

Ejercicios de técnicas estadísticas aplicadas a la Geografía Humana

Javier Colomo Ugarte

Año 1992

ÍNDICE

PRIMERA PARTE

EJERCICIO NÚMERO 1

Del porcentaje de paro de las provincias españolas según datos de 1986.

Apartado a). Elabora una distribución de frecuencias de los valores relativos a las PROVINCIAS españolas.

Apartado b). Dibuja el histograma de frecuencias y el polígono de frecuencias relativas.

Apartado c). Calcula las principales medidas de tendencia central, media aritmética, mediana, moda y valores cuartiles.

Apartado d). Calcula las principales medidas de dispersión, rango, varianza y desviación típica.

Apartado e). Calcula las principales medidas de variación relativa, de asimetría y curtosis.

TÉCNICAS ESTADÍSTICAS

1 (Cuadro de frecuencias)

2 (Media aritmética)

3 (Mediana)

4 (Cuartiles - Deciles)

5 (Moda)

6 (Desviación media)

7 (Media aritmética)

8 (Varianza)

9 (Desviación típica)

10 (Índice de variabilidad)

11 (Coeficiente de variación)

12 (Asimetría o sesgo)

13 (Curtosis)

EJERCICIO NÚMERO 2

A partir de la distribución del empleo sectorial de las provincias andaluzas y del total nacional relativas al año 1975 expresadas en valores porcentuales en el cuadro correspondiente al ejercicio número 2, calcular: el índice de diversificación de Rodgers y el de Gibbs- Martin, de la primera provincia de la tabla, correspondiente a Almería. Igualmente calcular el índice de especialización según el índice de Nelson.

TÉCNICAS ESTADÍSTICAS

14 (Índice bruto de diversificación)

[15 \(Índice neto de diversificación\)](#)

[16 \(Índice de diversificación de Gibbs-Martin\)](#)

[17 \(Índice de Nelson\)](#)

[EJERCICIO NÚMERO 3](#)

Cálculo del coeficiente de localización de Sargent Florence para las ramas industriales 2, (Textil, cuero, calzado), 5, (Químicas), 9 (Material de transporte), de la provincia de Albacete, sobre el total de la Comunidad Autónoma de Castilla-La Mancha.

TÉCNICAS APLICADAS

[18 \(Coeficiente de localización de Sargentz Florence\)](#)

[EJERCICIO NÚMERO 4](#)

Con las cifras de ventas de las cincuenta primeras empresas españolas durante el año 1985, elabora una distribución de frecuencias y calcula el Índice de Concentración de Gini.

TÉCNICAS ESTADÍSTICAS

[19 \(Coeficiente de concentración de Gini\)](#)

[20 \(La curva de Lorenz\)](#)

[EJERCICIO NÚMERO 5](#)

Aplicación del índice de Weaver, a los aprovechamientos agrarios de un municipio medidos en términos de sus necesidades día/ hombre (valores porcentuales), calculando a que distribución de categorías ideal (1 a 6 aprovechamientos) se aproximaría mas.

TÉCNICAS ESTADÍSTICAS

[21 \(Índice de Weaver\)](#)

[EJERCICIO NÚMERO 6](#)

Con los datos de las variables correspondientes a los porcentajes de la población activa dedicada a la agricultura y la industria de las comunidades autónomas españolas (EPA, 1986), calcula:

Apartado a). El coeficiente de correlación de Pearson:

Apartado b). El coeficiente de determinación

Apartado c). Convierte los valores en datos ordinales y calcula el coeficiente de correlación de SPEARMAN.

Apartado d). Halla la recta de regresión de una variable sobre la otra

TÉCNICAS ESTADÍSTICAS

[22 \(Coeficiente de correlación de Pearson\)](#)

[23 \(Coeficiente de determinación\)](#)

[24 \(Coeficiente de Spearman\)](#)

[25 \(Coeficiente de correlación de Kendall\)](#)

[26 \(Recta de regresión\)](#)

[27 \(Residuales\)](#)

SEGUNDA PARTE

EJERCICIO 1

Halla el centro de gravedad de la distribución de ciudades de Nottinghamshire y contrástalo con el centro de gravedad de la superficie delimitada por sus límites administrativos. ¿Cual es la distancia existente entre ambos? Calcular el valor del radio dinámico.

TÉCNICAS ESTADÍSTICAS

[28 \(Centro medio y centro medio ponderado\)](#)

[29 \(Centro de gravedad\)](#)

[30 \(Centro de gravedad ponderado\)](#)

[31 \(Centro de gravedad de la superficie\)](#)

[32 \(Radio dinámico\)](#)

EJERCICIO NÚMERO 2

Aplicando el análisis de vecindad al plano de la figura decidir si su estructura espacial se adecua a una distribución dispersa o concentrada.

TÉCNICAS ESTADÍSTICAS

[33 \(Análisis de vecindad\)](#)

[34 \(Vecino más próximo\)](#)

EJERCICIO NÚMERO 3

El mapa de la figura represente las relaciones por carretera existentes entre diversos núcleos urbanos del Sur de Gales y Suroeste de Inglaterra. A partir del mismo dibujar la red topológica, Calculando las principales medidas de conexión y accesibilidad indicando la importancia relativa de sus arcos.

TÉCNICAS ESTADÍSTICAS

[35 \(Análisis red topológica\)](#)

EJERCICIO NÚMERO 4

Aplicar el análisis de formas según los índices descritos al condado de Gloucestershire antes y después de la reforma de la administración local en 1974.

TÉCNICAS ESTADÍSTICAS

[36 \(Índice de compacidad\)](#)

[37 \(Índice de Cole\)](#)

[38 \(Índice de Gibbs\)](#)

[39 \(Índice de I \(s\) Índice I \(T\)\)](#)

EJERCICIOS NÚMERO 5 y 6

A partir de la matriz de la estructura de los factores, (cuadro 3) aplicado a algunos aspectos geográficos de México sobre una tabla de datos de 18 variables 32 lugares.

