

# Estadística IV

13/11/2024

Segundo Semestre 2024

Profesor: Kevin Carrasco

Daniela Olivares

Ayudante: María Fernanda Núñez

Universidad Alberto Hurtado

Departamento de Sociología

# **Análisis de Conglomerados**

## **(Análisis de tipologías o clusters)**

\*\*\*\**Bibliografía recomendada*\*\*\*\*

- Cea D' Ancona, M. A. 2002. *Análisis multivariable: teoría y práctica*. Madrid: Ed. Síntesis (Cap 3, Análisis de Conglomerados)
- Hair J, F. et al (2004): *Análisis multivariante*. Madrid: Pearson (Cap. "Análisis de conglomerados")

# Antecedentes y objetivos

- Sentido genérico: técnica enmarcada en el “Análisis de tipologías” (cluster analysis, Taxonomía Numérica, Análisis Q, etc.)
- Reúne un conjunto de procedimientos orientados a **agrupar casos** (individuos u objetos) en función de una serie de variables de clasificación
- **Objetivo:**
  - agrupar objetos o individuos en grupos o conglomerados (clusters) de modo tal de que los objetos en el mismo grupo se parezcan más entre sí que otros pertenecientes a otro grupo

# Antecedentes y objetivos

- *Idea clave:* maximizar la homogeneidad de los objetos dentro de un grupo y maximizar la heterogeneidad entre los grupos.
- Con ello se busca identificar un patrón de agrupamiento subyacente a los casos.

# Funcionamiento

- El análisis de conglomerados requiere 6 tareas fundamentales:
  1. Resolución de aspectos previos: diseño de investigación, revisión de supuestos y elección de las variables y su forma de medición
  2. **Medición de la similitud** entre los objetos o sujetos (casos): hay diferentes formas que llevan a resultados diferentes
  3. **Formación de los grupos: dos grandes procedimientos**
    - Procedimientos jerárquicos
    - Procedimientos no-jerárquicos
  4. Toma de decisión: ¿con cuántos grupos quedarse?
  5. Caracterización de los grupos
  6. Validación de los grupos

# Funcionamiento

- El análisis de conglomerados requiere 6 tareas fundamentales:
  1. ...
  2. **Medición de la similitud** entre los objetos o sujetos (casos): hay diferentes procedimientos que llevan a resultados diferentes

## 2. Medidas de similitud entre los objetos

- Dependiendo del tipo de investigación y variables, existen tres tipos de medidas de similitud
  1. Medidas de asociación como Chi-cuadrado...
  2. Correlación...
  3. **Medidas de distancia:** son las más usadas para medir similitud. Existen varias medidas de distancia.



## 2. Medidas de similitud entre los objetos

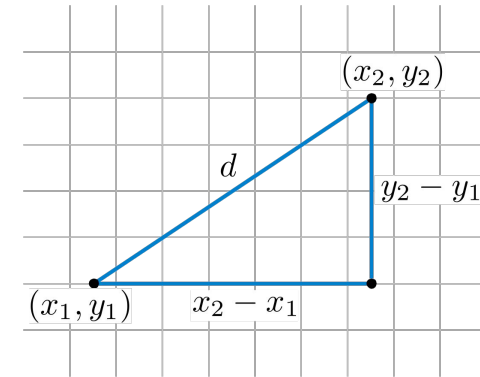
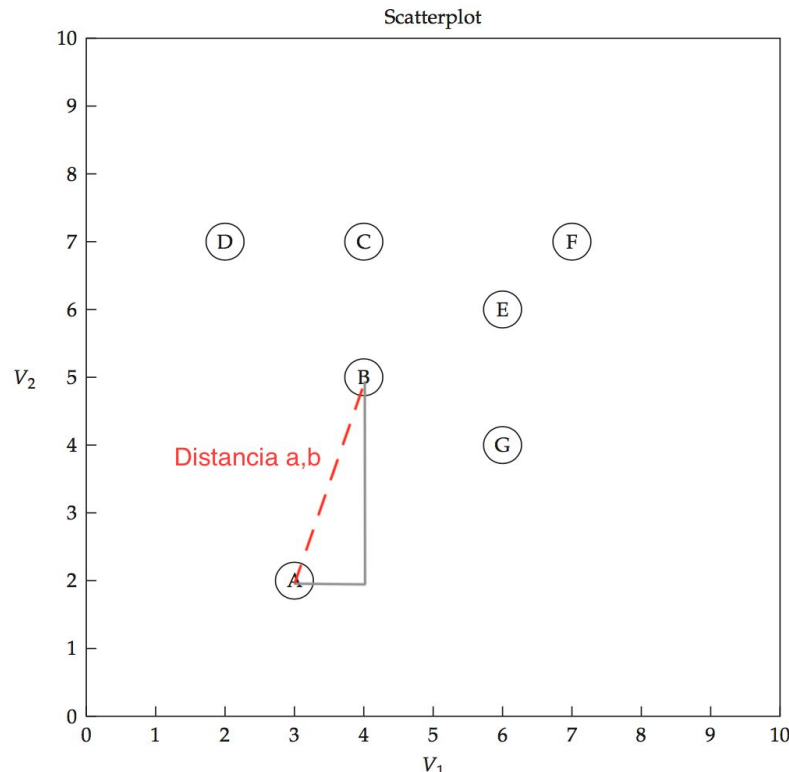
- **Medidas de distancia (similitud)**

1. **Distancia euclídea:** una de las más usadas en la investigación empírica, define similitud a través de una línea recta (cercanía en un plano)

## 2. Medidas de similitud entre los objetos

- Medidas de distancia (similitud)

1. **Distancia euclídea:** una de las más usadas en la investigación empírica, define similitud a través de una línea recta (cercanía en un plano)



$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## 2. Medidas de similitud entre los objetos

- **Medidas de distancia (similitud)**

1. **Distancia euclídea...**
2. **Distancia euclídea al cuadrado:** se desarrolló para hacer las cosas más rápidas; estadísticamente es lo mismo que la anterior
3. **Distancia de Manhattan....**
4. **Distancia de Mahalanobis....**

# III. Funcionamiento

- El análisis de conglomerados requiere 6 tareas fundamentales:

1. ...

2. ...

