



Diplomado: Análisis Estadístico para Estudios Agropecuarios

Análisis Multivariado  
Análisis de Componentes Principales (ACP)

Módulo 4  
Análisis Multivariado

Erika Kanis Kuhl  
Ing. Agr. Dr.

Archivo Proteínas

Objetivo: estudiar los alimentos que se utilizan como fuentes proteicas en las dietas de los países europeos.

Los datos corresponden a la composición proteica de dietas de habitantes de países europeos según los alimentos consumidos (Carne Vacuna, Carne de Cerdo, Huevos, Leche, Pescado, Cereal, Embutidos, Frutos Secos y Frutas y Vegetales)

El archivo contiene 9 variables correspondientes a los alimentos consumidos y un criterio de clasificación con el nombre de los países

Matriz de diagramas de dispersión

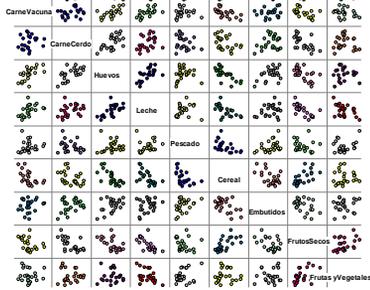


Gráfico de estrella

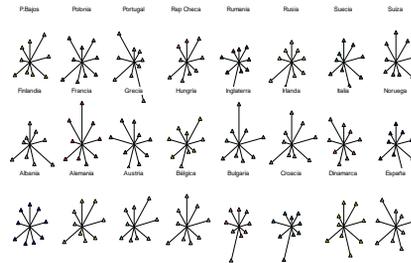


Gráfico en dos dimensiones

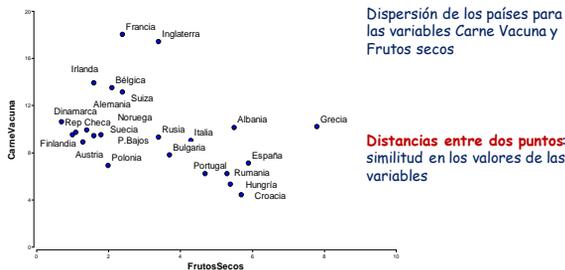
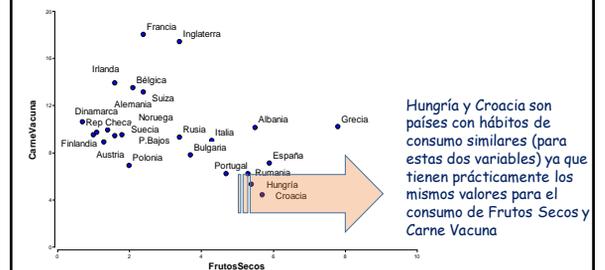
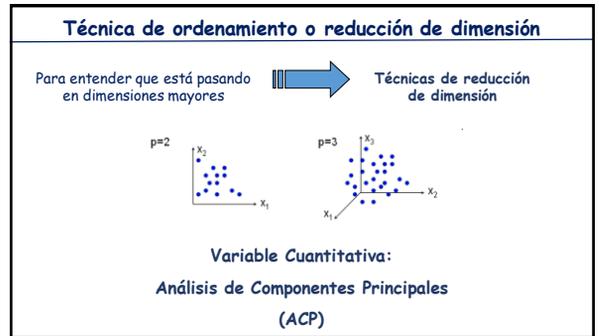
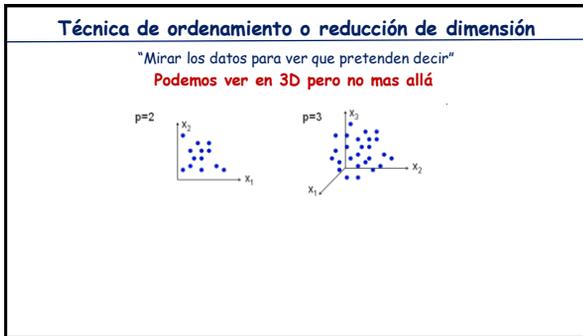
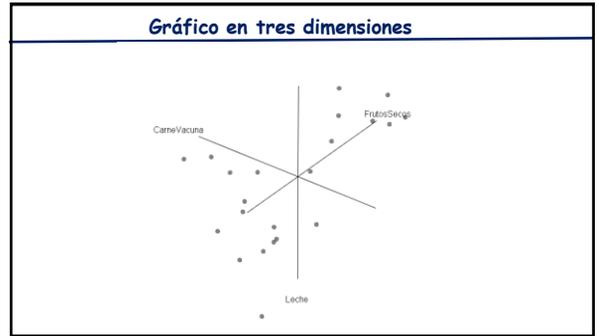
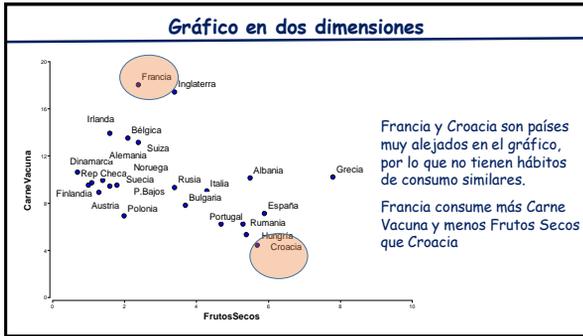


Gráfico en dos dimensiones





### Archivo Proteínas

*¿Será posible describir el conjunto de países utilizando un número menor de dimensiones sin perder información importante?*



Si proyectamos la foto en otra dirección.....



Nube de puntos



¿Podemos tomar una foto de la nube de puntos, de forma tal que se conserve la mayor parte de la información posible?

Video

Explicación intuitiva del  
Análisis de Componentes Principales

**Veamos una explicación  
intuitiva del Análisis de  
Componentes Principales  
(ACP)**

**¿Que vemos acá?**



Una Tetera

¿Cuál es la idea?

Queremos tomar una foto del objeto, o buscar la posición en la que se conserve la mayor parte de la información posible

¿Cómo lo hacemos?

Rotando la tetera hasta que consideremos la posición en que tengamos la mayor parte de la información posible

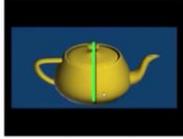
A ese eje le vamos a llamar  
Componente Principal 1  
(CP1)

**Componente Principal 1**



Luego rotamos el objeto hasta recoger la mayor parte de la información restante, y a eso le llamaremos  
Componente Principal 2 (CP2)

**Componente Principal 2**



Luego, lo que era un objeto en tres dimensiones, lo tenemos ahora en dos dimensiones, manteniendo prácticamente toda la información contenida en la representación del objeto.

En Resumen...

CP1: es la dirección del espacio que recoge la mayor parte de la variabilidad de los datos.