1º Calcular el % de la varianza

2º Deducir el sentido de cada factor

TÉCNICAS ESTADÍSTICAS

[40 \(Análisis factorial\)](#)

EJERCICIOS DE LA PRIMERA PARTE

EJERCICIO NÚMERO 1

Del porcentaje de paro de las provincias españolas según datos de 1986.

Apartado a)

Elabora una distribución de frecuencias de los valores relativos a las PROVINCIAS españolas.

ORDENACIÓN DE LOS DATOS POR PROVINCIAS:

VALOR MÍNIMO, PROVINCIA DE LÉRIDA 6,3

VALOR MÁXIMO, PROVINCIA DE CÓRDOBA 34,1

INTERVALOS DE CLASE ELEGIDOS 5-10, 10-15, 15-20, 20-25, 25-30, 30-35.

El número de intervalos de clase lo he elegido teniendo en cuenta la indicación de situarse por debajo de cinco veces el log. del número de observaciones y la fórmula de Brook Carruthers para establecer el número de intervalos de clase que debe ser igual a $1 + 3,3 \log_{10}$ del total del número de observaciones. Que es $= \log_{10}$ de 50 = $1,69 \cdot 3,3 + 1 = 6,57$.

TABLA DE DATOS

Almería 19,5	Zaragoza 16,1	Ávila 18,9	Gerona 10,5	Orense 8,3
Cádiz 34,1	Baleares 16,6	Burgos 13,6	Lérida 6,3	Pontevedra 17
Córdoba 31	Las Palmas 28,1	León 14,5	Tarragona 16,3	Madrid 18,6
Granada 31,3	Tenerife 24,6	Palencia 19,9	Alicante 20,9	Murcia 20,6
Huelva 33,6	Cantabria 18,2	Salamanca 18,7	Castellón 11,6	Navarra 18,3
Huesca 14,6	Guadalajara 18	Zamora 19	La Coruña 14,1	Asturias 19,7
Jaén 27,5	Albacete 17,9	Segovia 15	Valencia 19,9	Álava 19
Málaga 31,7	Ciudad Real 18,4	Soria 9,7	Badajoz 31,9	Guipúzcoa 23,8
Sevilla 33,3	Cuenca 11,7	Valladolid 24,8	Cáceres 22,6	Vizcaya 24,4
Teruel 11	Toledo 13,1	Barcelona 23,7	Lugo 8,7	La Rioja 14,6

*

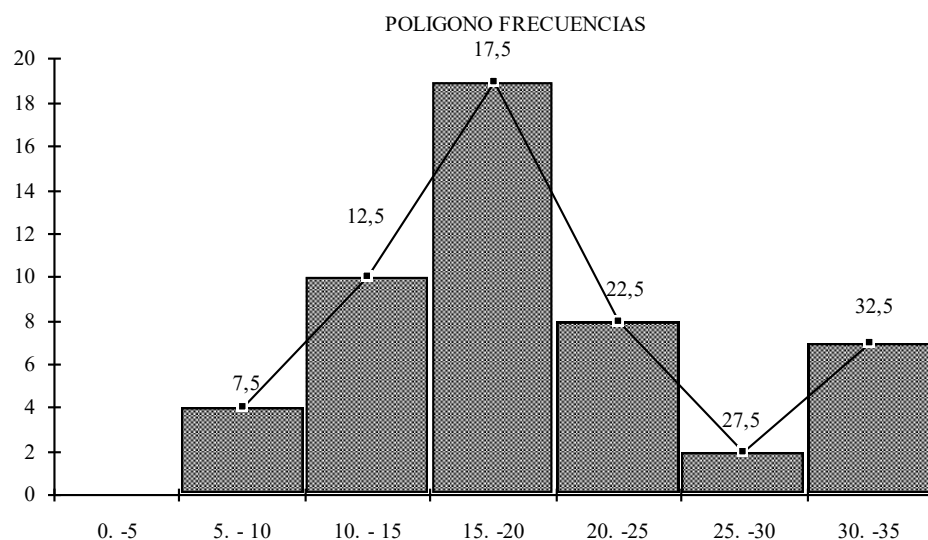
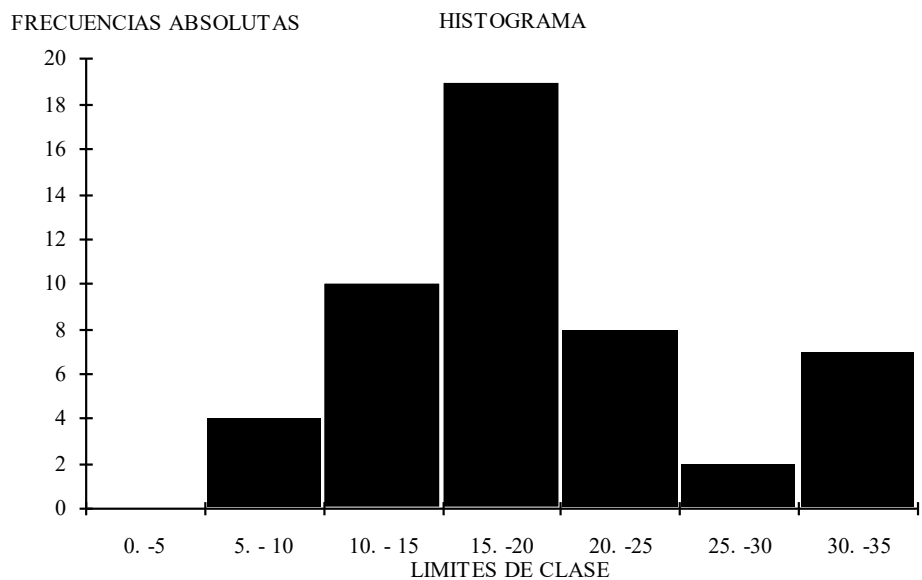
Intervalos	RECuento	Frecuencias
5 - 10	////	4
10 - 15	///// /////	10
15 - 20	///// ///// ///// ///	19
20 - 25	///// ///	8
25 - 30	//	2
30 - 35	///// //	7
		50

