Análisis de componentes principales

Capítulo 14 de

McCune y Grace 2002

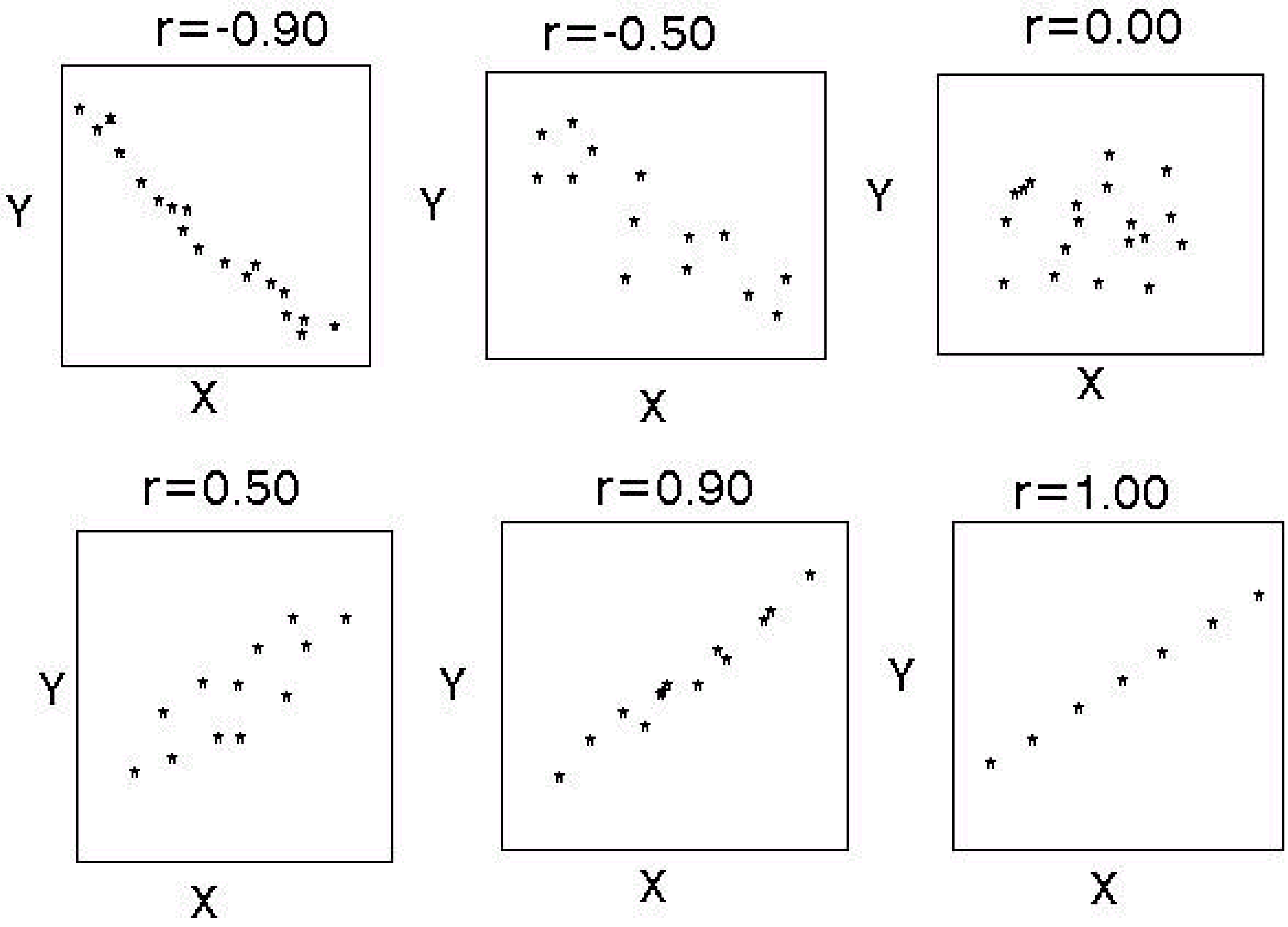
Algunas técnicas estadísticas

Relación entre 2 variables

Correlación

- Examina el grado en que 2 variables varían a la par.
- Por ejemplo, ¿existe una variación a la par entre el largo de la nariz (*x*) y el largo de la oreja izquierda (*y*)?
- La hipótesis nula sería:
  H_0: *x* no se correlaciona con *y*