3. **Formación de los grupos: dos grandes procedimientos**

- Procedimientos jerárquicos: revisados semana pasada

- Procedimientos no-jerárquicos: esta clase

# Métodos de aglomeración jerárquicos

- Clasifican y agrupan los casos por etapas mediante un proceso en forma de “árbol” (dendograma)
  - Originalmente, hechos para variables métricas, también se usan con variables ordinales
  - Ventajas: permiten observar gráficos de aglomeración (dendogramas) y matrices de distancia □ especialmente útiles con n pequeños
- Existen 2 tipos de procedimientos jerárquicos:
  1. **Aglomerativos (más usado)**
  2. Divisivos

# Métodos de aglomeración jerárquicos: algoritmos

- **Dentro de los métodos de aglomeración jerárquicos, existen diversos *algoritmos de agrupamiento***
  - Procedimientos que definen cuándo los grupos se van a combinar entre sí
  - Ellos, por tanto, entregan información sobre cuántos grupos escoger como “solución óptima”

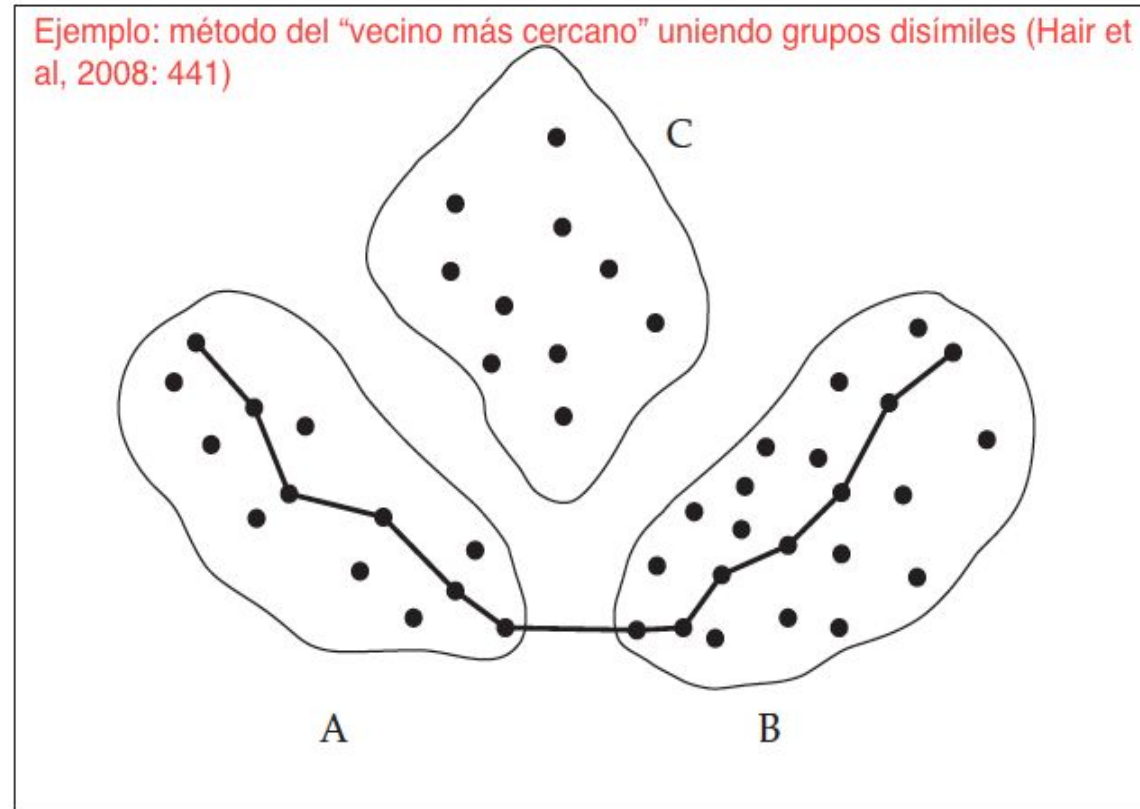
# 3. Métodos de aglomeración jerárquicos: algoritmos

## 1. Método del “Vecino más cercano” (single-linkage)

- Similitud entre grupos = distancia más corta entre el objeto de uno y otro grupo (ver ejemplo)
- Problema: riesgo de encadenamiento