CUADRO DE FRECUENCIAS

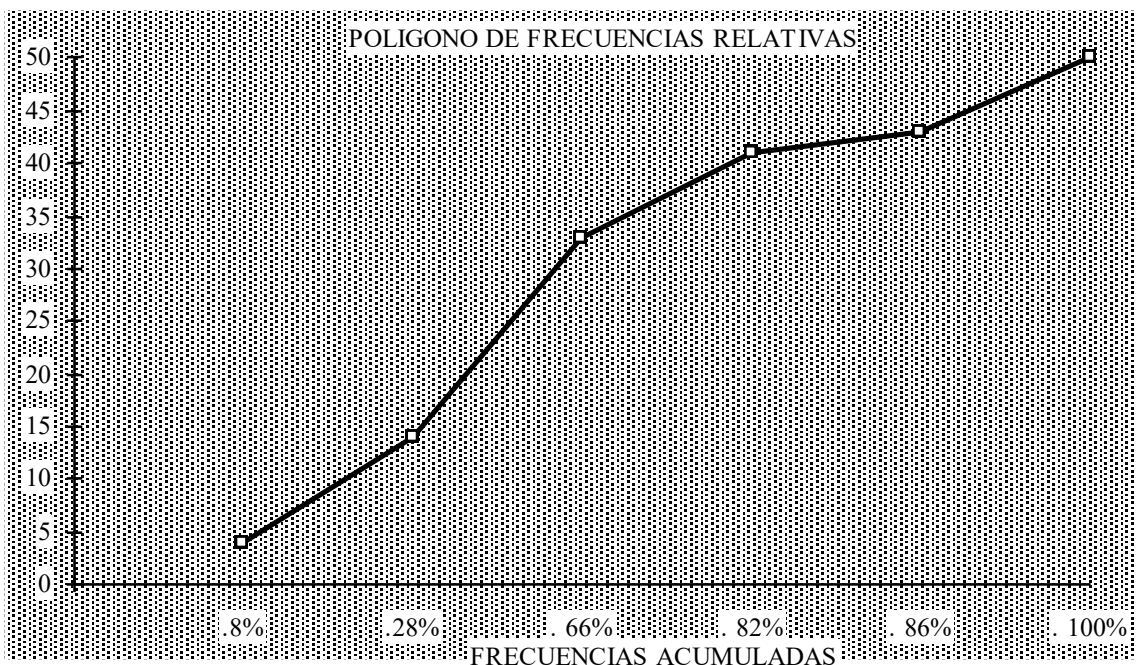
Intervalo de clase	Marca de clase	Frecuencia. absoluta	Frecuencias absolutas acumuladas	Frecuencia relativa = $F/50$	Frecuencia relativa acumulada
	x	f	F	r	R
5. - 10	7,5	4	4	0,08	0,08
10. - 15	12,5	10	14	0,2	0,28
15. -20	17,5	19	33	0,38	0,66
20. -25	22,5	8	41	0,16	0,82
25. -30	27,5	2	43	0,04	0,86
30. -35	32,5	7	50	0,14	1
		50		1	

Apartado b)

Dibuja el histograma de frecuencias y el polígono de frecuencias relativas.



POLÍGONO FRECUENCIAS RELATIVAS EN %



Apartado c)

Calcula las principales medidas de tendencia central, media aritmética, mediana, moda y valores cuartiles.

MEDIA ARITMÉTICA

Intervalo de clase	Marca de clase	Frecuencias absolutas	PRODUCTOS
	X	F	F*X
5. - 10	7,5	4	30
10. - 15	12,5	10	125
15. -20	17,5	19	332,5
20. -25	22,5	8	180
25. -30	27,5	2	55
30. -35	32,5	7	227,5
		50	950

Nota. La media aritmética tiene en el presente trabajo la forma de **X** o **Y** en negrita.

$$\sum(F*X) = \text{Sumatorio de } F*X$$

$$\sum F = n = \text{Sumatorio de } F$$

FÓRMULA MEDIA ARITMÉTICA EN DATOS AGRUPADOS: **X** =

$$\frac{\sum(F*X)}{n} = \frac{950}{50} = \mathbf{19}$$

MEDIANA

Es el valor medio de los valores centrales de un conjunto de datos, dividiendo el total de datos en dos partes iguales

En el caso de datos agrupados se halla según la siguiente fórmula:

MD = Mediana

L = Término inferior de la clase mediana (o del cuartil o decil correspondiente)

n = El total de frecuencias

$\sum F$ = El total de frecuencias de las clases más bajas a la clase mediana (o del cuartil o decil)

f = La frecuencia de la clase mediana (o del cuartil o decil correspondiente)

C = La amplitud del intervalo de clase.

$$\text{FÓRMULA MD} = L + \frac{(n/2 - \sum F) * C}{f = \text{mediana}}$$

CUARTILES - DECILES

Por extensión del concepto de mediana se pueden determinar los valores que dividen a los valores en cuatro partes iguales (cuartiles) o en diez (deciles).

Los valores Q1, MD, Q3, Q10, Q90, se sitúan según el siguiente cuadro

El número de orden de la frecuencia en cada caso corresponde a un intervalo de clase.