Correlación



Correlación

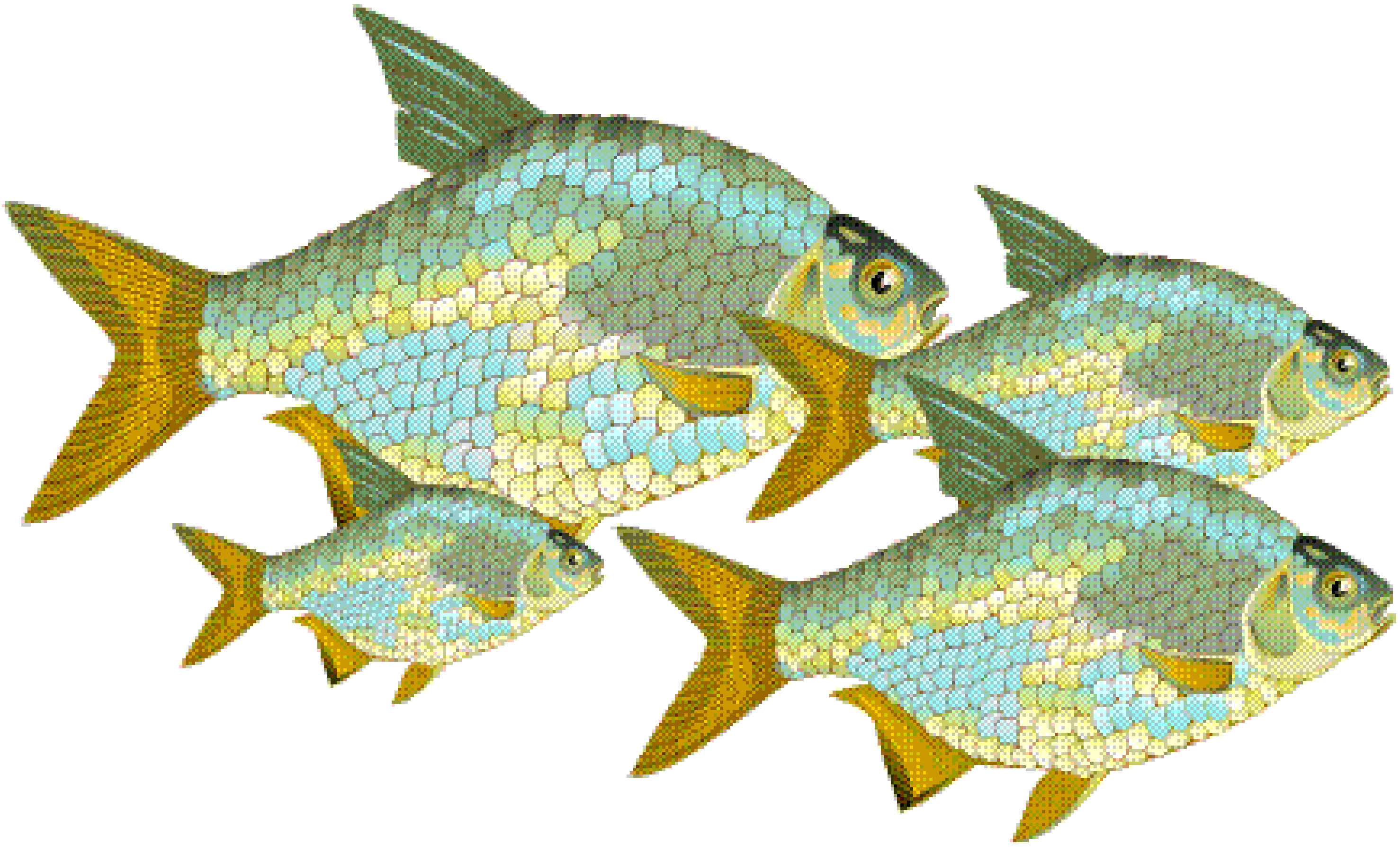
- r = coeficiente de correlación; provee una medida de la dispersión de los valores desde la línea de mejor correlación
- $y = a + bx$; define la línea de mejor correlación
- a = intercepto en y
- b = pendiente de la línea de correlación