### 3. Métodos de aglomeración jerárquicos: algoritmos

- Problema: riesgo de encadenamiento



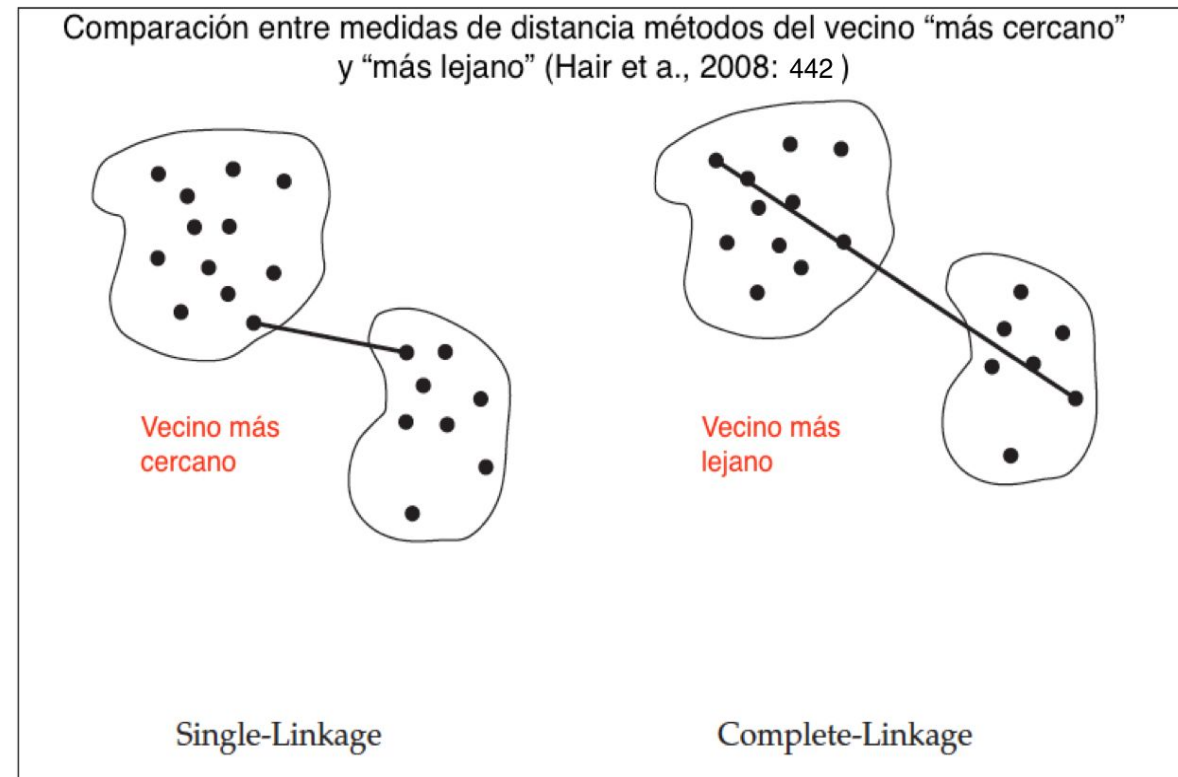


# 3. Métodos de aglomeración jerárquicos: algoritmos

## 2. Método del “Vecino más lejano” (complete-linkage)

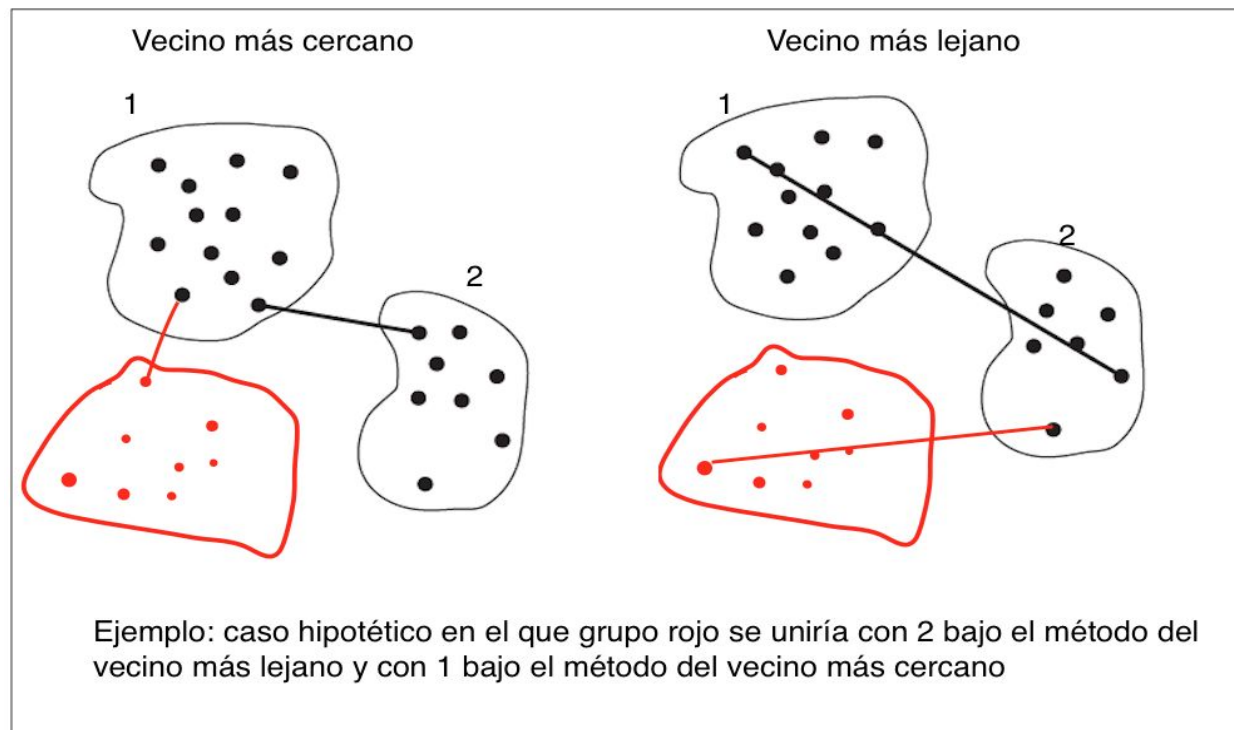
- Similitud entre grupos:
  - Análisis las distancias máximas entre las observaciones de cada grupo
  - Se unen grupos cuyas distancias “más lejanas” son las menores

### 3. Métodos de aglomeración jerárquicos: algoritmos



### 3. Métodos de aglomeración jerárquicos: algoritmos

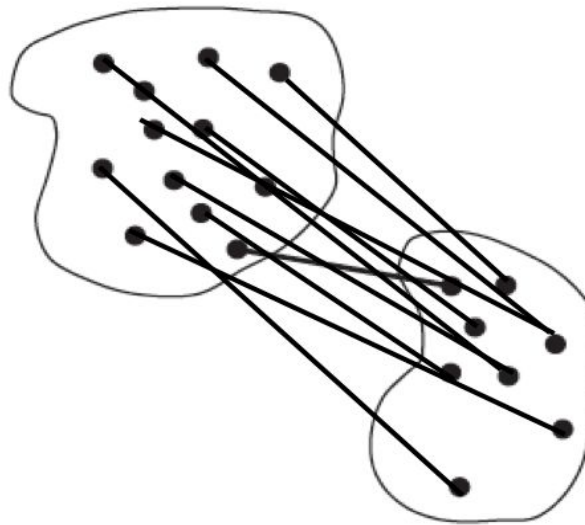
Diferencias entre métodos del "Vecino más cercano" y "Vecino más lejano"



# 3. Métodos de aglomeración jerárquicos: algoritmos

## 3. Promedio de las distancias entre grupos (average linkage between groups)

- Se calcula la distancia promedio de *todos* los integrantes de un grupo respecto de los integrantes de otro grupo; dos grupos se combinan cuando el promedio distancia es la menor posible (algoritmo definido por defecto en algunas versiones de SPSS)



Promedio de las distancias entre grupos

# 3. Métodos de aglomeración jerárquicos: algoritmos

## 4. Promedio intragrupal

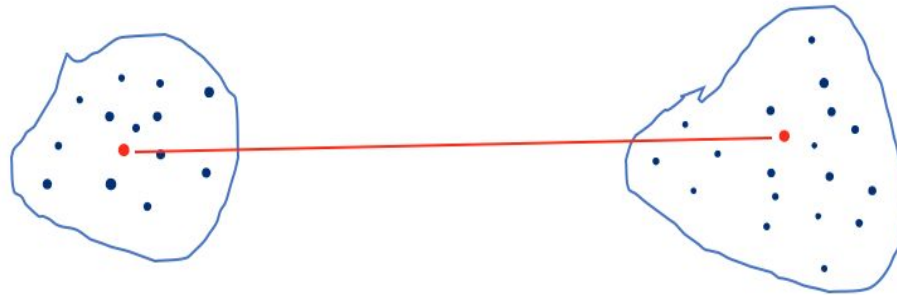
- Variante del algoritmo anterior
- Se parte combinando de dos en dos los grupos, de manera sucesiva.
- Luego se calcula la distancia promedio entre los miembros de esos grupos.
- Se mantienen sólo los grupos cuya distancia promedio intragrupal es la menor

