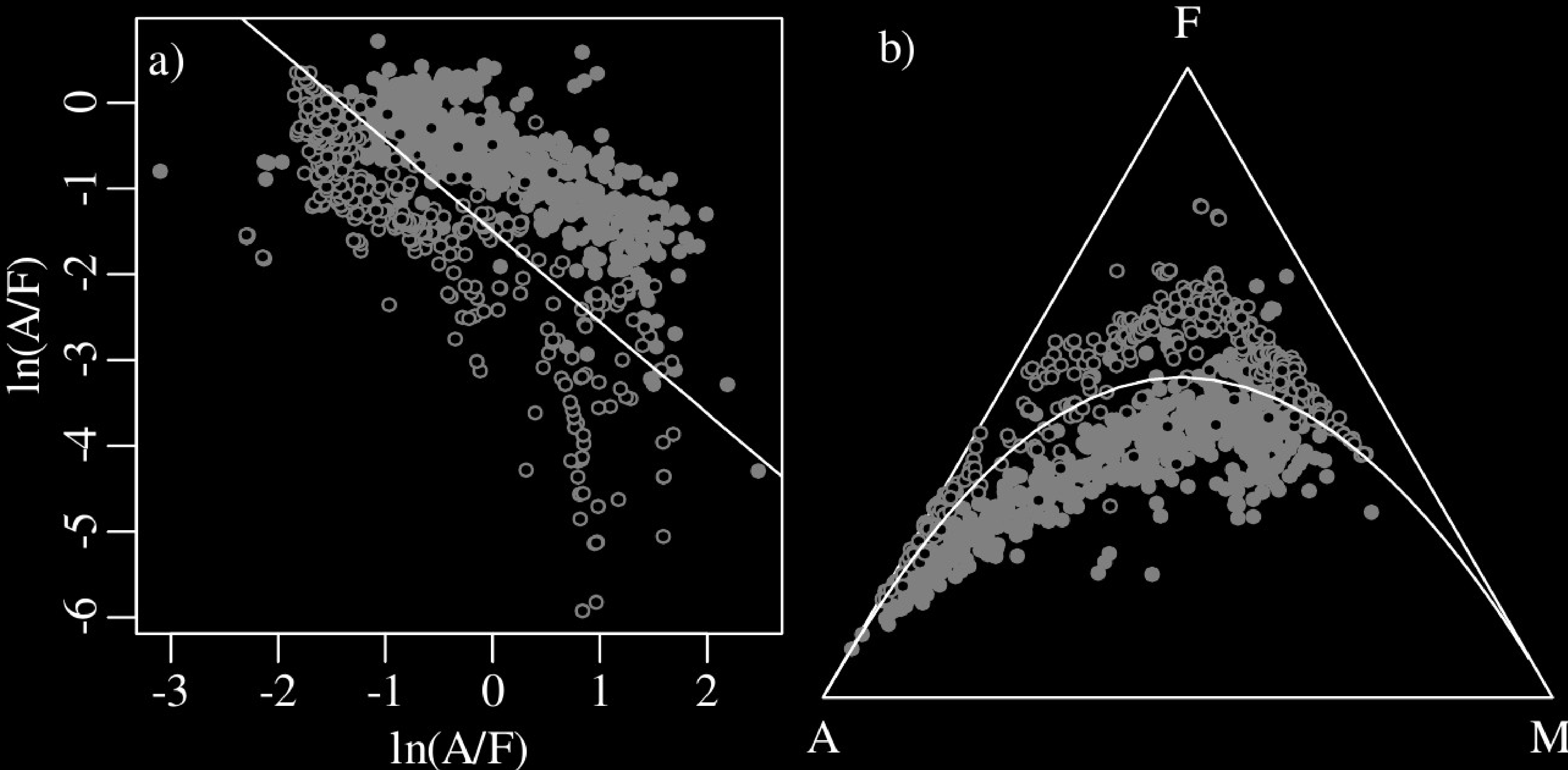


$$x = \frac{\exp[v]}{\exp[v] + \exp[w] + 1}, \; y = \frac{\exp[w]}{\exp[v] + \exp[w] + 1}, \; z = \frac{1}{\exp[v] + \exp[w] + 1}$$


F
A
M



ln(M/F)
0
-1
-2
-3
-4
-5
-6
-3
-2
-1
0
1
2
ln(A/F)



a)
b)
0
-1
-2
-3
-4
-5
-6
ln(A/F)
-3
-2
-1
0
1
2
ln(A/F)
F
A
M

1. `mtcars` is one of `R`'s built-in datasets. It contains a table with the fuel consumption and 10 aspects of automobile design and performance for 32 automobiles features in a 1974 edition of 'Motor Trend' magazine.

   (a) Calculate the average fuel consumption of these 32 vehicles –which are listed in miles per gallon– using the arithmetic mean.

   (b) Convert the data to European units, namely litres per 100km. To convert from mpg to litre/100km, use the following formula:

   $$1 \text{ litre/100km} = \frac{235.21}{1 \text{ mpg}}$$

   (c) Calculate the arithmetic mean fuel consumption in litre/100km.

   (d) Convert the arithmetic mean number of miles per gallon (from step 1a) to units of litre/100km. How does the resulting value compare with that obtained from step 1c.

   (e) Compute the geometric mean fuel consumption in mpg and litre/100km. Then convert the units of these mean values and compare with the result observed under step 1d.

2. Load `geostats`' `test` data into memory and plot the CaO–K_2O–Na_2O **subcomposition** on a ternary diagram. Add the arithmetic and geometric mean compositions to the plot. Experiment with the arguments to the `ternary` function to make the plot look as clear as possible.