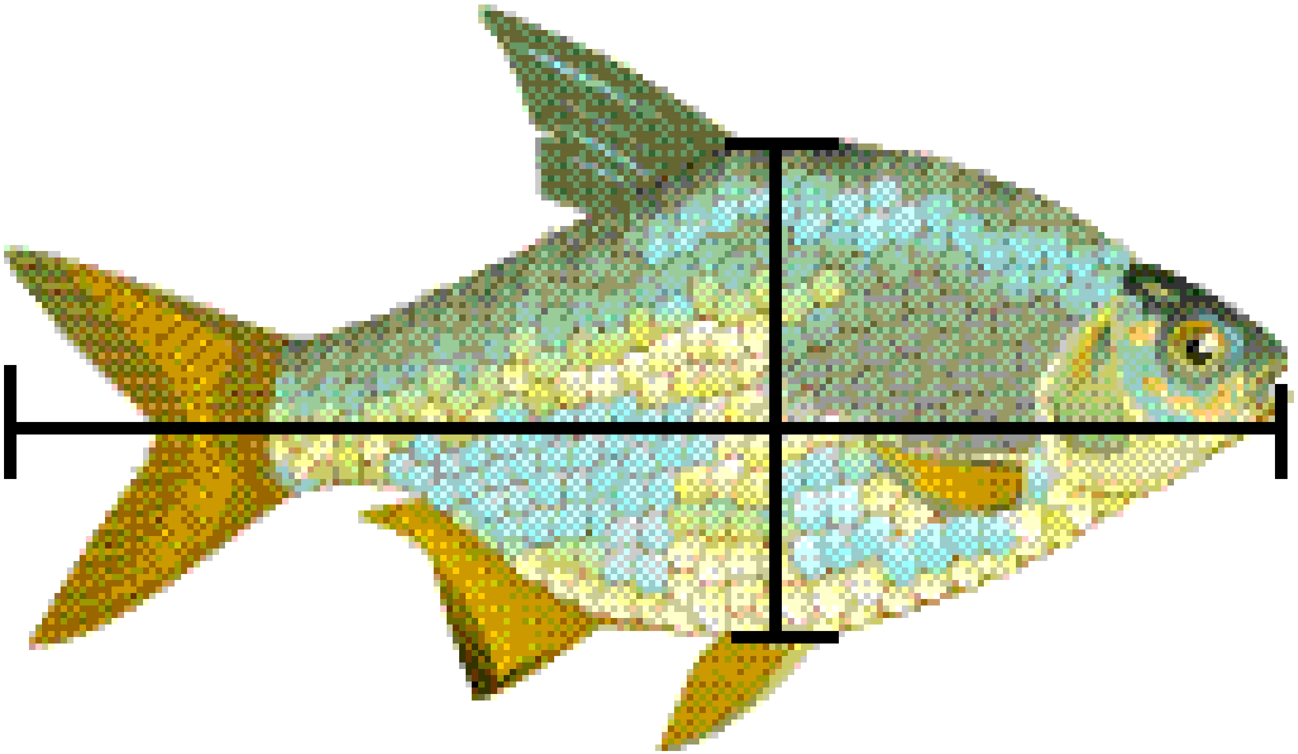
Cuando tratamos con más de 2 variables

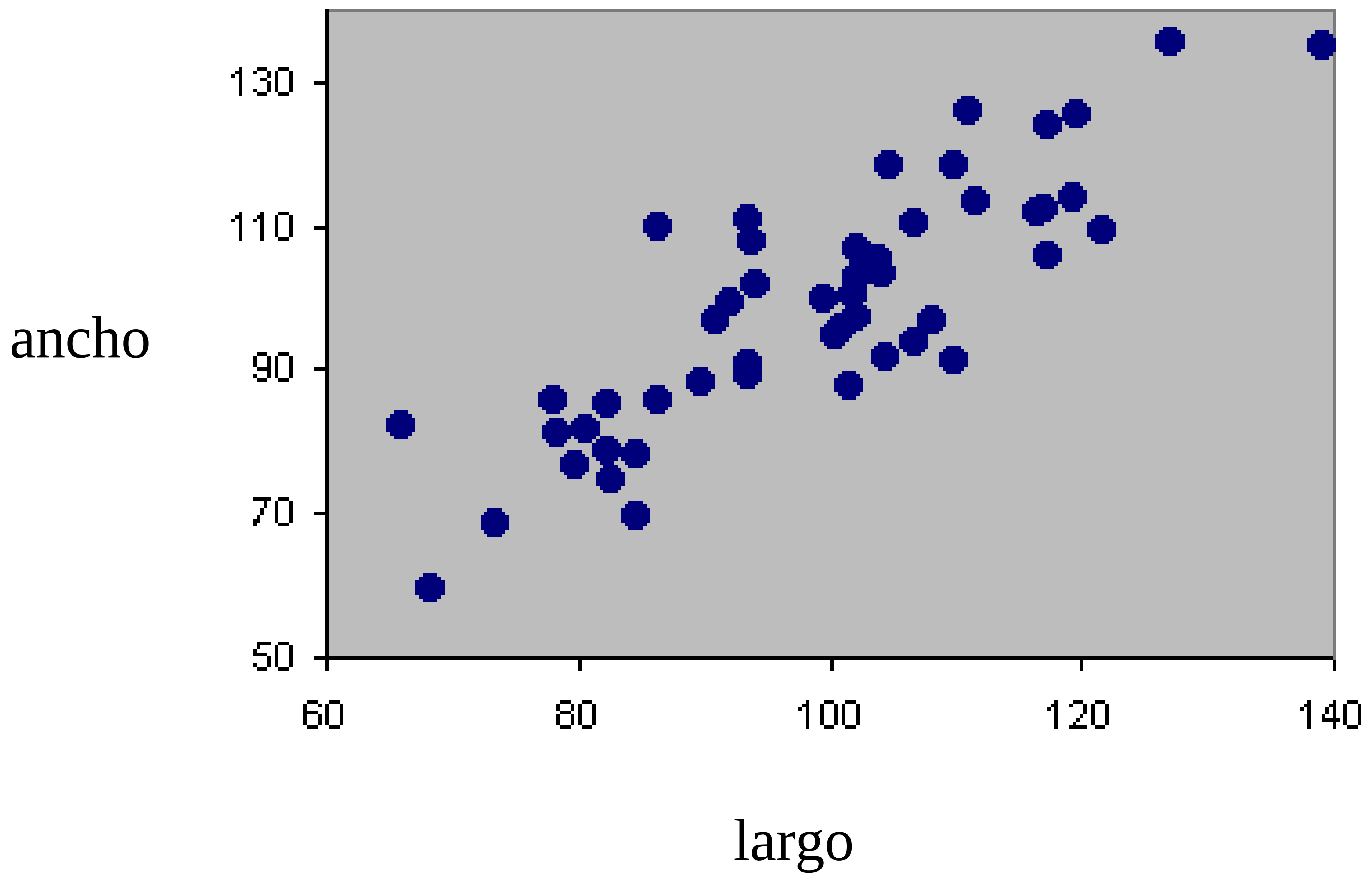
Análisis de componentes principales

Reducción de 3 dimensiones a sólo 2 dimensiones