CP2: es la dirección del espacio que recoge la mayor parte de la variabilidad de los datos restante

**Técnica de ordenamiento o reducción de dimensión**

Variable Cuantitativa:  
Análisis de Componentes Principales  
(ACP)



"Simplificar la interpretación de un conjunto complejo de datos"

**Técnica de ordenamiento o reducción de dimensión**

Variable Cuantitativa:  
Análisis de Componentes Principales  
(ACP)



Realizan una proyección de todos los datos a un espacio de menor dimensión al espacio original de las variables, en el que podamos visualizar los datos.

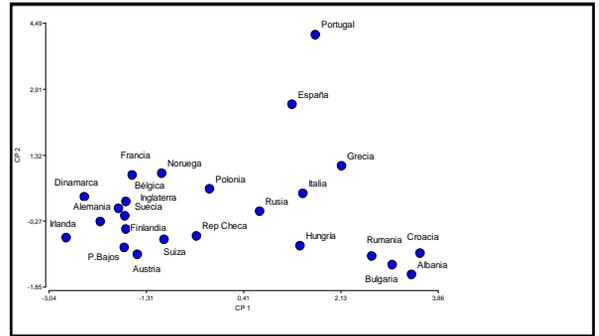


Rescatar un alto % de la variabilidad de las observaciones del espacio multidimensional

### ¿Cómo funciona un ACP?

Caso	País	CarneVacuna	CarneCerdo	Huevos	Leche	Pescado	Cereal	Embutidos	FrutosSecos	Frutas yVegetales	CP 1	CP 2
1	Alemania	10,10	1,40	0,50	8,90	0,20	42,30	0,60	5,50	1,70	3,39	-1,56
2	Austria	8,90	14,00	4,30	19,50	2,10	28,00	3,60	1,30	4,30	-1,46	-1,08
3	Bélgica	13,50	9,30	4,10	17,50	4,50	29,60	5,70	2,10	4,00	-1,66	0,20
4	Bulgaria	7,80	6,00	1,60	8,30	1,20	56,70	1,10	3,70	4,20	3,05	-1,33
5	Rep Checa	9,70	11,40	2,80	12,50	2,00	34,30	5,00	1,10	4,00	-0,41	-0,63
6	Dinamarca	10,60	10,80	3,70	25,00	9,90	21,90	4,80	0,70	2,40	2,41	0,31
7											-1,66	-0,47
8											-1,55	0,83
9											2,15	1,04
10											1,42	-0,87
11											-2,72	-0,67
12	Italia	9,00	5,10	2,90	13,70	3,40	30,90	2,10	4,30	0,70	1,47	0,38
13	P Bajos	9,50	13,60	3,60	23,40	2,50	22,40	4,20	1,80	3,70	-1,69	-0,91
14	Noruega	9,40	4,70	2,70	23,30	9,70	23,00	4,60	1,60	2,70	-1,03	0,88
15	Polonia	6,90	10,20	2,70	19,30	3,00	36,10	5,90	2,00	6,60	-0,19	0,49
16	Portugal	6,20	3,70	1,10	4,90	14,20	27,00	5,90	4,70	7,90	1,69	4,21
17	Rumania	6,20	6,30	1,50	11,10	1,00	49,00	3,10	5,30	2,80	2,68	-1,11
18	España	7,10	3,40	3,10	8,60	7,00	29,20	5,70	5,90	7,20	1,28	2,53
19	Suecia	9,90	7,80	3,50	24,70	7,50	19,50	3,70	1,40	2,00	-1,69	-0,16
20	Suiza	13,10	10,10	3,10	23,90	2,30	25,00	2,80	2,40	4,90	-0,99	-0,71
21	Inglaterra	17,40	5,70	4,70	20,60	4,30	24,30	4,70	3,40	3,30	-1,79	0,03
22	Rusia	9,30	4,00	2,10	16,60	3,00	43,60	6,40	3,40	2,90	0,71	-0,04
23	Alemania	11,40	12,50	4,10	18,60	3,40	18,00	5,20	1,50	3,80	-2,12	-0,29
24	Croacia	4,40	5,00	1,20	9,50	0,60	55,90	3,00	5,70	3,20	3,54	-0,66

Una CP es una combinación lineal de las variables originales



### ¿Cómo funciona un ACP?

Construye ejes artificiales (nuevas variables sintéticas llamadas CP)

Las observaciones son luego graficadas, usando estas nuevas variables sintéticas.

Como el ordenamiento de las observaciones se produce en un espacio de menor dimensión, esta técnica recibe el nombre de:

"Técnica de ordenamiento" o "Técnica de reducción de dimensión"

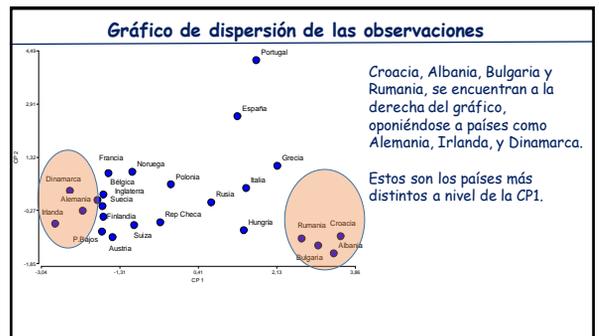
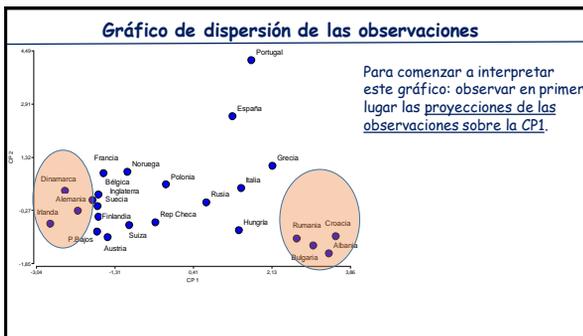
### ¿Que son la Componentes Principales (CP) ?