# 3. Métodos de aglomeración jerárquicos: algoritmos

## 5. Método del centroide

- Similitud = distancia entre centroides (valores promedios) de los grupos
- Cálculo: cada vez que los casos son reagrupados, un nuevo centroide se reclacula

Método de centroide

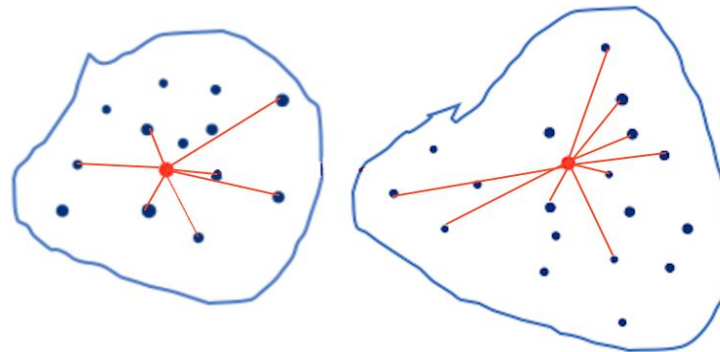


# 3. Métodos de aglomeración jerárquicos: algoritmos

## 6. Método de Ward

- Se calculan las distancias al cuadrado de cada caso a la media de su grupo al cuadrado
- Luego se calcula la *suma total* de dichas distancias
- Dos grupos se unen cuando esa suma total sea la menor posible, es decir, cuando se cumpla el criterio de minimización de “varianza” (suma de distancias al cuadrado), asumiendo que *siempre* el valor de las distancias al cuadrado va a aumentar

Método de Ward



# 3. Métodos de aglomeración jerárquicos: Ejemplo

## Ejemplo

- Problema de investigación: ¿Es posible agrupar países según sus modalidades de relaciones laborales?
- Casos: 45 países
- Base de datos: Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts (ICTWWS) de Jelle Visser

## *Variables y medición*

1. Nivel de centralización de las relaciones laborales: valor teórico 0 - 100
2. Derecho a sindicalización (índice): valor teórico 0 - 100
3. Derecho a negociación colectiva (índice): valor teórico 0 - 100
4. Derecho a huelga (índice): valor teórico 0 - 100
5. % Cobertura negociación colectiva: valor teórico 0 - 100

Variable relevante excluida: intervención del estado en relaciones laborales (multicolinealidad)



### 3. Métodos de aglomeración jerárquicos: ejemplo

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
1. Centralization	45	,00	100,00	30,0000	25,89314
2. Right of association	45	33,33	100,00	88,5185	15,00094
3. Right of collective bargaining	45	33,33	100,00	74,4444	19,00292
4. Right to strike	45	33,33	100,00	64,8148	19,85497
5. Bargaining coverage rate	45	6,00	98,00	44,5647	29,06477
Valid N (listwise)	45				

### 3. Métodos de aglomeración jerárquicos: ejemplo

- Análisis de conglomerado jerárquico
  - Medida de distancia: distancia euclídea al cuadrado
  - Algoritmo de agrupamiento: Ward
  - Especificación rango de soluciones: 2 a 4

# 3. Métodos de aglomeración jerárquicos: ejemplo

Coef. algoritmo de Ward  
(suma al cuadrado de distancias)

Matriz de proximidad (extracto)

Proximity Matrix

Case	Squared Euclidean Distance									
	1:Argentina	2:Australia	3:Austria	4:Belgium	5:Brazil	6:Bulgaria	7:Canada	8:Chile	11:Cyprus	12:Czech Republic
1:Argentina	,000	3691,744	3855,214	8286,682	1669,395	4306,568	5347,462	5656,947	4289,090	4164,789
2:Australia	3691,744	,000	3403,598	9696,686	2038,857	1518,358	2975,993	3212,773	3275,835	1985,078
3:Austria	3855,214	3403,598	,000	3632,472	2076,877	6167,250	9499,031	10561,164	6695,439	6737,685
4:Belgium	8286,682	9696,686	3632,472	,000	4292,730	12336,222	15320,826	18561,709	8484,716	10584,003
5:Brazil	1669,395	2038,857	2076,877	4292,730	,000	2756,196	3797,053	5253,556	1574,127	1998,419
6:Bulgaria	4306,568	1518,358	6167,250	12336,222	2756,196	,000	1458,333	1020,652	2762,026	2157,616
7:Canada	5347,462	2975,993	9499,031	15320,826	3797,053	1458,333	,000	673,674	1442,218	768,316
8:Chile	5656,947	3212,773	10561,164	18561,709	5253,556	1020,652	673,674	,000	3578,611	2394,964
11:Cyprus	4289,090	3275,835	6695,439	8484,716	1574,127	2762,026	1442,218	3578,611	,000	438,470
12:Czech Republic	4164,789	1985,078	6737,685	10584,003	1998,419	2157,616	768,316	2394,964	438,470	,000
13:Denmark	6692,795	5591,967	3807,111	1828,028	2182,934	8042,361	8596,680	12132,280	3468,237	4680,606
14:Estonia	8852,278	4773,929	11753,472	15329,000	5099,507	2883,222	869,468	2523,590	1395,051	1302,042
15:Finland	8395,064	8604,444	4261,111	443,028	3855,469	11613,361	13417,477	17148,706	6747,867	8547,300
16:France	4688,547	4653,598	3125,000	4188,028	1799,099	7417,250	6721,254	10005,609	2945,439	3473,796
17:Germany	3245,144	2523,261	2187,716	3436,366	489,171	2779,766	5001,345	6296,436	2393,660	3161,753