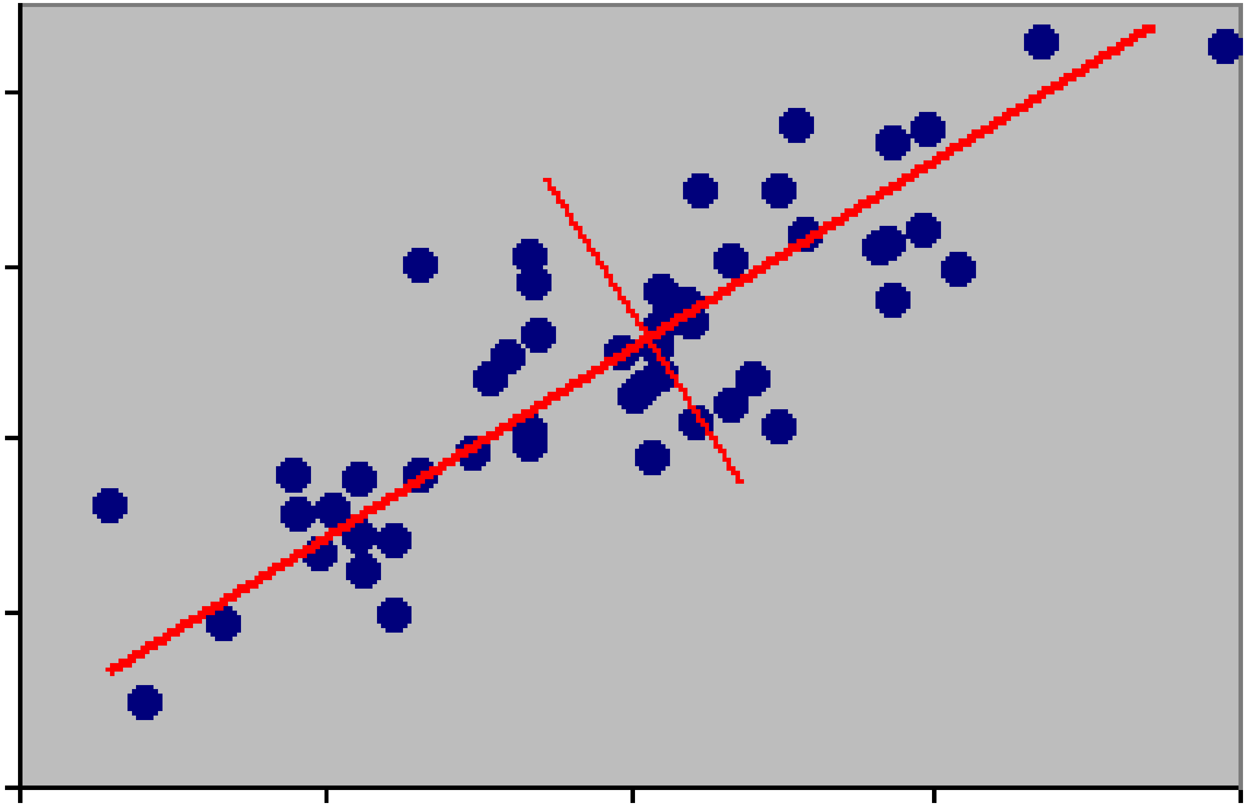
Tamaño





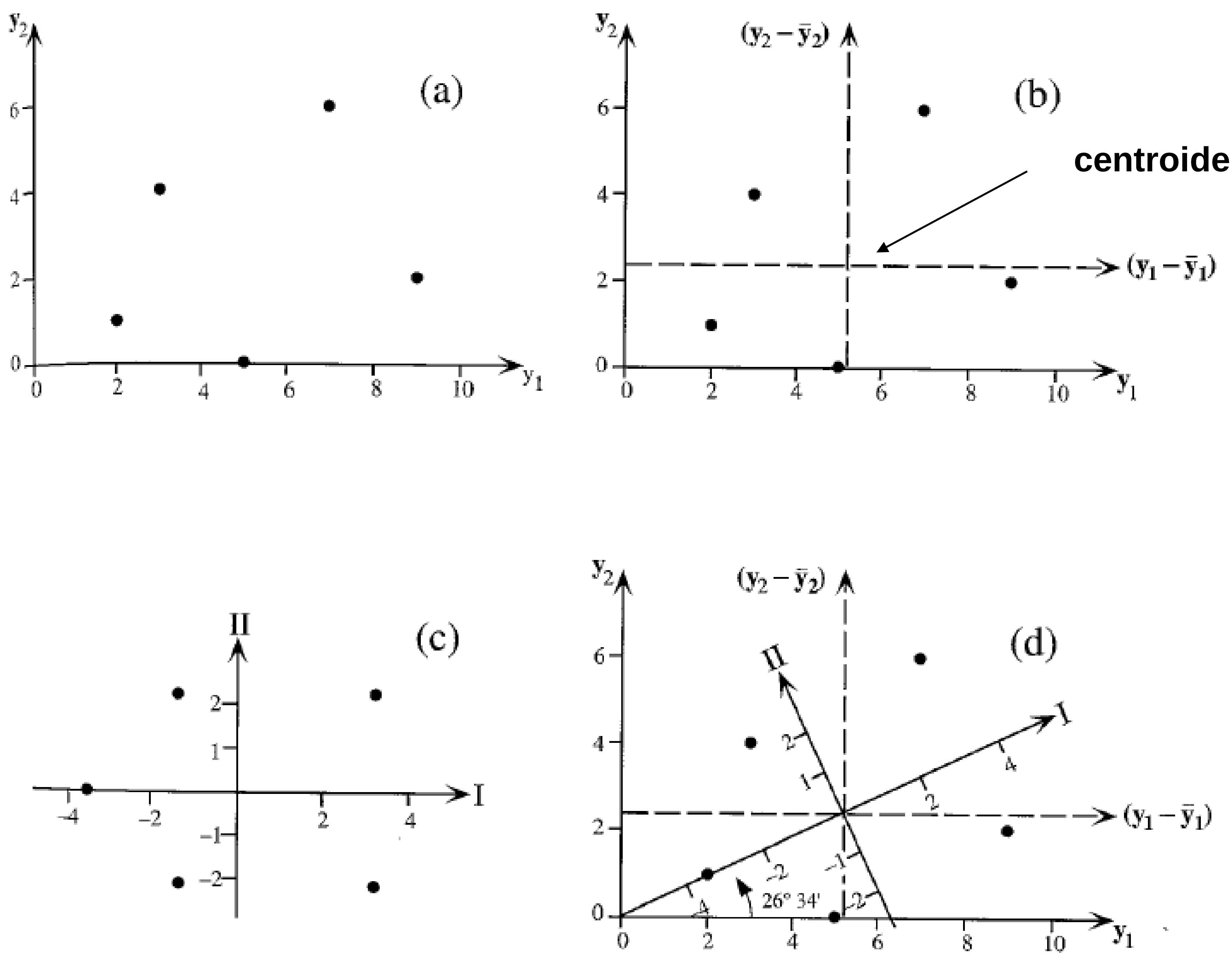


ancho
130
110
90
70
50
60
80
100
120
140
largo



Primer componente resume ambas variables:

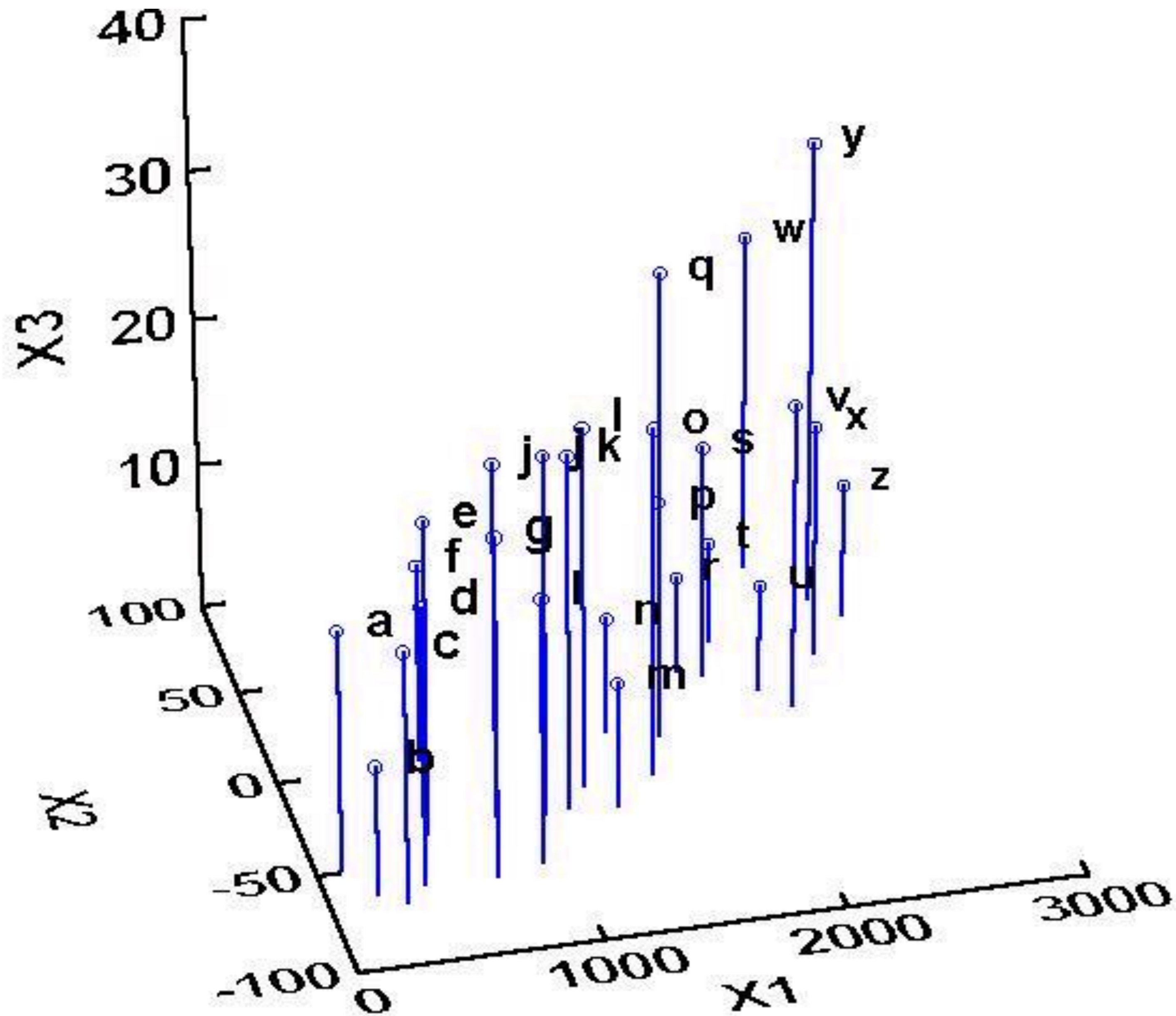
Tamaño = largo + ancho


(a)
(b)
(c)
(d)
centroide
y_1
y_2
$(y_1 - \bar{y}_1)$
$(y_2 - \bar{y}_2)$
I
II
26° 34'

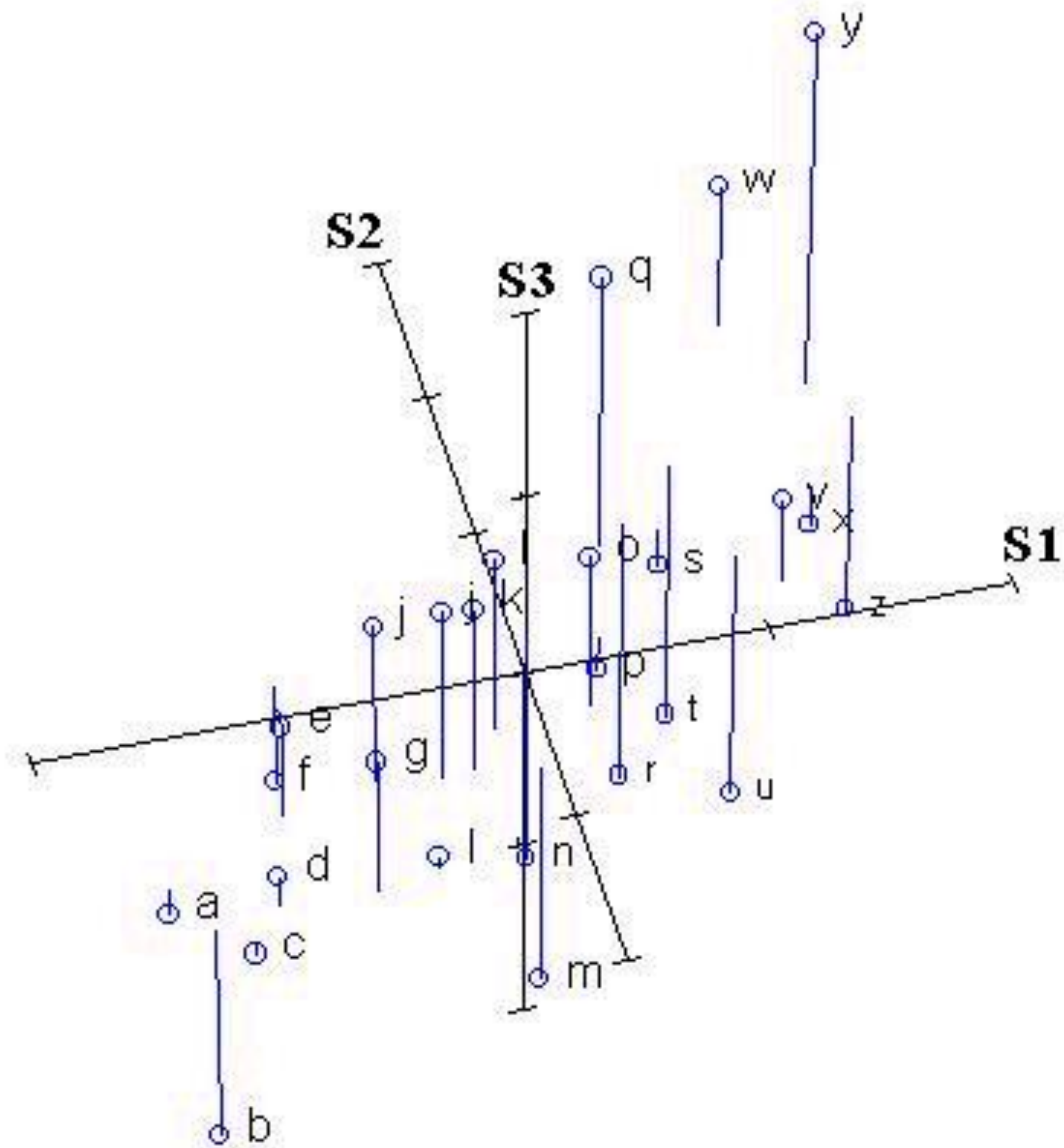
Otro ejemplo con 3 variables en 28 muestras