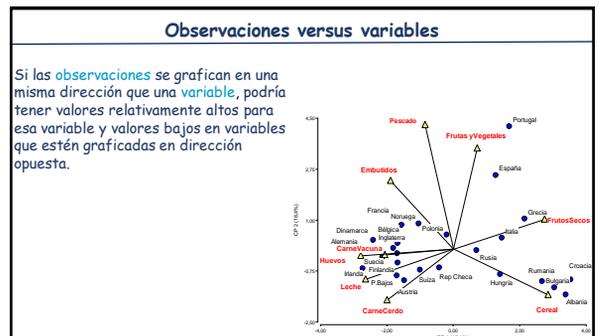
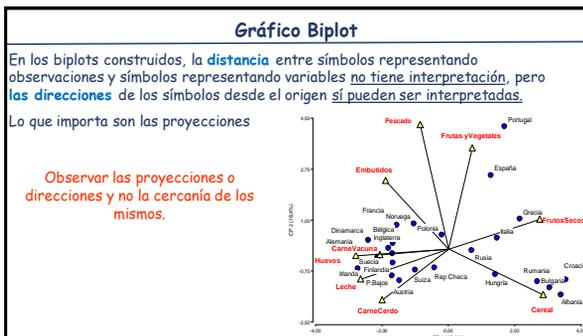
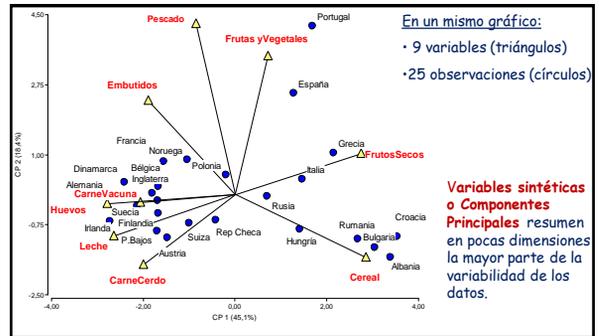
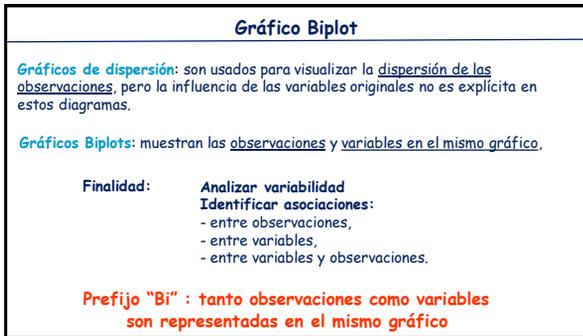
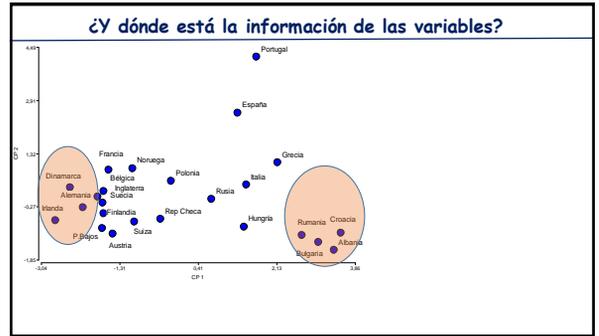
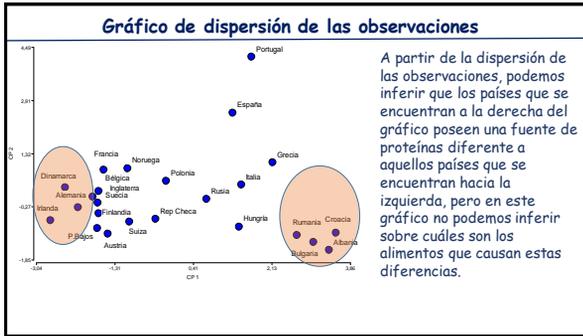
Una CP es una combinación lineal de las variables originales, en donde a cada variable se le asigna un peso diferente.

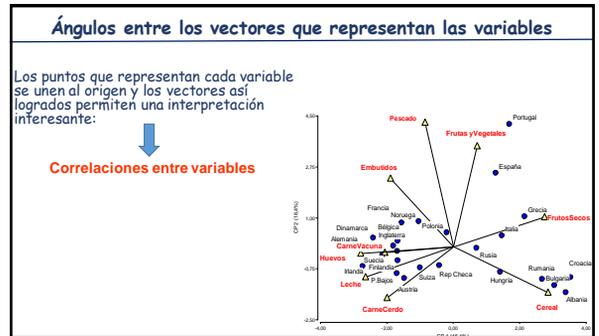
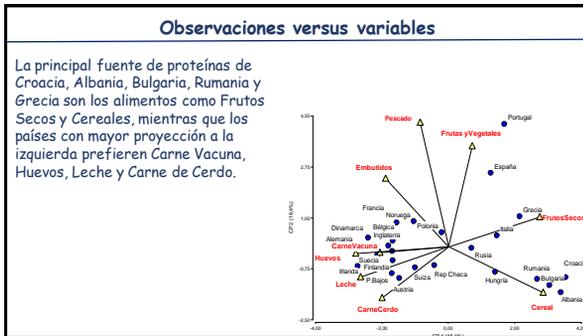
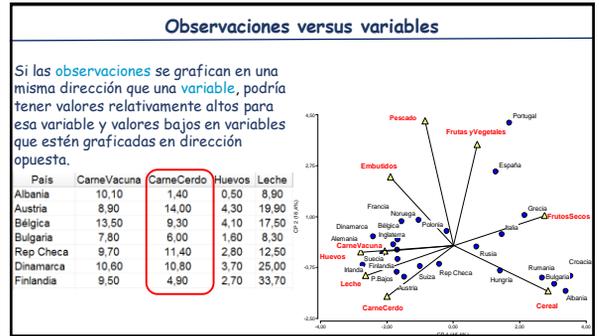
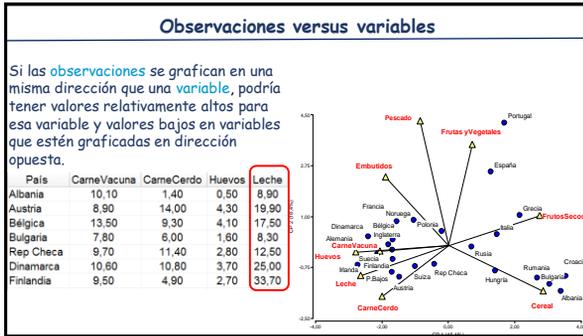
**CP1:  $ax_1+bx_2+cx_3+\dots$**

↓

Con la finalidad de hacer algunas de ellas más importantes en la descripción de la variabilidad de los datos.