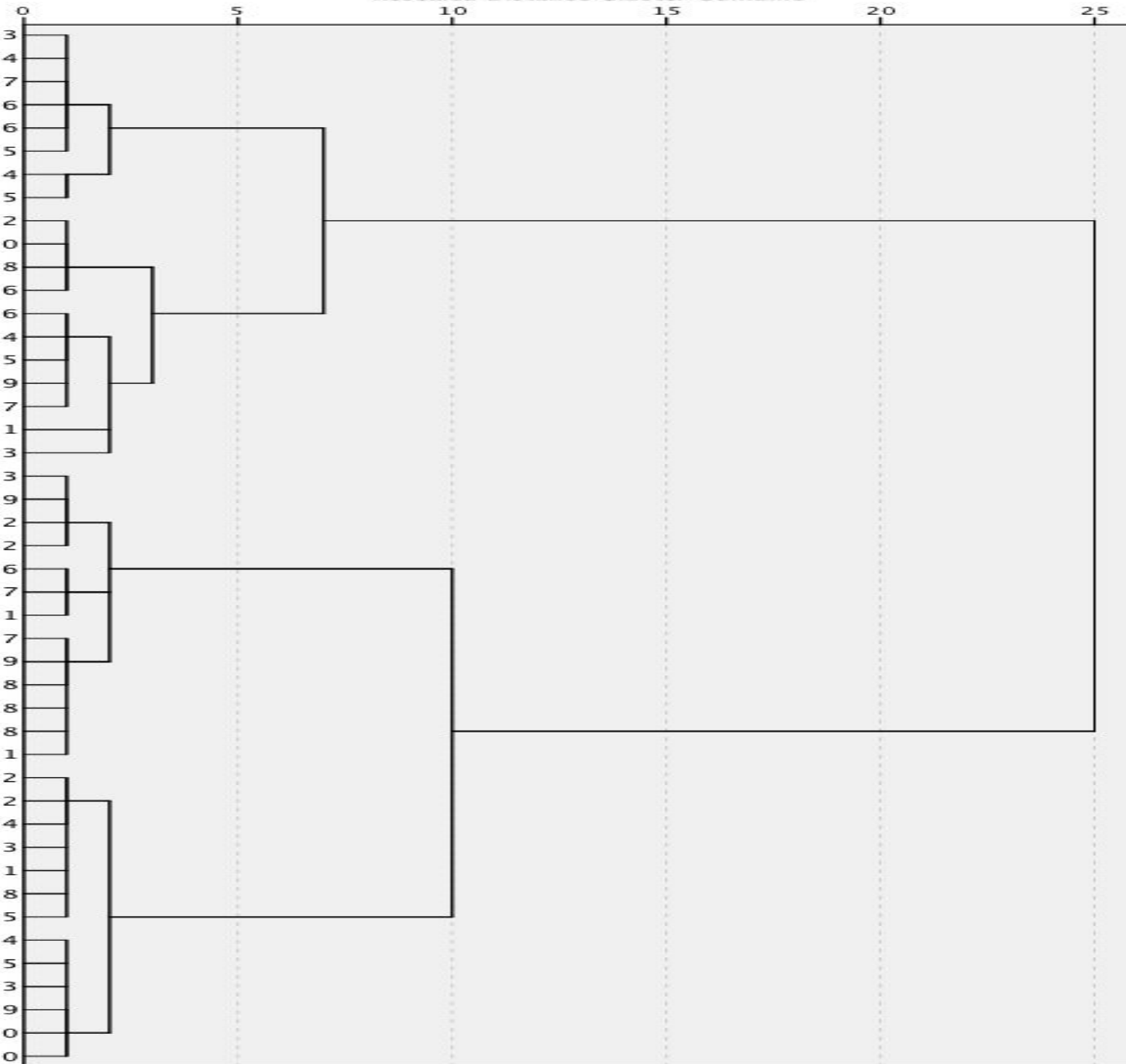
Historial de aglomeración (extracto)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	13	34	,350	0	0	3
2	8	28	5,289	0	0	23
3	13	47	19,283	1	0	7
4	14	35	48,928	0	0	5
5	14	23	165,970	4	0	8
6	11	18	309,938	0	0	22
7	13	46	479,634	3	0	30
8	14	19	688,211	5	0	26
9	24	43	905,950	0	0	27
10	5	39	1124,913	0	0	24
11	36	44	1343,927	0	0	32
12	4	15	1565,441	0	0	37
13	26	27	1796,864	0	0	20
14	33	49	2030,158	0	0	19
15	16	25	2270,283	0	0	30
16	10	22	2522,122	0	0	21

**Dendrogram using Ward Linkage**

Rescaled Distance Cluster Combine



Dendrograma

## 4. Definición número de grupos: ejemplo

Caracterización/Validación de grupos: ¿tiene sentido formar 2, 3 o 4 grupos? ¿A qué grupo pertenece cada país?

Case	Cluster Membership		
	4 Clusters	3 Clusters	2 Clusters
1:Argentina	1	1	1
2:Australia	1	1	1
3:Austria	1	1	1
4:Belgium	2	1	1
5:Brazil	1	1	1
6:Bulgaria	1	1	1
7:Canada	3	2	2
8:Chile	3	2	2
11:Cyprus	4	3	2
12:Czech Republic	4	3	2
13:Denmark	2	1	1
14:Estonia	4	3	2
15:Finland	2	1	1
16:France	2	1	1
17:Germany	1	1	1
18:Greece	4	3	2
19:Hungary	4	3	2
21:India	2	2	2

**Grupo 1:** Argentina, Australia, Austria, Bélgica, Alemania, Suecia, etc. □ relaciones “corporativistas”

**Grupo 2:** Canadá, Chile, Corea de Sur, Singapur, USA, etc. □ relaciones liberales

**Grupo 3:** Chipre, República Checa, Estonia, Grecia, Hungría, etc. □ intermedios

## 2 pasos restantes

### 5. Caracterización de los conglomerados

- ¿Cómo definir los grupos formados?
- ¿Qué características tienen sus miembros?

### 6. Validación de los conglomerados

- Idea clave: si los grupos formados son “reales”, entonces ellos deberían generar diferencias significativas en otras variables no utilizadas para generarlo (ej. ingresos, orientaciones políticas, etc.)

# 5-6. Caracterización/validación de grupos: ejemplo

## Caracterización/Validación de grupos:

Ejemplo: solución de 3 grupos.

- Cruce con variable “tasa de sindicalización”

		Union density rate		
		Count	Mean	Standard Deviation
Ward	1	19	32,33	19,81
Method	2	13	13,64	6,09
	3	13	24,14	13,49
Total		45	24,57	16,78

## ANOVA

Union density rate, net union membership as a proportion of wage earners in employment (Num\*100/WSEE)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2698,845	2	1349,423	5,846	,006
Within Groups	9695,372	42	230,842		
Total	12394,217	44			

**Grupo 1:** relaciones “corporativistas”

**Grupo 2:** relaciones liberales

**Grupo 3:** países intermedios

### 3. Método de aglomeración no jerárquico

- Procedimientos que no implican la formación sucesiva de grupos en “forma de árbol”
- **El número de grupos lo especifica de antemano quien investiga**
  - Los objetos (casos) no se van agrupando entre sí, sino que se van asignando a los grupos de modo sucesivo
- **Algoritmos de aglomeración:** hay varios, pero lejos el más usado es el de K-medias (SPSS)