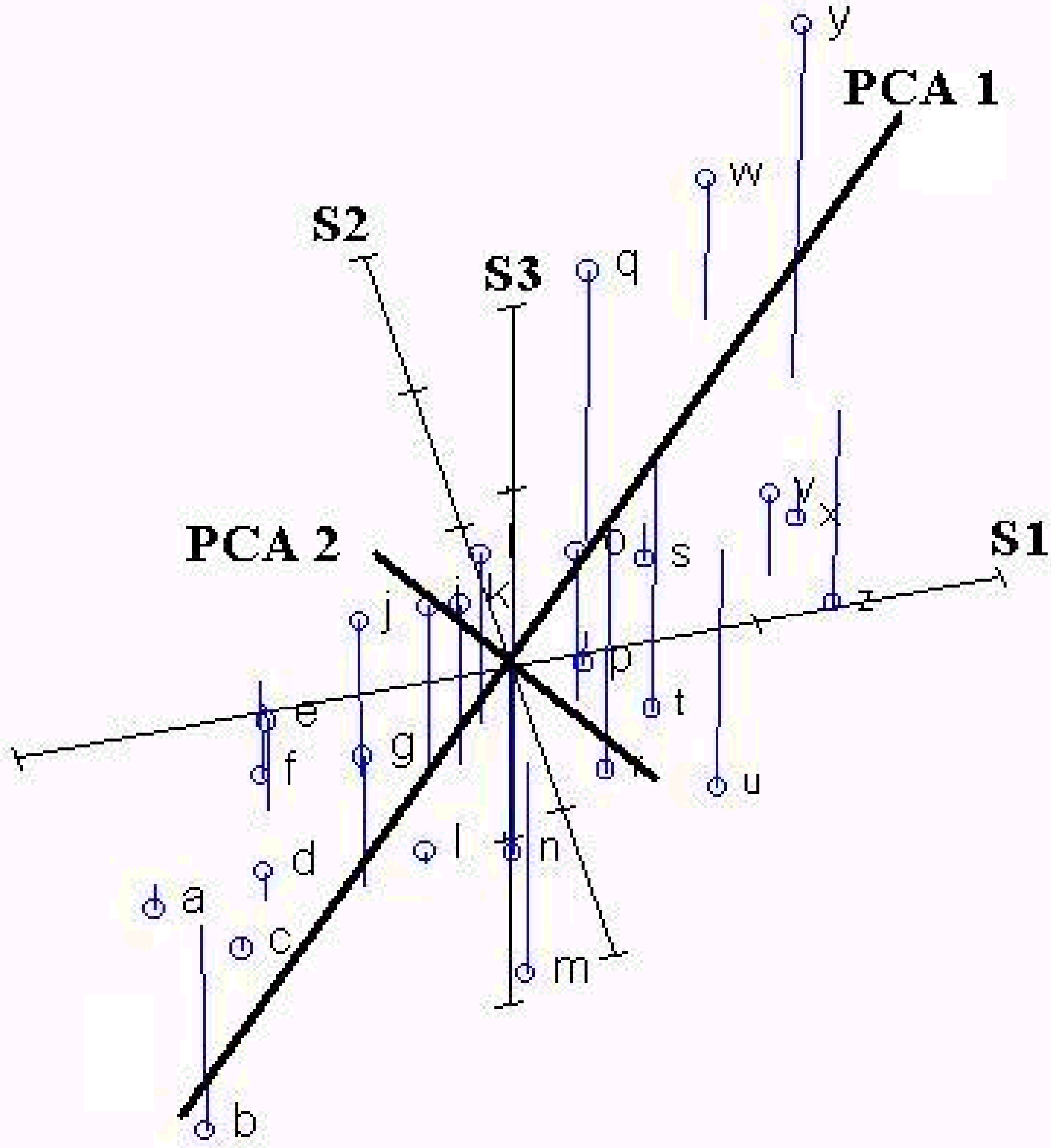
X1, X2, X3

a - z

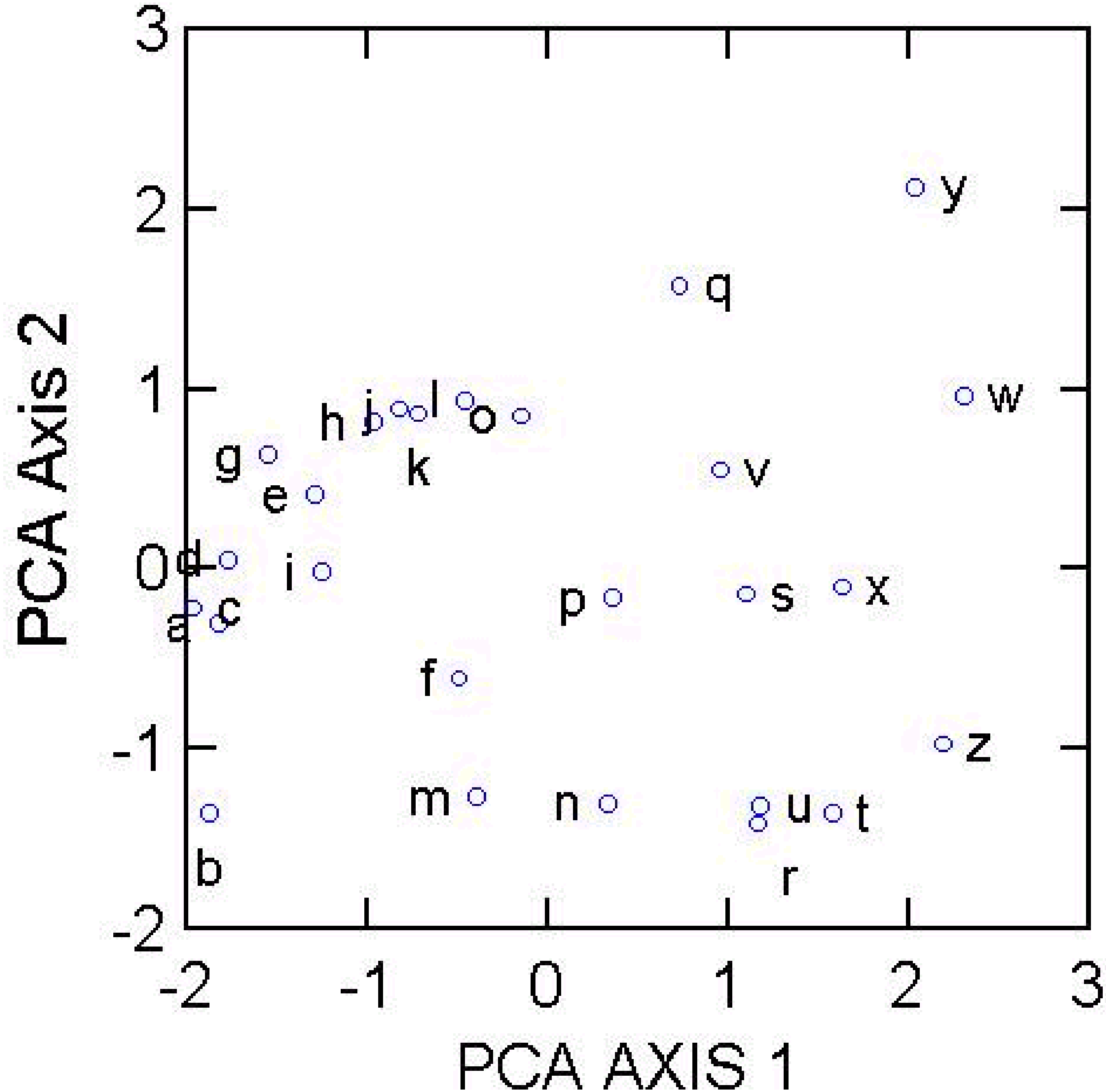


X3
X2
X1
40
30
20
10
100
50
0
-50
-100
0
1000
2000
3000
a
b
c
d
e
f
g
j
j
k
l
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z


S1
S2
S3
a
b
c
d
e
f
g
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z



PCA 1
PCA 2
S1
S2
S3
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z


PCA Axis 2
PCA AXIS 1
3
2
1
0
-1
-2
-2
-1
0
1
2
3
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

PCA

- Ecuación general para uno de los componentes principales:
- Posición en 1er componente (eje) principal $= a_1y_1 + a_2y_2 + a_3y_3 \ldots a_ny_n$
- Donde "a_1" = eigenvector de especie 1, y "y_1" = valor (abundancia) de especie 1
- eigenvalor = porción de la varianza total explicada por un componente

VARIANCE EXTRACTED, FIRST 9 AXES