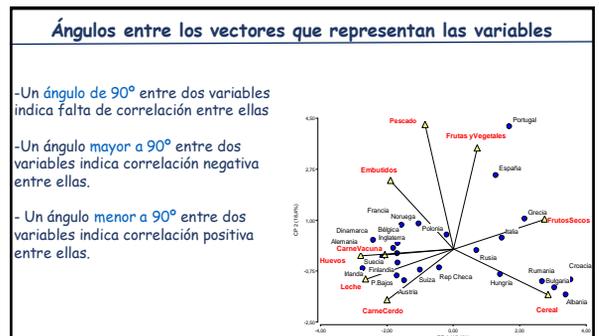
### Tabla de Resultados

#### Matriz de correlación/Coefficientes

	CarneVacuna	CarneCerdo	Huevos	Leche	Pescado	Cereal	Embutidos	FrutosSecos	Frut yVeget
CarneVacuna	1								
CarneCerdo	0,18	1							
Huevos	0,63	0,61	1						
Leche	0,5	0,33	0,62	1					
Pescado	0,07	-0,25	0,06	0,15	1				
Cereal	-0,52	-0,4	-0,71	-0,68	-0,52	1			
Embutidos	0,17	0,22	0,43	0,29	0,4	-0,52	1		
FrutosSecos	-0,38	-0,62	-0,55	-0,69	-0,14	0,64	-0,44	1	
Frutas yVegetales	-0,08	-0,05	-0,04	-0,43	0,27	0,04	0,11	0,37	1

#### Matriz de correlación/Probabilidades

	CarneVacuna	CarneCerdo	Huevos	Leche	Pescado	Cereal	Embutidos	FrutosSecos	Frut yVeget
CarneVacuna									
CarneCerdo	0,4103								
Huevos	0,0017	0,0015							
Leche	0,0135	0,1145	0,0013						
Pescado	0,7541	0,2301	0,793	0,4763					
Cereal	0,0091	0,0552	0,0001	0,0009	0,009				
Embutidos	0,4322	0,1995	0,0341	0,1715	0,0516	0,0094			
FrutosSecos	0,0646	0,0013	0,0057	0,0002	0,5252	0,0007	0,0329		
Frutas yVegetales	0,7095	0,8181	0,8632	0,0375	0,1993	0,8599	0,6203	0,0738	



### Tabla de Resultados

#### Matriz de correlación/Coefficientes

	CarneVacuna	CarneCerdo	Huevos	Leche	Pescado	Cereal	Embutidos	FrutosSecos	Frut yVeget
CarneVacuna	1								
CarneCerdo	0,18	1							
Huevos	0,61	0,61	1						
Leche	0,5	0,33	0,62	1					
Pescado	0,07	-0,25	0,06	0,15	1				
Cereal	-0,52	-0,4	-0,71	-0,63	-0,52	1			
Embutidos	0,17	0,27	0,43	0,29	0,4	-0,52	1		
FrutosSecos	-0,38	0,62	-0,55	-0,69	-0,14	0,64	-0,44	1	
Frutas yVegetales	-0,08	-0,05	-0,04	-0,43	0,27	0,04	0,11	0,37	1

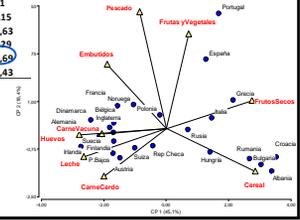
#### Matriz de correlación/Probabilidades

	CarneVacuna	CarneCerdo	Huevos	Leche	Pescado	Cereal	Embutidos	FrutosSecos	Frut yVeget
CarneVacuna	0,4103								
CarneCerdo	0,0017	0,0015							
Huevos	0,0135	0,1145	0,0013						
Leche	0,7541	0,2301	0,733	0,4763					
Pescado	0,0091	0,0552	0,0001	0,0009	0,009				
Cereal	0,4322	0,1995	0,0341	0,1715	0,0516	0,0094			
Embutidos	0,0646	0,0013	0,0057	0,0002	0,5292	0,0007	0,0329		
FrutosSecos	0,7095	0,8181	0,8632	0,0375	0,1993	0,8399	0,6203	0,0738	

### Ángulos entre los vectores que representan las variables

#### Matriz de correlación/Coefficientes

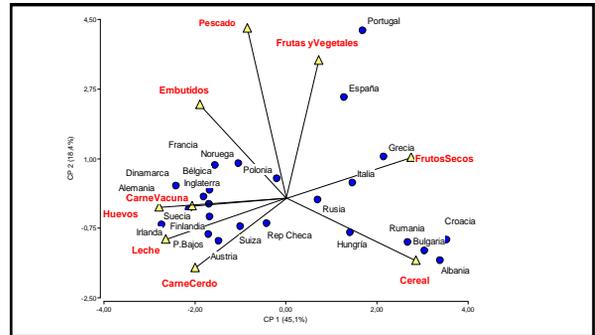
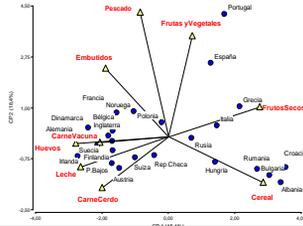
	CarneVacuna	CarneCerdo	Huevos	Leche
CarneVacuna	1			
CarneCerdo	0,18	1		
Huevos	0,61	0,61	1	
Leche	0,5	0,33	0,62	1
Pescado	0,07	-0,25	0,06	0,15
Cereal	-0,52	-0,4	-0,73	-0,63
Embutidos	0,17	0,27	0,43	0,29
FrutosSecos	-0,38	0,62	-0,55	-0,69
Frutas yVegetales	-0,08	-0,05	-0,04	-0,43



### Ángulos entre los vectores que representan las variables

Si el gráfico resume mas de un 70-80 % de la variabilidad total, la interpretación leída con respecto al ángulo de las variables es muy consistente con la matriz de correlación.

Si el gráfico resume menos de un 70 % de la variabilidad total, pueden existir inconsistencias entre las interpretaciones leídas con respecto al ángulo de las variables y la matriz de correlación



### Componentes Principales

- Las CP se obtienen de forma ordenada, en función de la cantidad de varianza que son capaces de explicar.
- En este sentido, el primer componente será el más importante por ser el que explica un mayor porcentaje de la varianza de los datos.
- En el ACP se obtienen tantas CPs como variables analizas.
- Las CP no están correlacionadas entre sí, es decir son independientes. Esto es importante al momento de la interpretación, ya que cada CP puede analizarse separadamente o independientemente, entregando nueva información e independiente a la de las demás CP
- Un gráfico de dispersión construido a partir de la CP1 y la CP2 proyecta la nube de puntos en el sentido de máxima variación.

### Tabla de Resultados

#### Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Porcentaje de la variabilidad total que existe en el espacio multidimensional que explica cada eje

**Tabla de Resultados**

Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Porcentaje de la variabilidad total que existe en el espacio multidimensional que explica cada eje

Valores propios de cada componente principal → Son las varianzas de cada CP

Proporción de la varianza explicada por cada CP → Varianza explicada acumulada

**Tabla de Resultados**

Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Porcentaje de la variabilidad total que existe en el espacio multidimensional que explica cada eje

Porcentaje de variabilidad explicado por la CP1: Varianza CP1/Varianza total  $4,06/8,99 = 0,45$

Valores propios de cada componente principal → Son las varianzas de cada CP

Proporción de la varianza explicada por cada CP → Varianza explicada acumulada

**Tabla de Resultados**

Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Estas nuevas variables o componentes principales resumen en pocas dimensiones la mayor parte de la variabilidad de un gran número de variables.