Intervalo de clase	Frecuencia. absoluta	Frecuencias absolutas acum.	Posición de los valores	Orden de la frecuencia
	f	F		
5. - 10	4	4		
10. - 15	10	14	Q10 Q1	(5°-6°) (12°-13°)
15. -20	19	33	Mediana MD	25°-26°
20. -25	8	41	Q3	38°-39°
25. -30	2	43		
30. -35	7	50	Q90	45°-46°
	50		1	

$$\text{FÓRMULA MD} = L + \frac{(n/2 - \sum F) * C}{f = \text{mediana}} = 15 + \frac{(50/2 - 14) * 5}{19} = 2,89 + 15 = \mathbf{17,89}$$

1 CUARTIL

$$\text{FÓRMULA Q1} = L + \frac{(n/4 - \sum F) * C}{f = Q1} = 10 + \frac{(50/4 - 4) * 5}{10} = 4,25 + 10 = \mathbf{14,25}$$

2 CUARTIL Es igual a la mediana

3 CUARTIL

$$\text{FÓRMULA Q3} = L + \frac{(n/3/4 - \sum F) * C}{f = Q3} = 20 + \frac{(50/ 3/4 - 33) * 5}{8} = 2,81 + 20 = \mathbf{22,81}$$

1 DECIL

$$\text{FÓRMULA } Q_{10} = L + \frac{(n/1/10 - \sum F) * C}{f = Q_{10}} = 10 + \frac{(50 / 1/10 - 4) * 5}{10} = 0,5 + 10 = \mathbf{10,5}$$

9 DECIL

$$\text{FÓRMULA } Q_{90} = L + \frac{(n/9/10 - \sum F) * C}{f = Q_{90}} = 30 + \frac{(50 / 9/10 - 43) * 5}{7} = 1,42 + 30 = \mathbf{31,42}$$

$$Q_{90} - Q_{10} = 31,42 - 10,45 = 20,97$$

MODA

Se define como el valor que tiene una mayor frecuencia en un conjunto de datos

En el caso de datos agrupados como es el presente ejemplo se halla con la siguiente fórmula:

L1 = Límite inferior de la clase de la moda

d1 = Diferencia de la frecuencia del intervalo de clase de la moda respecto de la anterior

d2 = Diferencia de la frecuencia del intervalo de clase de la moda respecto de la posterior.

C = Intervalo de clase:

$$\text{FÓRMULA } MO = L1 = 15 \quad (d1 = 19 - 10 = 9) \quad (d2 = 19 - 8 = 11) \quad C = 5$$

$$L + \frac{d1}{d1 + d2} * C = 15 + \frac{9}{9 + 11} * 5 = 2,25 + 15 = \mathbf{17,25}$$

OTRA FÓRMULA DE LA MODA

MO = MODA

L = Término superior de la clase modal

f1 = Frecuencia inferior de la clase modal

f2 = Frecuencia superior de la clase modal

C = Intervalo de clase

$$\text{FÓRMULA } MO = L - \frac{f1 * C}{f1 + f2} = 20 - \frac{10}{10 + 8} * 5 = 20 - 2,77 = \mathbf{17,23}$$

Apartado d)

Calcula las principales medidas de dispersión, rango, varianza y desviación típica.

La desviación intercuartílica o diferencia entre el Q1 y Q3 = 22,81 - 14,25 = **8,56**

Esta desviación contiene el 50% de los datos

Para obtener el promedio de todos los datos respecto a la mediana, se promedia la diferencia entre el cuartil superior e inferior. Este índice se conoce con el nombre de desviación semiintercuartílica.

RANGO (o desviación semiintercuartílica)

$$\text{FÓRMULA Q} = \frac{Q3 - Q1}{2} = \frac{22,81 - 14,25}{2} = 4,28$$

DESVIACIÓN MEDIA

Se denomina a la media, de las desviaciones entre las marcas de clase y la media aritmética por el número de frecuencias de cada clase (En la suma se prescinde del signo para sumar)

Clases	Frecuencia	Marca clase				
	F	X	X*F	X -X	X -X	X -X *F
5-10	4	7,5	30	-11,5	11,5	46
10-15	10	12,5	125	-6,5	6,5	65
15-20	19	17,5	332,5	-1,5	1,5	28,5
20-25	8	22,5	180	3,5	3,5	28
25-30	2	27,5	55	8,5	8,5	17
30-35	7	32,5	227,50	13,5	13,5	94,5
	50		950			279

MEDIA ARITMÉTICA

$$\text{MEDIA ARITMÉTICA } X = \frac{\sum X*F}{\sum F = n} = \frac{950}{50} = 19$$

$$\text{MEDIA ARITMÉTICA FÓRMULA DESVIACIÓN MEDIA} = \frac{\sum |X-X|*F}{n} = \frac{279}{50} = 5,58$$

VARIANZA

LA VARIANZA Se denomina a la media, de los cuadrados de las desviaciones entre las marcas de clase y la media aritmética por el número de frecuencias de cada clase.

Clases	Frecuencia	Marca clase	Desviación	Desviación2	Productos
	F	X	X -X	(X -X)2	(X -X)2*F
5-10	4	7,5	-11,5	132,25	529
10-15	10	12,5	-6,5	42,25	422,5
15-20	19	17,5	-1,5	2,25	42,75
20-25	8	22,5	3,5	12,25	98
25-30	2	27,5	8,5	72,25	144,5
30-35	7	32,5	13,5	182,25	1275,75
	50			443,5	2512,5

$$\text{FÓRMULA DE LA VARIANZA} = \frac{\sum (X-X)^2*F}{n} = \frac{2512,5}{50} = 50,25$$

DESVIACIÓN TÍPICA

La desviación típica es la raíz cuadrada de la varianza. Por la tanto se puede definir como la raíz cuadrada del promedio de los cuadrados de las desviaciones.

$$\text{DESVIACIÓN TÍPICA FÓRMULA} = S = \sqrt{\frac{\sum (X-X)^2*F}{n}} = \sqrt{50,25} = 7,088$$

Apartado e)

Calcula las principales medidas de variación relativa, de asimetría y curtosis.