AXIS	Eigenvalue	%Variance	Cum.%Var.	Broken-stick Eigenvalue
1	2.806	31.179	31.179	2.829
2	2.009	22.323	53.502	1.829
3	1.456	16.182	69.684	1.329
4	1.070	11.884	81.568	0.996
5	0.650	7.226	88.794	0.746
6	0.593	6.588	95.382	0.546
7	0.268	2.981	98.364	0.379
8	0.147	1.636	100.000	0.236
9	0.000	0.000	100.000	0.111

FIRST 6 EIGENVECTORS

	Eigenvector					
Species	1	2	3	4	5	6
Abgr-t	0.3746	0.4312	0.1875	0.0539	-0.1382	0.0486
Acma-t	0.3673	0.3561	0.2293	-0.3022	-0.3680	0.2511
Conu	0.3321	-0.5293	0.0516	0.1086	-0.1865	-0.1157
Frla	-0.0186	-0.2620	0.6508	-0.0473	-0.2761	-0.5382
Prav-t	0.3754	-0.4691	-0.2251	0.1103	0.0793	0.2340
Psme-t	0.2895	0.3066	0.0466	0.5199	0.4811	-0.4419
Pyco-t	-0.0943	-0.1363	0.6348	-0.0783	0.5731	0.4644
Quga-t	-0.5824	0.0749	-0.0112	-0.0208	-0.0808	-0.1369
Rhpu-t	0.2030	-0.0162	-0.1732	-0.7764	0.4022	-0.3836

¿Cuándo es apropiado?