**Tabla de Resultados**

Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Un 64 % de la variabilidad que existe en la nube de puntos multidimensional es explicada por las dos primeras CP

CP1: Combinación lineal que explica la mayor variabilidad de los datos. El 45 % de la variabilidad total de los datos esta representada por la CP1

"La sombra de las observaciones en el eje 1, esta representando un 45 % de la variabilidad de los datos que hay en el hiperespacio que no podemos ver"

**Tabla de Resultados**

Autovalores: valores propios (eigenvalue)

	Lambda	Valor	Proporción	Prop Acum
CP1	1	4,06	0,45	0,45
CP2	2	1,66	0,18	0,64
CP3	3	1,1	0,12	0,76
CP4	4	0,9	0,1	0,86
CP5	5	0,48	0,05	0,91
CP6	6	0,32	0,04	0,95
CP7	7	0,25	0,03	0,98
CP8	8	0,12	0,01	0,99
CP9	9	0,1	0,01	1

Se pueden construir tantas CP como variables originales tengamos

**Selección del número de CP**

Autovalores: valores propios (eigenvalue)

Lambda	Valor	Proporción	Prop Acum
1	4,06	0,45	0,45
2	1,66	0,18	0,64
3	1,1	0,12	0,76
4	0,9	0,1	0,86
5	0,48	0,05	0,91
6	0,32	0,04	0,95
7	0,25	0,03	0,98
8	0,12	0,01	0,99
9	0,1	0,01	1

Existen diferentes propuestas para decidir la selección del número de CP con que trabajar:

- Gráfico de autovalores. Se busca un "quiebre" a partir del cual todos los autovalores posteriores son iguales entre si
- Fijar la proporción de la varianza explicada (60-70 % o más)
- Guardar tantos ejes como autovalores mayores al valor promedio de los autovalores haya.

### Tabla de Resultados

**Autovectores (eigenvector)** Información sobre el peso de cada variable para conformar cada CP (importancia)

Variables	e1	e2	e3	etc...
CarneVacuna	-0,31	-0,03	-0,1	
CarneCerdo	-0,3	-0,26	0,59	
Huevos	-0,42	-0,03	0,26	
Leche	-0,4	-0,16	-0,33	
Pescado	-0,13	0,65	-0,34	
Cereal	0,43	-0,24	0,08	
Embutidos	-0,29	0,36	0,14	
FrutosSecos	0,42	0,16	0,06	
Frutas yVegetales	0,11	0,53	0,56	

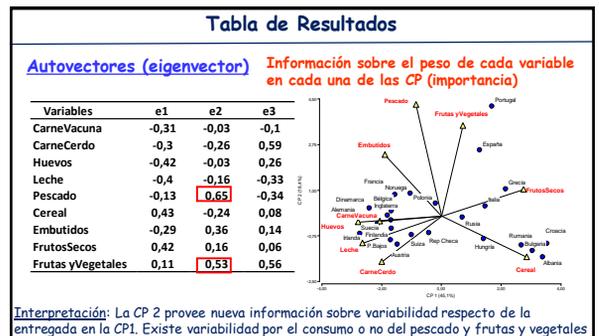
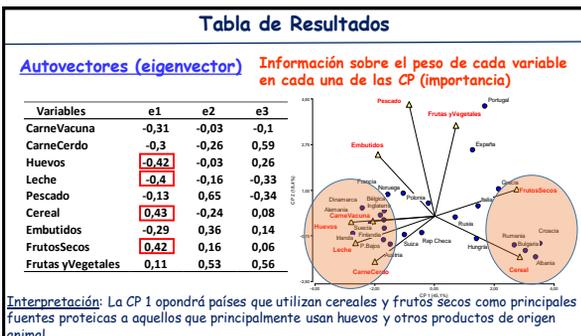
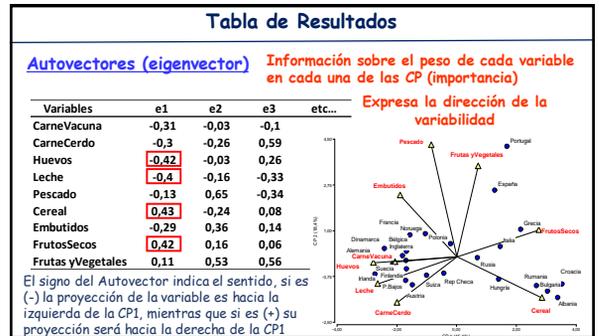
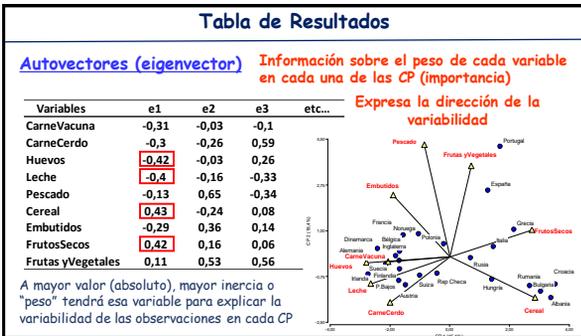
### Tabla de Resultados

**Autovectores (eigenvector)** Información sobre el peso de cada variable para conformar cada CP (importancia)

Variables	e1	e2	e3	etc...
CarneVacuna	-0,31	-0,03	-0,1	
CarneCerdo	-0,3	-0,26	0,59	
Huevos	-0,42	-0,03	0,26	
Leche	-0,4	-0,16	-0,33	
Pescado	-0,13	0,65	-0,34	
Cereal	0,43	-0,24	0,08	
Embutidos	-0,29	0,36	0,14	
FrutosSecos	0,42	0,16	0,06	
Frutas yVegetales	0,11	0,53	0,56	


**Coefficientes con que cada variable fue ponderada para conformar las CPs**  
 $CP1\ Albania: (-0,31 \cdot 0,06) + (-0,3 \cdot 1,72) + \dots + (0,11 \cdot 1,34) = 3,39$

	CV	CC	H	L	P	C	E	FS	FvV	CP1	CP2
Albania	10,10	1,40	0,50	8,90	0,20	42,30	0,60	5,50	1,70	3,39	-1,56
Datos estandarizados	0,06	-1,72	-2,13	-1,18	-1,16	0,88	-2,24	1,18	-1,34		



**Tabla de Resultados**

**Correlación cofenética:**

Es una medida para saber que tan bien anduvo la reducción de dimensión

Cercana a 1: no hay mucha deformación en la proyección de las observaciones del hiperespacio al plano de dos o tres ejes.