# 3. Método de aglomeración no jerárquico

- **Método de K-medias**

1. Investigador/a define número de grupos
2. Se designan centroides arbitrarios (semillas de los conglomerados)
3. A través de diversas iteraciones, los casos se van agrupando en torno a esos centroides (cada vez que se asigna un caso, se calcula un nuevo centroide y se reasignan los casos)
4. El procedimiento se detiene hasta que se cumple el *requisito de minimizar las distancias entre las observaciones dentro del grupo y maximizar las distancias entre los grupos*

- A pesar de que puede funcionar con variables de distinta medida, se recomienda estandarizar cuando hay unidades de medida muy diversas

### 3. Método de aglomeración no jerárquico

- **Método de K-medias: ventajas**

- Comparativamente, sus resultados están menos afectados por outliers, unidades de medidas e inclusión de variables irrelevantes
- Es ideal cuando se trabaja con  $n$  grande (no se necesita de dendogramas ni de matrices de similitud; sólo requiere cálculo de distancias caso – centroide)

- **Método de K-medias: desventajas**

- Requiere de conocimiento previo para la formación de grupos

### 3. Método de aglomeración no jerárquico: ejemplo

- **Problema de investigación:** ¿Es posible identificar grupos de la población según su disposición a la acción colectiva?
- Base de datos: Latinobarómetro 2015 (sólo casos de Chile, 18 años y más); n = 1.200 (1.131 finalmente considerados)
- Variables utilizadas para formar grupos: 4 respuestas a la pregunta “¿Cuán dispuesto estaría Ud. de salir a marchar y protestar por... ? (respuestas de 1 a 10)
  1. Aumento de salario y mejores condiciones de trabajo
  2. Mejoras a la salud y educación
  3. La explotación de los recursos naturales
  4. La defensa de los derechos democráticos

### 3. Método de aglomeración no jerárquico: ejemplo

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
1. Aumento de salario y mejores condiciones de trabajo	1150	1	10	6,76	3,450
2. La mejora de la salud y la educación	1152	1	10	6,97	3,409
3. La explotación de recursos naturales	1145	1	10	6,41	3,363
4. Defender los derechos democráticos	1135	1	10	5,99	3,394
Valid N (listwise)	1131				

### 3. Método de aglomeración no jerárquico: ejemplo

- Análisis de K-medias
- Se decidió formar 3 grupos
- Máximo de 10 iteraciones (opción predefinida)
  - Supuesto: si no se logra convergencia después de la iteración número 10, entonces no es posible formar grupos con los datos que se tiene)

### 3. Método de aglomeración no jerárquico: ejemplo

**Initial Cluster Centers**

	Cluster		
	1	2	3
1. Aumento de salario y mejores condiciones de trabajo	10	1	1
2. La mejora de la salud y la educación	10	1	10
3. La explotación de recursos naturales	5	1	10
4. Defender los derechos democráticos	1	1	10

Medias de los grupos iniciales  
(información no muy relevante)

### 3. Método de aglomeración no jerárquico: ejemplo

Historial de iteraciones: indica las medias de los grupos y cómo ellas cambian en cada paso

**Iteration History<sup>a</sup>**

Iteration	Change in Cluster Centers		
	1	2	3
1	5,631	2,589	7,962
2	,996	,177	,791
3	,538	,599	,060
4	,451	,263	,170
5	,322	,098	,156
6	,255	,148	,084
7	,114	,051	,044
8	,018	,021	,000
9	,000	,000	,000

a. [Convergence achieved due to no or small change in cluster centers.](#)

The maximum absolute coordinate change for any center is ,000. The current iteration is 9. The minimum distance between initial centers is 13,342.

### 3. Método de aglomeración no jerárquico: ejemplo

**“Center” = media de los grupos finales.** Permite chequear que las medias en cada variable difieran

#### Final Cluster Centers

	Cluster		
	1	2	3
1. Aumento de salario y mejores condiciones de trabajo	7	2	10
2. La mejora de la salud y la educación	7	2	10
3. La explotación de recursos naturales	6	2	9
4. Defender los derechos democráticos	5	2	9

Entonces, podemos decir que:

- Grupo 3 = activos
- Grupo 2 = inactivos
- Grupo 1 = intermedios



### 3. Método de aglomeración no jerárquico: ejemplo

#### ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
1. Aumento de salario y mejores condiciones de trabajo	5573,807	2	2,111	1128	2639,827	,000
2. La mejora de la salud y la educación	5579,030	2	1,871	1128	2982,255	,000
3. La explotación de recursos naturales	5289,866	2	2,007	1128	2636,223	,000
4. Defender los derechos democráticos	5322,351	2	2,095	1128	2540,353	,000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

ANOVA: En este caso se usa sólo para fines descriptivos (para ver la “relevancia” de cada variable en la conformación del conglomerado)

### 3. Método de aglomeración no jerárquico: ejemplo

**Number of Cases in each Cluster**

Cluster	1	331,000
	2	288,000
	3	512,000
Valid		1131,000
Missing		69,000

Chequear si hay algún grupo con muy pocos casos

### 3. Método de aglomeración no jerárquico: ejemplo

- ***Validación resultados. ¿Tienen sentido estos grupos?*** Recordar que:
  - Grupo 3 = activos
  - Grupo 2 = inactivos
  - Grupo 1 = intermedios
- Cruce con variable de interés: ¿Ha realizado o estaría dispuesto a realizar reclamos a través de redes sociales?
  - Hipótesis: los más activos tienen más disposición a realizar reclamos a través de redes sociales

# 3. Método de aglomeración no jerárquico: ejemplo

P21N.E Acciones que ha realizado: Hacer un reclamo a través de las redes sociales \* Cluster  
Number of Case Crosstabulation

% within Cluster Number of Case

	Cluster Number of Case			Total
	Grupo 1: intermedios	Grupo 2: inactivos	Grupo 3: activos	
P21N.E Hacer un reclamo a través de las redes sociales	9,6%	2,8%	12,7%	9,2%
Ha realizado Podría realizar	42,9%	15,0%	61,4%	44,0%
Nunca; bajo ninguna circunstancia	47,4%	82,2%	25,9%	46,8%
Total	100,0%	100,0%	100,0%	100,0%

### 3. Método de aglomeración no jerárquico: ejemplo

**P21N.E Acciones que ha realizado: Hacer un reclamo a través de las redes sociales \* Cluster**  
**Number of Case Crosstabulation**

% within Cluster Number of Case

		Cluster Number of Case			Total
		Grupo 1: intermedios	Grupo 2: inactivos	Grupo 3: activos	
P21N.E Hacer un reclamo a través de las redes sociales	Ha realizado	9,6%	2,8%	12,7%	9,2%
	Podría realizar	42,9%	15,0%	61,4%	44,0%
	Nunca; bajo ninguna circunstancia	47,4%	82,2%	25,9%	46,8%
Total		100,0%	100,0%	100,0%	100,0%

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	231,102 <sup>a</sup>	4	,000
Likelihood Ratio	245,079	4	,000
Linear-by-Linear Association	45,425	1	,000
N of Valid Cases	1093		

a. 0 cells (0,0%) have expected count less than 5.  
The minimum expected count is 26,43.