MEDIDAS DE VARIACIÓN RELATIVA

Estas medidas comparan las medidas de centralidad con las de dispersión. Se expresan en términos porcentuales , y se definen de la siguiente manera

$$\begin{aligned}\text{ÍNDICE DE VARIABILIDAD} &= \frac{\text{Desviación intercuartílica}}{\text{Mediana}} * 100 = \frac{Q1-Q3}{MD} * 100 \\ &= \frac{22,81-14,25}{17,89} * 100 = \mathbf{47,84\%}\end{aligned}$$

$$\begin{aligned}\text{VARIABILIDAD RELATIVA} &= \frac{\text{Desviación media}}{\text{Media aritmética}} * 100 = \frac{DM}{X} * 100 \\ &= \frac{5,58}{19} * 100 = \mathbf{29,36\%}\end{aligned}$$

COEFICIENTE DE VARIACIÓN

$$\begin{aligned}\text{COEFICIENTE DE VARIACIÓN} &= \frac{\text{Desviación típica}}{\text{media aritmética}} * 100 = \frac{S}{X} * 100 \\ &= \frac{7,088}{19} * 100 = \mathbf{37,30\%}\end{aligned}$$

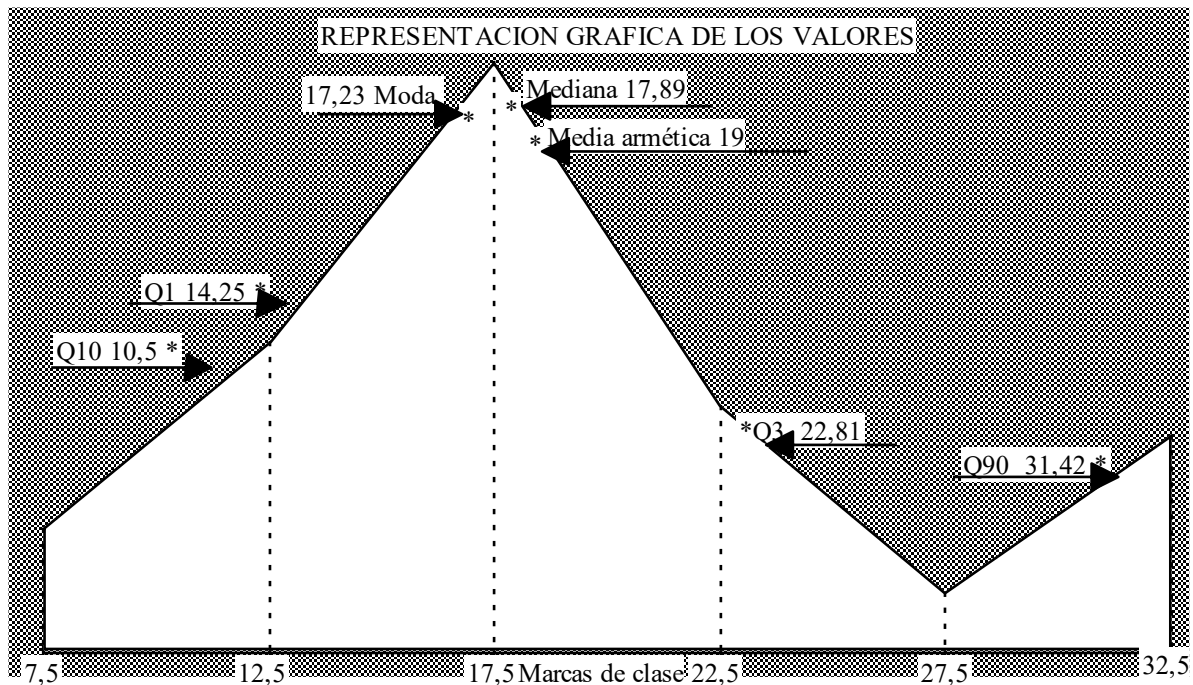
ASIMETRÍA

La asimetría o sesgo es el grado de asimetría o falta de asimetría. En el caso de asimetría con sesgo positivo , la moda se encuentra siempre a la izquierda de la distribución, a continuación se sitúa la mediana y la media. En el caso opuesto, de asimetría negativa, la cola está a la izquierda y las medidas de tendencia central se disponen de forma contraria que el caso anterior. Por último en el caso de la distribución simétrica media, moda y mediana coinciden.

El coeficiente se calcula teniendo en cuenta los valores de la moda, la mediana y la desviación típica. Sus valores se sitúan entre ± 1

$$\begin{aligned}\text{LA ASIMETRÍA O SESGO} &= \frac{\text{Media Arit.} - \text{moda}}{\text{Desviación típica}} = \frac{19 - 17,23}{7,088} = \frac{1,77}{7,088} = \mathbf{0,24}\end{aligned}$$

El coeficiente demuestra una asimetría de carácter positivo



El gráfico refleja las características de una distribución de frecuencias con asimetría positiva al situarse los valores de la moda, media y mediana, de izquierda a derecha.

CURTOSIS

Es el grado de apuntamiento de una distribución

Para definir la misma se toma como referencia la curva normal de una distribución,. Un grado de apuntamiento grande y con fuerte desarrollo de sus colas, se llama leptocúrtica, las que tienen forma aplanada son platocúrticas y a la curva normal se le denomina mesocúrtica.

una de las formulas para medir la curtosis es la siguiente:

DEFECTO DE CURTOSIS =

$$\frac{1/2 \cdot (Q3 - Q1)}{Q90 - Q10} = \frac{1/2 \cdot (22,81 - 14,25)}{31,42 - 10,5} = \frac{8,56}{20,92} = 0,409 / 2 = \mathbf{0,204}$$

Para la interpretación de esta fórmula se tiene en cuenta los siguientes valores

Si el resultado es superior a 0,263 es leptocúrtica

Si es igual a 0,263 es mesocúrtica

Si está por debajo de 0,263 es platocúrtica

En el presente ejemplo LA DISTRIBUCIÓN ES PLATOCÚRTICA.