Correlación cofenética= **0,910**

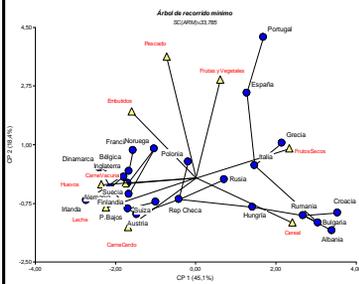
**Árbol de recorrido mínimo**

Al trabajar con proyecciones a un plano de dos ejes, existen deformaciones de la nube de puntos original (estamos "aplastando la nube")

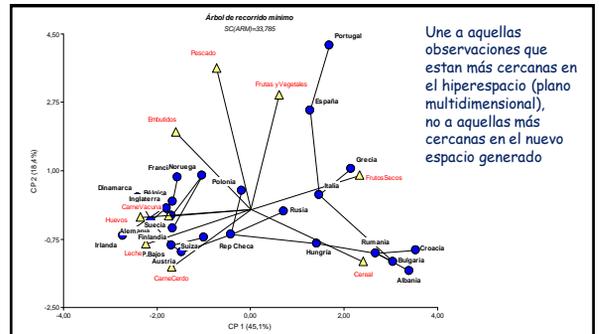
Árbol de recorrido mínimo: es una unión de segmentos de todas las observaciones según la distancia que ellas tenían en el espacio multidimensional

Las uniones de estos segmentos respetan lo que se estaría viendo en el plano multidimensional.

**¿Cuándo presentar un árbol de recorrido mínimo?**



Muy recomendable para planos en donde el % de variabilidad es bajo, y por lo tanto hay mas chance de que existan deformaciones debidas a la proyección.



Une a aquellas observaciones que estan más cercanas en el hiperespacio (plano multidimensional), no a aquellas más cercanas en el nuevo espacio generado

**Existen dos opciones:**

Realizar el ACP sobre la:



Matriz de correlación (R)



Matriz de varianzas y covarianzas (S)

**ACP sobre la Matriz de correlación (R)**

El ACP a partir de la matriz de correlación (R), es útil cuando:

- Las unidades de medida de las variables son diferentes y/ó
- Cuando las variables (que están en la misma unidad de medida) no tienen varianzas similares, de otro modo las variables con mayor varianza (no necesariamente las más informativas) tendrán demasiada influencia en los resultados del análisis

Obtener las CP a partir de la matriz de correlación

Trabajar con los datos estandarizados (Dato-Media/DE)

### ACP sobre la Matriz de varianzas y covarianzas (S)

El ACP a partir de la matriz de varianzas y covarianzas (S), es útil cuando:

- Las unidades de medida de las variables son iguales y sus varianzas similares
- Las unidades de medida de las variables son iguales y el objetivo del estudio esta centrado en estudiar las varianzas de cada variable



Obtener las CP a partir de la matriz de varianzas y covarianzas



Trabajar con los datos sin estandarizar

### ACP sobre la Matriz de varianzas y covarianzas (S)

Al trabajar con los datos no estandarizados, el % de explicación de los ejes es mayor ya que:

- No solo explica la variabilidad de las observaciones
- Sino que también se ven explicadas las varianzas de las variables

Los rayos de los vectores tienen distintas longitudes, y esas longitudes son proporcionales a las varianzas de las variables originales.

### En Resumen....

#### Objetivos de un ACP:

- Estudiar o explorar variabilidad entre observaciones (teniendo en cuenta todas las variables)
- Estudiar correlaciones entre variables
- Identificar variables de mayor contribución (de mayor peso) en la explicación de la variabilidad de las observaciones

### Resumiendo....

- La CP1 permite visualizar más variabilidad en los datos que cualquier otra CP.
- La CP2 no esta correlacionada con la CP1 (aporta nueva información) y explica mayor variabilidad que cualquier otra CP que no sea la CP1.
- Un gráfico de dispersión construido a partir de la CP1 y la CP2 proyecta la nube de datos en el sentido de máxima variación. Ideal para estudiar variación.

### Pasos a seguir para interpretar un Biplot

#### 1. Observar el porcentaje de variabilidad total explicado por el Biplot.

Si el Biplot conformado por las CP1 y CP2 no explica más del 60% de la variabilidad total, juzgar la necesidad de explorar los patrones de variabilidad en un segundo Biplot conformado por las CP1 y CP3. Si son necesarios muchos Biplot para explicar un porcentaje razonable de la variabilidad total, digamos mayor a 60-70%, habrá indicios de que el ACP no es suficiente para representar confiablemente las relaciones entre los casos y las variables (Arroyo et al., 2005).

### Pasos a seguir para interpretar un Biplot

#### 2. Concentrarse en la CP1 que por construcción, siempre explicará el mayor porcentaje de variabilidad total.

- a) Analizar las proyecciones perpendiculares a la CP1 de los puntos que representan las observaciones. Identificar las observaciones de mayor inercia, es decir los puntos que se encuentran más a la derecha o más a la izquierda. Interpretar "similitudes/disimilitudes" entre las observaciones

### Pasos a seguir para interpretar un Biplot

2. Concentrarse en la CP1 que por construcción, siempre explicará el mayor porcentaje de variabilidad total.

b) Analizar las proyecciones de los puntos que representan las variables sobre la CP1.

- Identificar las variables de mayor inercia.
- Interpretar "correlaciones" entre variables según los ángulos de los vectores que los representan.

Nota: La longitud de los vectores correspondientes a las variables no son de interés cuando los datos han sido previamente estandarizados. Si no se estandarizan los datos, las longitudes de los vectores son proporcionales a las varianzas de las variables.

### Pasos a seguir para interpretar un Biplot

2. Concentrarse en la CP1 que por construcción, siempre explicará el mayor porcentaje de variabilidad total.

c) Interpretar asociaciones entre observaciones y variables en función de la orientación, pero no de la cercanía entre puntos filas y columnas, es decir las variables orientadas hacia la derecha tendrán altos valores en las observaciones orientadas en la misma dirección y las variables orientadas hacia la izquierda tendrán altos valores en las observaciones orientadas hacia la izquierda.

**La distancia entre símbolos representando observaciones y símbolos representando variables no tiene interpretación.**

### Pasos a seguir para interpretar un Biplot

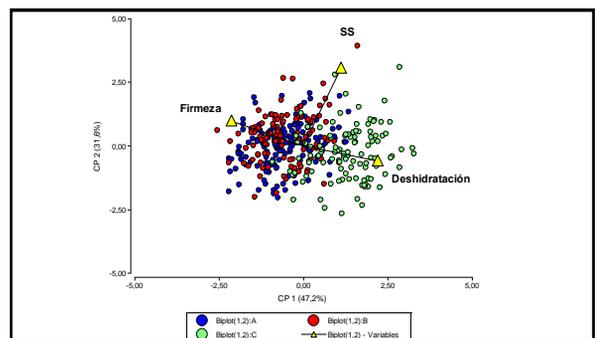
3. Concentrarse en la CP2 y realizar las interpretaciones siguiendo un procedimiento análogo al realizado para la CP1 pero teniendo en cuenta que las variables en esta dimensión son de menor importancia que los realizados sobre la CP1 según indican los porcentajes de variabilidad total explicados por cada CP.