### 3. Comparación métodos jerárquicos / no jerárquicos

- ¿Qué método (jerárquico/no jerárquico) es mejor? Depende de los objetivos de la investigación, del tamaño de la muestra, del conocimiento previo, etc.

Procedimientos jerárquicos (aglomerativos)		Procedimientos no jerárquicos
Número de conglomerados final es decisión de quien investiga <i>después</i> de mirar gráficos y analizar medidas de heterogeneidad		Número de conglomerados es definido <i>antes</i> de ejecutar el análisis

### 3. Comparación métodos jerárquicos / no jerárquicos

- ¿Qué método (jerárquico/no jerárquico) es mejor? Depende de los objetivos de la investigación, del tamaño de la muestra, del conocimiento previo, etc.

Procedimientos jerárquicos (aglomerativos)		Procedimientos no jerárquicos
Número de conglomerados final es decisión del investigador <i>después</i> de mirar gráficos y analizar medidas de heterogeneidad		Número de conglomerados es definido <i>antes</i> de ejecutar el análisis
Los objetos se unen a los grupos sucesivamente <i>determinados</i> por la solución inicial (su pertenencia a un grupo <i>no</i> puede cambiar)		Los objetos se unen a los grupos en diferentes iteraciones (su pertenencia a un grupo puede cambiar)

### 3. Comparación métodos jerárquicos / no jerárquicos

- ¿Qué método (jerárquico/no jerárquico) es mejor? Depende de los objetivos de la investigación, del tamaño de la muestra, del conocimiento previo, etc.

Procedimientos jerárquicos (aglomerativos)		Procedimientos no jerárquicos
Número de conglomerados final es decisión del investigador <i>después</i> de mirar gráficos y analizar medidas de heterogeneidad		Número de conglomerados es definido <i>antes</i> de ejecutar el análisis
Los objetos se unen a los grupos sucesivamente <i>determinados</i> por la solución inicial (su pertenencia a un grupo <i>no</i> puede cambiar)		Los objetos se unen a los grupos en diferentes iteraciones (su pertenencia a un grupo puede cambiar)
Se usan cuando: <ul style="list-style-type: none"><li>• se quiere obtener una representación gráfica de la clasificación;</li><li>• interesa explorar la estructura de los datos</li><li>• se trabaja con un n relativamente bajo (<math>n \leq 200</math>)</li></ul>		Se usan cuando: <ul style="list-style-type: none"><li>• no interesa la representación gráfica</li><li>• no interesa explorar la estructura jerárquica de la clasificación</li><li>• se tiene una muestra demasiado grande (<math>n &gt; 200</math>)</li></ul>

¿Se pueden combinar ambos métodos?



### 3. Comparación métodos jerárquicos / no jerárquicos

- ¿Qué método (jerárquico/no jerárquico) es mejor? Depende de los objetivos de la investigación, del tamaño de la muestra, del conocimiento previo, etc.

Procedimientos jerárquicos (aglomerativos)		Procedimientos no jerárquicos
Número de conglomerados final es decisión del investigador <i>después</i> de mirar gráficos y analizar medidas de heterogeneidad		Número de conglomerados es definido <i>antes</i> de ejecutar el análisis
Los objetos se unen a los grupos sucesivamente <i>determinados</i> por la solución inicial (su pertenencia a un grupo <i>no</i> puede cambiar)		Los objetos se unen a los grupos en diferentes iteraciones (su pertenencia a un grupo puede cambiar)
Se usan cuando: <ul style="list-style-type: none"><li>• se quiere obtener una representación gráfica de la clasificación;</li><li>• interesa explorar la estructura de los datos</li><li>• se trabaja con un n relativamente bajo (<math>n &lt; 200</math>)</li></ul>		Se usan cuando: <ul style="list-style-type: none"><li>• no interesa la representación gráfica</li><li>• no interesa explorar la estructura jerárquica de la clasificación</li><li>• se tiene una muestra demasiado grande (<math>n &gt; 200</math>)</li></ul>

¿Se pueden combinar ambos métodos? Sí (por ejemplo comenzar con una técnica jerárquica y luego ejecutar una no jerárquica)

# **Ejemplo k-medias (y análisis factorial)**

# Evaluating Industrial Relations Systems of OECD Countries from 1993 to 2005: A Two-Dimensional Approach

*Dong-One Kim, Yoon-Ho Kim, Paula Voos, Hiromasa Suzuki and Young Doo Kim*

## Abstract

*This article uses both cross-sectional and longitudinal methods to evaluate the national industrial relations systems of 30 Organisation for Economic Co-operation and Development (OECD) countries from 1993 to 2005. We adopt a pluralistic view of industrial relations that gives equal weight to efficiency and equity, along with a general systems model consisting of input, process and output. We rank each country in terms of a combined score of efficiency and equity. We find that the 30 OECD countries can be separated into three distinct groups (high on both equity and efficiency; high on efficiency but low on equity; moderate on equity and low on efficiency), and that these groups exhibit considerable stability over time.*

## 1. Introduction

As economic globalization has accelerated and the need for social regulation has escalated, progress has been made in collecting and adding the economic and social indexes of different nations to international databases. Various international organizations, such as the Organisation for Economic Co-operation and Development (OECD), the International Labour Organization (ILO), the World Economic Forum (WEF), and the International Institute for Management Development (IMD) have undertaken this work. Scholars, too, have classified, measured and analysed the industrial relations of countries. Their activities have been prompted at times by various theoretical debates, such as the one over corporatism. Hence, in recent years,

Dong-One Kim and Young Doo Kim are at Korea University. Yoon-Ho Kim is at the Korea University of Technology and Education. Paula Voos is at Rutgers University. Hiromasa Suzuki is at Waseda University.

## Estudio de 30 países OCDE (datos entre 1993 y 2005)

Pregunta de investigación: ¿cómo los países combinan, a través de sus tipos de relaciones laborales, eficiencia económica y equidad social?