NOTA. La desviación típica en el programa del ordenador es de 7,038.

EJERCICIO NÚMERO 2

A partir de la distribución del empleo sectorial de las provincias andaluzas y del total nacional relativas al año 1975 expresadas en valores porcentuales en el cuadro correspondiente al ejercicio número 2, calcular el índice de diversificación de Rodgers y el de Gibbs- Martin, de la primera provincia de la tabla, correspondiente a Almería. Igualmente calcular el índice de especialización según el índice de Nelson.

ÍNDICE BRUTO DE DIVERSIFICACIÓN

PROVINCIA ALMERÍA

NÚMERO DE RAMAS = 9

RAMAS DE LA PRODUCCIÓN	% de empleo ordenados	totales acumulados
Comercio	27,2	27,2
Agricultura, selvicultura y pesca	16,7	43,9
Construcción y obras públicas	16,1	60
Diversificadas	13,3	73,3
Agua, gas y electricidad	11,2	84,5
Transportes y comunicaciones	9,6	94,1
Industrias fabriles	3,6	97,7
Servicios	1,3	99
Minería	1	100

Se considera Índice Bruto de Diversificación a la suma de los % acumulados

ÍNDICE BRUTO DE DIVERSIFICACIÓN PROVINCIAL 679,7

ESPAÑA

RAMAS DE LA PRODUCCIÓN	% de empleo ordenados	totales acumulados
Diversificadas	24,5	24,5
Agricultura, selvicultura y pesca	22,3	46,8
Construcción y obras públicas	13,7	60,5
Agua, gas y electricidad	11,1	71,6
Comercio	8,1	79,7
Transportes y comunicaciones	8,1	87,8
Minería	4,4	92,2
Servicios	3,9	96,1
Industrias fabriles	3,9	100
ÍNDICE BRUTO DE DIVERSIFICACIÓN NACIONAL		659,2

Se considera Índice de Diversificación Mínima al concepto teórico que supone a toda la población activa trabajando en una sola rama de la producción. En el presente ejemplo al considerarse nueve ramas sería =

Índice de diversificación mínima: $9 \times 100 = 900$

ÍNDICE NETO DE DIVERSIFICACIÓN

El Índice Neto de Diversificación de Rodgers varía entre 0 y 1000 . Un valor 0 indica una diversificación extrema, (falta de especialización) y 1000 supone una diversificación mínima o (especialización) . Este índice puede proporcionar valores negativos, lo que indica que la distribución que se mide esta más diversificada que la que se compara.

SE CALCULA:

FÓRMULA = $1000 \times \frac{\text{Índice Bruto de Diversificación Prov.} - \text{Índice Bruto D.N.}}{\text{Índice Bruto de Diversificación mínima} - \text{Índice Bruto D.N.}}$

$$= 679,7 - 659,2 = 20,5 \times 1000 = 85,13$$

$$= 900 - 659,2 = 240,8$$

El valor indica que la distribución sectorial en Almería está menos diversificada que en el conjunto del país.

ÍNDICE DE DIVERSIFICACIÓN DE GIBBS-MARTIN

Este índice sirve para medir el grado de concentración o de diversificación de una distribución. El presente ejemplo la distribución sectorial de la provincia de Almería.

NÚMERO DE RAMAS 9

RAMAS DE LA PRODUCCIÓN	% de empleo = X	X ²
Minería	1	1
Servicios	1,3	1,69
Industrias fabriles	3,6	12,96
Transportes y comunicaciones	9,6	92,16
Agua, gas y electricidad	11,2	125,44
Diversificadas	13,3	176,89
Construcción y obras públicas	16,1	259,21
Agricultura, selvicultura y pesca	16,7	278,89
Comercio	27,2	739,84
TOTAL 9	100	1688,08

$$\text{FÓRMULA I.G.M.} = 1 - \frac{\sum X^2}{(\sum X)^2} = \frac{1688,08}{100^2} = 0,168 = 1 - 0,168 = \mathbf{0,832}$$

Según el índice de G.M. cuanto más se aproxime este al valor 0 tanto más concentrada será la distribución y cuanto más se aproxime al valor 1 tanto más diversificada será la distribución.

En el caso de Almería, el valor del ÍNDICE DE GIBBS MARTIN es de 0,832 lo que indica que la población activa se reparte de manera bastante uniforme en las diferentes ramas.

ÍNDICE DE NELSON

Este índice sirve para evaluar el grado de especialización de una actividad, en un ámbito espacial determinado. Este índice se basa en las propiedades de la desviación típica. El grado de especialización se mide sumando la desviación típica a la media aritmética

El Índice de Nelson se aplica en el presente caso a la rama de Agricultura, selvicultura y pesca de la Comunidad Autónoma andaluza.

RAMA 1	AGRICULTURA	ANDALUCÍA	
F	X	X-X	(X-X) ²
Almería	16,70	-10,98	120,72
Cádiz	28,40	0,71	0,50
Córdoba	32,20	4,51	20,36
Granada	29,90	2,21	4,89
Huelva	26,90	-0,78	0,62
Jaén	30,20	2,51	6,31
Málaga	27,80	0,11	0,01
Sevilla	29,40	1,71	2,93
TOTAL 8	221,50		156,36
		Varianza = 19,54	Desv. típica = S
			4,42
	Media aritmética	X = 27,68	
VALOR 1	X+1S = A	32,1	CÓRDOBA,.
VALOR 2	X+2S = A2	36,52	
VALOR 3	X+3S = A3	40,94	

$$\text{MEDIA ARITMÉTICA } X = \frac{\sum X}{F} = \frac{221,50}{8} = \mathbf{27,68}$$

DESVIACIÓN TÍPICAS

$$S = \sqrt{\frac{\sum (X-X)^2}{n}} = \sqrt{\frac{156,36}{8}} = \mathbf{4,42}$$

El valor 1 es igual a la media aritmética más la desviación típica

El valor 2 es igual a la media aritmética más dos veces la desviación típica

El valor 3 es igual a la media aritmética más tres veces la desviación típica

Cada valor supone un grado de especialización de la rama económica que se trate, en el presente caso:

El valor de 1 = $X + 1S = 27,68 + 4,42 = 32,1$

El valor de 2 = $X + 2S = 27,68 + (4,42 \cdot 2) = 36,52$

El valor de 3 = $X + 3S = 27,68 + (4,42 \cdot 3) = 40,94$

La provincia que tienen un valor de especialización es CÓRDOBA, que se sitúa dentro de $X + 1S$.