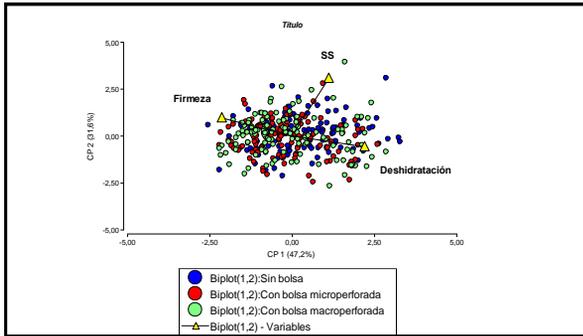
### Algunas aplicaciones del

### Análisis multivariado de Componentes Principales

### Ensayo Arándanos

Con el objetivo de mejorar la condición final de la fruta de arándanos, se montó un ensayo de postcosecha con tres variedades de arándanos que fueron manejados con tres diferentes tipos de bolsa en la caja embalada.





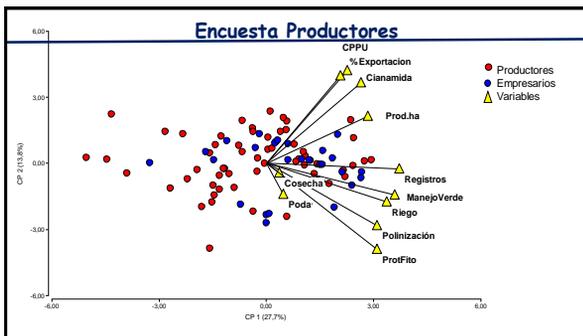
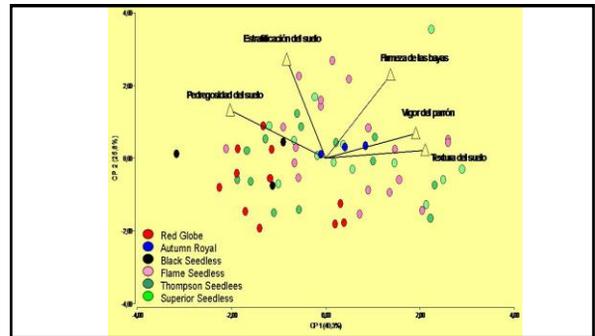
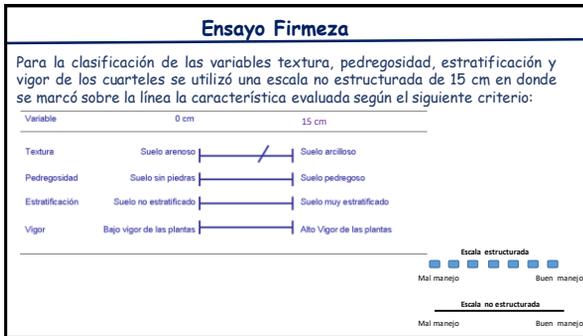
### Ensayo Firmeza

Objetivo: Evaluar la firmeza de las bayas de uva de mesa y establecer relaciones con la textura, pedregosidad, estratificación del suelo y vigor de las plantas.

El ensayo se realizó en 42 cuarteles del Valle del Huasco y 21 cuarteles del Valle de Copiapó, de las variedades Flame Seedless, Black Seedless, Thompson Seedless, Red Globe, Superior Seedless y Autumn Royal.

En cada cuartel se realizó una calicata representativa del sector, en la cual se determinó la textura, pedregosidad y estratificación del suelo.

Además en cada cuartel se determinaron las características de vigor de las plantas y firmeza de la fruta (en una muestra de 100 bayas al azar evaluadas con el instrumento medidor de firmeza Firmtech2®).



### Archivo Añjes

Se realizó un estudio con 21 variedades de añjes, en donde se midieron características de: altura de planta, materia seca, sólidos solubles, número de semillas por fruto, diámetro ecuatorial, longitud del fruto, peso del fruto y número de frutos por planta.

### Preguntas

Para analizar la variabilidad entre las variedades y las correlaciones entre las variables realice un ACP, estandarizando los datos, y presente los resultados mediante un gráfico Biplot

- ¿Cuál es el porcentaje de variabilidad total explicado por las dos primeras componentes?
- Identifique las observaciones de mayor inercia en la CP1. Interprete similitudes/disimilitudes entre las variedades
- Identifique y mencione las 4 variables de mayor importancia en el eje 1 (CP1). Interprete correlaciones entre las variables.
- ¿Qué conclusiones puede inferir respecto a la caracterización de las variedades? Interprete asociaciones entre observaciones y variables. ¿Cómo se caracterizan las variedad 20 y 21?

### Materiales y Método

Con el objetivo de estudiar o explorar variabilidad entre observaciones (variedades), estudiar las correlaciones entre las variables e identificar las variables de mayor contribución (de mayor peso) en la explicación de la variabilidad de las observaciones, se realizó un Análisis Multivariado de Componentes Principales

Dado que las variables están en diferentes unidades de medida, y por lo tanto sus medias y varianzas no son comparables, es que se procedió a estandarizar los datos, es decir se trabajó sobre la matriz de correlación R.

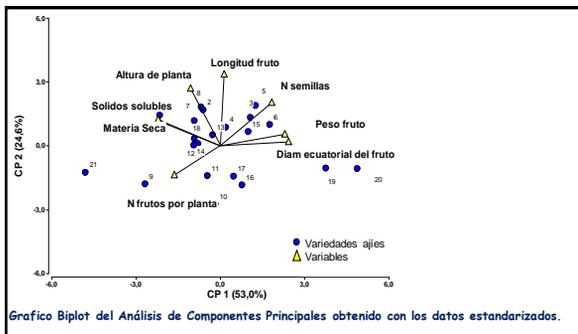
Estos resultados se muestran mediante un gráfico Bi-plot creados con el software estadístico InfoStat (Di Rienzo et al; 2017).

### Materiales y Método

El análisis de componentes principales (ACP) y los gráficos conocidos como Biplot son técnicas generalmente utilizadas para la reducción de la dimensión.

Las técnicas de reducción de dimensión permiten examinar todos los datos en un espacio de menor dimensión que el espacio original de las variables. Con el ACP se construyen ejes artificiales (Componentes Principales), que permiten obtener gráficos de dispersión y visualizar observaciones y variables en un mismo espacio con propiedades óptimas para la interpretabilidad (el prefijo "Bi" en el nombre Biplot refleja esta característica). Así es posible identificar asociaciones entre observaciones, entre variables y entre variables y observaciones.

Las variables fueron graficadas como vectores desde el origen (con terminaciones en triángulos amarillos) y las observaciones fueron graficadas con terminaciones en círculos azules según variedad.