- Ideal cuando las relaciones entre variables son lineales
- Las variables tienen distribuciones normales
- Ausencia de rezagados muy influyentes
- Pero…
  - Datos de comunidades generalmente no cumplen con esos requisitos

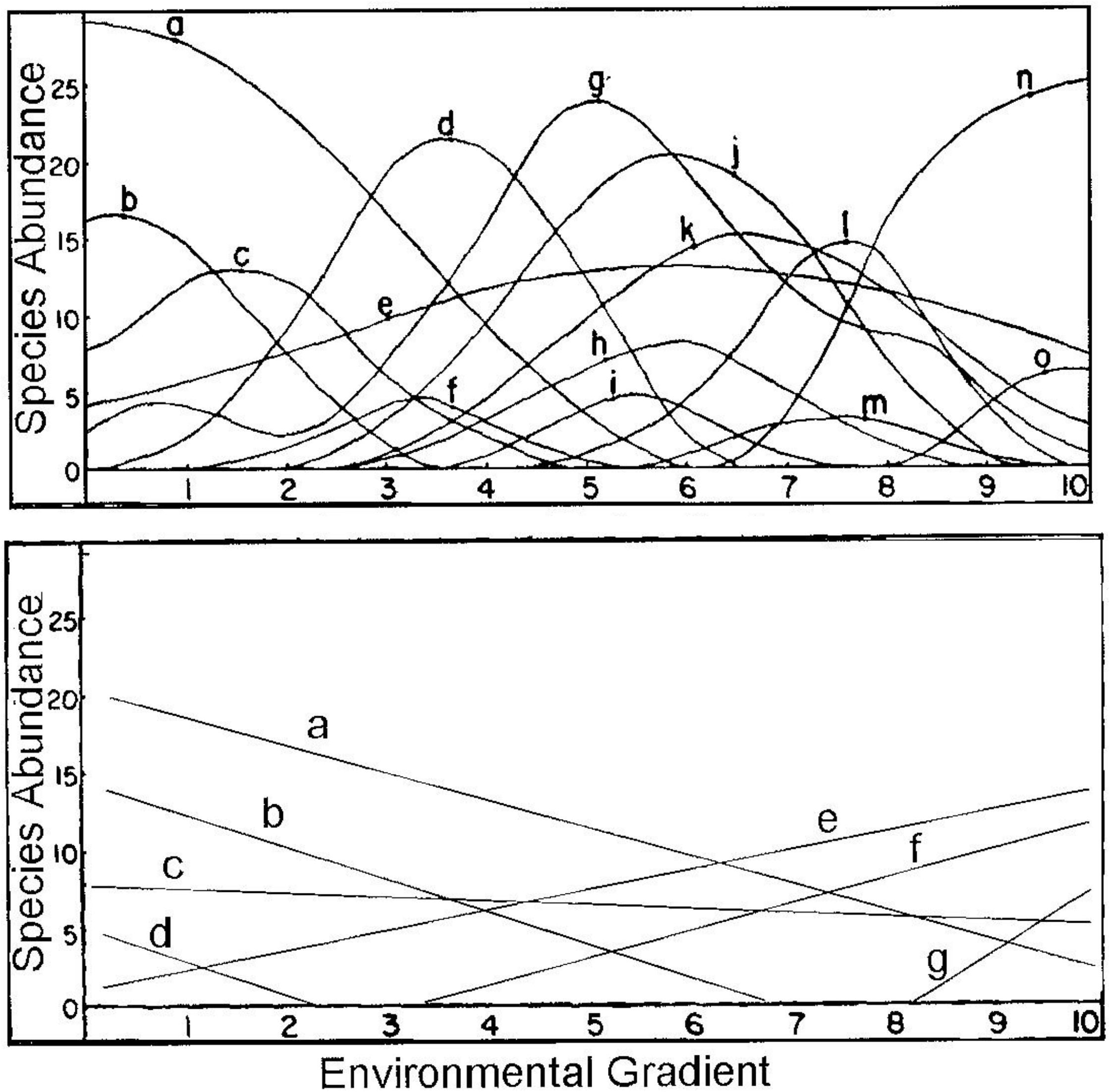
¿Qué informar?

- Forma de la matriz de productos cruzados: correlación o varianza/covarianza
- Justificación del modelo lineal
- Cuántos ejes fueron interpretados y la proporción de la varianza explicada
- Prueba de significancia para ejes
- Eigenvectores principales
- Ayudas para interpretación

Forma de la matriz de productos cruzados

- Correlación:
  - Estandariza las diferencias según la desviación estándar de cada variable.
  - Da igual peso a variables
  - Apropiada cuando las variables están en escalas distintas o hay mucha diferencia en su variación
- Varianza/covarianza:
  - Variables de mayor varianza tienen mas efecto en resultados

Justificación del modelo lineal



Cuántos ejes fueron interpretados

VARIANCE EXTRACTED, FIRST 9 AXES

AXIS	Eigenvalue	%Variance	Cum.%Var.	Broken-stick Eigenvalue
1	2.806	31.179	31.179	2.829
2	2.009	22.323	53.502	1.829
3	1.456	16.182	69.684	1.329
4	1.070	11.884	81.568	0.996
5	0.650	7.226	88.794	0.746
6	0.593	6.588	95.382	0.546
7	0.268	2.981	98.364	0.379
8	0.147	1.636	100.000	0.236
9	0.000	0.000	100.000	0.111