EJERCICIO COMPLEMENTARIO

En el presente cuadro muestro los cálculos y la especialización para todas las provincias de la Comunidad Autónoma Andaluza.

EMPLEO SECTORIAL DE LAS PROVINCIAS ANDALUZAS DEL AÑO 1975

	Agricult. 1	Minería 2	Constr. 3	Agua-gas 4	Ind fabr 5	Transp.. 6	Comun. 7	Servicios 8	Diver. 9
Almería	16,70	1,00	16,10	11,20	3,60	9,60	27,20	1,30	13,30
Cádiz	28,40	0,70	11,50	7,80	4,30	4,80	3,90	0,50	38,10
Córdoba	32,20	4,40	10,40	8,50	3,00	3,80	6,20	6,00	25,50
Granada	29,90	3,20	15,00	10,50	5,10	6,70	10,40	1,30	17,90
Huelva	26,90	1,40	7,40	11,50	4,30	20,20	6,70	6,80	15,00
Jaén	30,20	5,60	9,40	8,00	2,90	4,20	9,00	2,70	28,10
Málaga	27,80	7,60	18,50	7,20	2,60	6,10	5,90	1,20	23,30
Sevilla	29,40	5,00	9,50	7,20	3,40	7,10	8,50	1,60	28,40
C. Andalucía	221,50	28,90	97,80	71,90	29,20	62,50	77,80	21,40	189,60

M. aritmética	27,69	3,61	12,23	8,99	3,65	7,81	9,73	2,68	23,70
Des. Típica	4,42	2,31	3,62	1,68	0,80	4,99	6,87	2,23	7,66

VALORES DE ESPECIALIZACIÓN

	Agricult. 1	Minería 2	Const. 3	Aguas 4	Ind fabr. 5	Trans. 6	Comun. 7	Servicios 8	Diver. 9
$X + 1S$	32,11	5,92	15,84	10,66	4,45	12,81	16,60	4,91	31,36
$X + 2S$	36,53	8,24	19,46	12,34	5,25	17,80	23,47	7,14	39,02
$X + 3S$	40,95	10,55	23,07	14,02	6,05	22,80	30,35	9,38	46,69

ESPECIALIZACIÓN PROVINCIAL

	Agricult. 1	Minería 2	Const. 3	Agua-gas 4	Ind fabr. 5	Transp. 6	Comun. 7	Servicio 8	Diver. 9
Almería			$X + 1S$	$X + 1S$			$X + 2S$		
Cádiz									$X + 1S$
Córdoba	$X + 1S$							$X + 1S$	
Granada					$X + 1S$				
Huelva				$X + 1S$		$X + 2S$		$X + 1S$	
Jaén									
Málaga		$X + 1S$	$X + 1S$						
Sevilla									

EJERCICIO NÚMERO 3

Calcula el coeficiente de localización de Sargent Florence para las ramas industriales 2, (Textil, cuero, calzado), 5, (Químicas), 9 (Material de transporte), de la provincia de Albacete, sobre el total de la Comunidad Autónoma de Castilla-La Mancha.

COEFICIENTE DE LOCALIZACIÓN DE SARGENTZ FLORENCE

El coeficiente de localización de Sargentz Florence es uno de los coeficientes de localización más utilizado en Geografía. Sirve para comparar la estructura industrial de dos espacios económicos diferenciados

El valor del coeficiente oscila entre un valor 1, que indica una distribución idéntica y en la misma proporción que en el conjunto del espacio que se compara, por encima de 1 indica un valor superior, y por debajo de 1 un valor inferior.

En el presente ejercicio se aplica a las ramas (2, Textil, cuero, calzado), (5, Químicas), (9, Material de transporte), de la provincia de Albacete, en comparación con la Comunidad Autónoma de Castilla-la Mancha.

FÓRMULA

E i.j. = Empleo en el sector respectivo en ALBACETE

E.j. = Empleo en el sector respectivo en CASTILLA LA MANCHA

E i t. = Empleo en ALBACETE

E.t. = Empleo en CASTILLA LA MANCHA

FÓRMULA.

$$\frac{(E i.j./E.t.) \times 100}{(Ei.t./E.t.) \times 100}$$

DATOS DE APLICACIÓN

AÑO	TEXTIL CUERO CALZADO	QUÍMICAS	MATERIAL DE TRANSPORTE	TOTAL EMPLEO
CASTILLA LA MANCHA	22412	4520	415	84821
ALBACETE	8708	537	53	17847

SECTOR TEXTIL, CUERO CALZADO =

$$\frac{8708/17847 \times 100}{22412/84821 \times 100} = \mathbf{1,84}$$

SECTOR QUÍMICAS =

$$\frac{537/17847 \times 100}{4520/84821 \times 100} = \mathbf{0,56}$$

SECTOR MATERIAL DE TRANSPORTE =

$$\frac{53/17847}{415/84821} \times 100 = 0,60$$

OTRA FÓRMULA CIÓN =

Empleo de la provincia en la rama * Empleo en toda la Comunidad Autónoma

Empleo en la rama en la Comunidad Autónoma * Empleo en toda la provincia =

SECTOR TEXTIL, CUERO CALZADO =

$$8708*84821 = \frac{738621268}{22412*17847} = 1,84$$

$$22412*17847 = 39986964$$

SECTOR QUÍMICAS =

$$537*84821 = \frac{45548877}{4520*17847} = 0,56$$

$$4520*17847 = 80668440$$

SECTOR MATERIAL DE TRANSPORTE =

$$53*84821 = \frac{4495513}{415*17847} = 0,60$$

$$415*17847 = 7406505$$

COMENTARIO

El empleo en el SECTOR del TEXTIL, CUERO y CALZADO, en ALBACETE presenta un índice superior al de Castilla la Mancha al situarse por encima de 1.