### Presentación de Resultados

### Archivo Entrenudos

Se usaron entrenudos de cargadores y brotes de 15 cm de largo de la variedad Flame Seedless provenientes de la III Región y de la Zona Central del país

La unidad de observación correspondió a un cuartel

Objetivo:

- Relacionar la composición química del entrenudo de la vid con la composición química de los brotes pequeños (15 cm)
- Realizar un estudio de la variabilidad de las zonas considerando todas las variables analizadas.

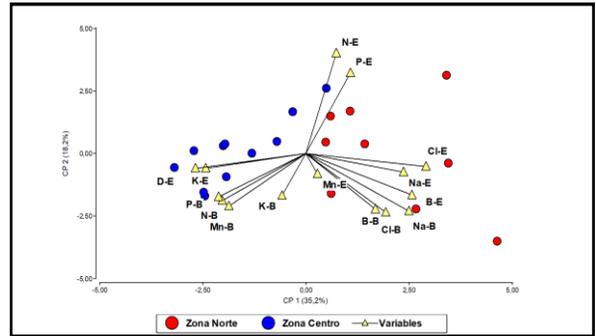
### Archivo Entrenados

Variables evaluadas:

En **entrenado**: diámetro de entrenudo (D-E), nitrógeno (N-E), fósforo (P-E), potasio (K-E), cloruro (Cl-E), boro (B-E), sodio (Na-E) y manganeso (Mn-E);

En **brotes (15 cm)**: se evaluó nitrógeno (N-B), fósforo (P-B), potasio (K-B), cloruro (Cl-B), boro (B-B), sodio (Na-B) y manganeso (Mn-B).

Zona	N-E	P-E	K-E	Cl-E	B-E	Na-E	Mn-E	D-E	N-B	P-B	K-B	Cl-B	B-B	Na-B	Mn-B
Norte															
Norte															
Norte															
Norte															
Norte															
Centro															
Centro															
Centro															
Centro															
Centro															



### Archivo Iris

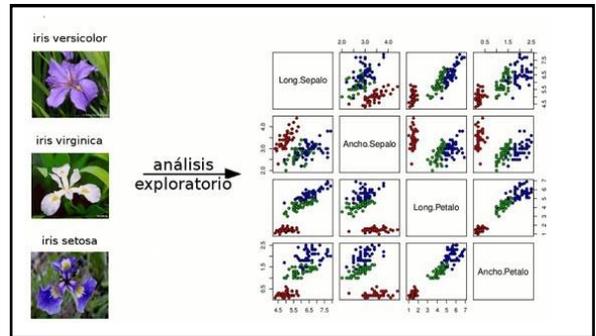


Los datos corresponden a 50 observaciones de 4 características de una flor para 3 especies del género Iris (Fisher, 1936), siendo el total de observaciones 150.

Descripción:

Iris: Especie de Iris  
 SepalLen: Longitud del sépalo  
 SepalWid: Ancho del sépalo  
 PetalLen: Longitud del pétalo  
 PetalWid: Ancho del pétalo

Aquí hay una clasificación definida a priori



### Archivo Contaminación Mercurio

La contaminación por Mercurio (Hg) de peces de agua dulce comestibles es una amenaza directa contra nuestra salud. Entre 1990-1991 se llevó a cabo un estudio en 53 lagos del Estado de Florida, con el fin de examinar los factores que influyen en el nivel de contaminación por Hg. Las variables que se midieron fueron:

Lago: Nombre del lago  
 Alcalinidad: Alcalinidad (mg/L de carbonato de calcio)  
 pH: pH  
 Calcio: Calcio (mg/L)  
 Clorofila: Clorofila (mg/L)  
 Mercurio: Concentración media de Hg (ppm) en el tejido muscular del grupo de peces estudiados en cada lago

### Archivo Cráneos Egipcios

Los datos corresponden a 4 medidas sobre cráneos de varones egipcios de cinco periodos históricos distintos

(Grupo 1: 4000 aC, Grupo 2: 3300 aC, Grupo 3: 1850 aC, Grupo 4: 200 aC, Grupo 5: 150 dC).

Para cada periodo temporal se midieron 30 cráneos. Las variables observadas son:

Ancho cráneo: Anchura máxima  
 Altura cráneo: Altura basibregmática  
 Long mandíbula: Longitud basialveolar (mandíbula)  
 Long nariz: Longitud de la nariz

### Archivo Gorriones

Se tienen las medidas de 5 variables biométricas sobre gorriones hembra, recogidos casi moribundos después de una tormenta. Los primeros 21 sobrevivieron mientras que los 28 restantes no lo consiguieron. Las variables evaluadas fueron:

Largo: Longitud total  
 Ext ala: Extensión del ala  
 L pico: Longitud del pico y de la cabeza  
 L húmero: Longitud del húmero  
 L esternón: Longitud del esternón

### Archivo Indicadores económicos y sociales

Los datos corresponden a 8 indicadores económicos y sociales de 96 países. Las variables evaluadas fueron:

Tmort inf: Tasa de mortalidad infantil por cada 1000 nacidos vivos  
 Porc mujer: Porcentaje de mujeres en la población activa  
 Prod electricidad: Producción de electricidad (millones de kW/h)  
 Lin telefonicas: Líneas telefónicas por cada 1000 habitantes  
 Consumo agua: Consumo de agua per cápita  
 Cobertura bosque: Proporción de la superficie del país cubierta por bosques  
 Consumo energía: Consumo de energía per cápita  
 Emision CO2: Emisión de CO2 per cápita

### Archivo Empleos

Los datos corresponden a los porcentajes de empleo en distintos sectores laborales para un conjunto de países Europeos. Las columnas del archivo correspondientes a los sectores laborales son:

AGR: agricultura  
 MIN: minería  
 MAN: manufactura  
 PS: previsión social  
 SER: servicios  
 FIN: finanzas  
 SPS: Seguros  
 TC: transporte y comunicación.

### Archivo Mamíferos

Los datos corresponden a la cantidad de dientes según tipo en distintos mamíferos.

Descripción:  
 Incisor\_S= cantidad de Incisivos superiores  
 Incisor\_I= cantidad de incisivos inferiores  
 Colmillo\_S= cantidad de colmillos superiores  
 Colmillo\_I= cantidad de colmillos inferiores  
 Premolar\_S= Cantidad de premolares superiores  
 Premolar\_I= cantidad de premolares inferiores  
 Molar\_S= cantidad de molares superiores  
 Molar\_I= cantidad de molares inferiores



FACULTAD DE CIENCIAS  
 AGRONÓMICAS  
 UNIVERSIDAD DE CHILE

Diplomado: Análisis Estadístico  
 para Estudios Agropecuarios

**Análisis Multivariado**  
**Análisis de Componentes Principales**  
**(ACP)**

Módulo 4  
 Análisis Multivariado

Erika Kanla Kuhl  
 Ing. Agr. Dr.