Hallazgo central: existen 3 tipos de países

1. Países con alta eficiencia y equidad
2. Países con alta eficiencia pero alta desigualdad
3. Países con baja eficiencia y niveles de equidad moderados

TABLE 2  
Results of Factor Analyses

<i>Input</i>	<i>Variables</i>	<i>Factor 1</i>	<i>Factor 2</i>
Efficiency-input	Freedom of dismissal	-0.282	<b>0.697</b>
	Freedom to employ temporary workers	-0.131	<b>0.775</b>
	Union decentralization	-0.388	<b>0.423</b>
	Expenditures on education	-0.006	<b>0.521</b>
	Literacy rate	0.299	<b>0.709</b>
Equity-input	Union density	<b>0.674</b>	0.146
	Ratification status of ILO fundamental conventions	<b>0.729</b>	-0.353
	Expenditure on ALMP	<b>0.828</b>	0.067
	Public social expenditures	<b>0.770</b>	0.253
	Centralization of wage-setting institution	<b>0.727</b>	-0.345
	Co-ordination of wage-setting institution	<b>0.747</b>	-0.242
	Maternity leave index	<b>0.523</b>	-0.265
Eigenvalue		4.27	2.20
Proportion of variance accounted for		35.54	18.32
<i>Process</i>	<i>Variables</i>	<i>Factor 1</i>	<i>Factor 2</i>
Efficiency-process	Industrial relations	0.042	<b>0.967</b>
	Labour-employer relations	-0.062	<b>0.939</b>
	Unit labour cost <sup>a</sup>	-0.492	<b>0.430</b>
Equity-process	Collective bargaining coverage	<b>0.860</b>	0.029
	National Tripartite Board or council for labour policies	<b>0.519</b>	-0.139
	Worker participation in management	<b>0.766</b>	-0.135
	Trade union rights index	<b>0.743</b>	0.282
Eigenvalue		2.48	2.04
Proportion of variance accounted for		35.40	29.11
<i>Output</i>	<i>Variables</i>	<i>Factor 1</i>	<i>Factor 2</i>
Efficiency-output	Labour productivity	-0.516	<b>0.418</b>
	Economic growth rates	-0.376	<b>0.583</b>
	Strikes and lock-outs <sup>a</sup>	0.074	<b>0.913</b>
	Workers involved <sup>a</sup>	0.125	<b>0.936</b>
Equity-output	Injuries: deaths <sup>a</sup>	<b>0.603</b>	0.002
	Average hours worked per person <sup>a</sup>	<b>0.809</b>	-0.242
	Gini coefficient <sup>a</sup>	<b>0.530</b>	-0.023
	Minimum relative to median wages of full-time workers	<b>0.460</b>	-0.339
	Ratio of estimated female-to-male earned income	<b>0.682</b>	-0.085
	Income distribution	<b>0.786</b>	-0.050
	Employment rates	<b>0.740</b>	0.133
	Unemployment replacement rates	<b>0.686</b>	-0.086
Eigenvalue		4.29	2.19
Proportion of variance accounted for		35.71	18.24

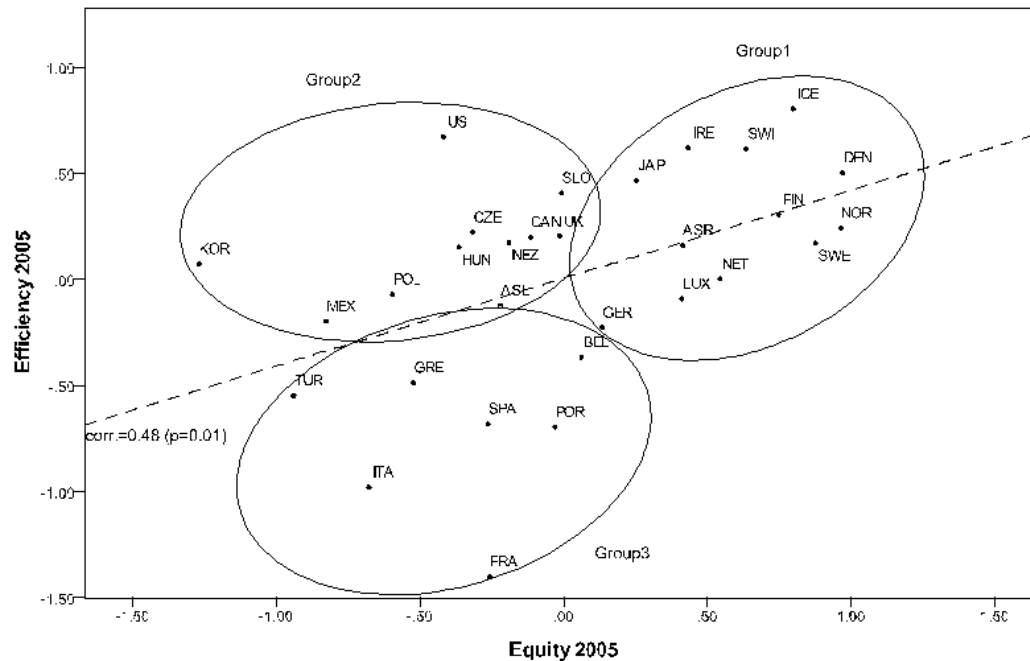
<sup>a</sup> Item was reverse-coded.

Note: Bold type indicates that the associated item loads at 0.40 or greater on a single factor.

ALMP, active labour market policies; ILO, International Labour Organization.

Figure 1 uses efficiency and equity along its two axes to illustrate industrial relations in 2005. Separate graphics, displaying patterns of industrial relations at the three specific points of time (i.e. 1993, 1999 and 2005) under study, show almost identical patterns so we illustrate only the final year.<sup>4</sup> The groups in the illustration were produced by K-means clustering analysis. Although the position of each country changes slightly over time, each country typically remains in the same group at each of the three points in time. In other words, national patterns in 1993 were very similar to those in

FIGURE 1  
Graphic Illustration of IR Systems of OECD Countries 2005.



*Notes:* Australia = ASL, Austria = ASR, Belgium = BEL, Canada = CAN, Czech Rep. = CZE, Denmark = DEN, Finland = FIN, France = FRA, Germany = GER, Greece = GRE, Hungary = HUN, Iceland = ICE, Ireland = IRE, Italy = ITA, Japan = JAP, Korea = KOR, Luxembourg = LUX, Mexico = MEX, Netherlands = NET, New Zealand = NEZ, Norway = NOR, Poland = POL, Portugal = POR, Slovakia = SLO, Spain = SPA, Sweden = SWE, Switzerland = SWI, Turkey = TUR, United Kingdom = UK, United States = US.

### 3 tipos de países

1. Países con alta equidad y eficiencia
2. Países con baja equidad y alta eficiencia
3. Países con niveles de equidad moderados y baja eficiencia