En el SECTOR QUÍMICO y de MATERIAL de TRANSPORTE se aprecia un índice menor al situarse este por debajo de 1.

EJERCICIO NÚMERO 4

Con las cifras de ventas de las cincuenta primeras empresas españolas durante el año 1985, elabora una distribución de frecuencias y calcula el Índice de Concentración de Gini.

La Curva de Lorenz es una curva de frecuencias acumulativas, que se puede emplear para estudiar distribuciones y para compararlas entre sí. Su expresión gráfica permite un comentario apreciativo visual en cuanto al grado de uniformidad o concentración de la distribución estudiada, sin ofrecer un índice que precise sus resultados. El Coeficiente de Concentración de Gini permite precisar el grado concentración.

COEFICIENTE DE CONCENTRACIÓN DE GINI

El Coeficiente de Concentración de GINI se define como el promedio ponderado de las diferencias absolutas entre todos los valores.

Una forma de calcularlo, es a partir de la Curva de Lorenz.

Se divide el cuadrado por la mitad tomando el área del triángulo inferior que forma la diagonal.

Se halla el área situada por debajo de la Curva de Lorenz y se resta al área total del triángulo, la diferencia es igual al área de concentración.

El presente ejercicio lo realizo en una distribución de frecuencias, tomando tres intervalos de clase.

Dividiendo por tanto los datos en tres grupos A , B , C , (en miles).

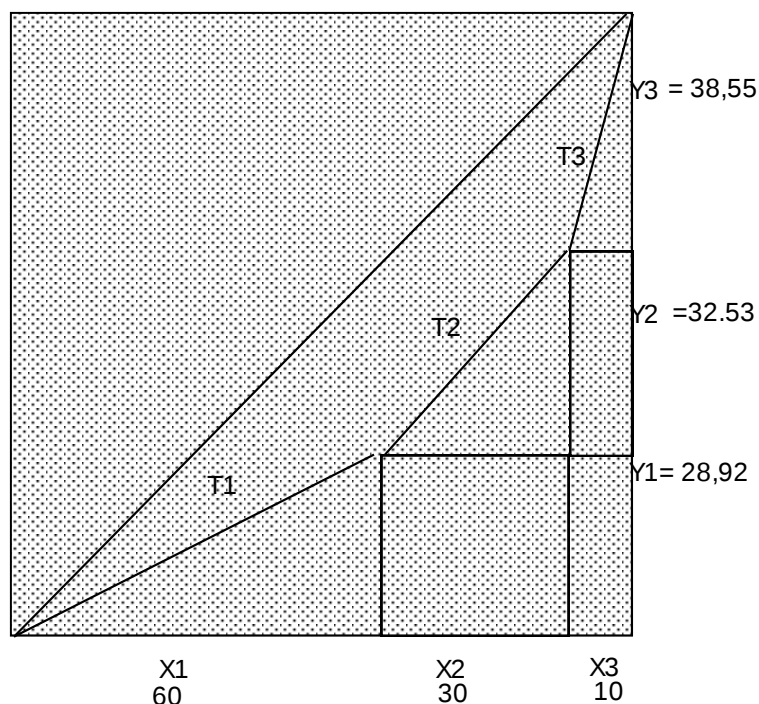
In. de clase	Marca de clase	Frec. Absolutas	Frec. abs. acu.	Frec. relativas %	F. r.elt ac. en %
	X	F			
50 150	100	30	30	60,00	60
150 300	225	15	45	30,00	90
300 1300	800	5	50	10,00	100

X*F	X*F Acumulado	% X*F	% Acumulados
3000	3000	28,92	28,92
3375	6375	32,53	61,45
4000	10375	38,55	100,00

CURVA DE LORENZ

GRUPO DE EMPRESAS % Acumulados CIFRAS DE VENTAS % Acumulados

<u>EN %</u>		<u>EN %</u>	
A 60 %	60%	28,92 %	28,92 %
B 30 %	90%	32,53 %	61,45 %
C 10 %	100%	38,55	100 %



CÁLCULOS PARA OBTENER EL COEFICIENTE DE CONCENTRACIÓN DE GINI
CÁLCULO DE LOS TRIÁNGULOS Y RECTÁNGULOS:

TRIÁNGULOS	RECTÁNGULOS
$1^{\circ} = 1/2 (X1*Y1) = 1/2 (60* 28,92) =$	$1^{\circ} X2*Y1 =$
$= \frac{1735,2}{2} = 867,6$	$= 30* 28,92 = 867,6$
$2^{\circ} = 1/2 (X2*Y2) = 1/2 (30*32,53) =$	$2^{\circ} 3^{\circ} = (Y1+Y2) *X3 =$
$= \frac{975,9}{2} = 487,95$	$= (28,2+32,53) * 10 = 614,5$
$3^{\circ} = 1/2 (X3*Y3) = 10*38,55 =$	
$\frac{385,5}{2} = 192,75$	

SUMA DE LAS ÁREAS

867,6
487,95
192,75
867,6
614,5
3030,4

El área del triángulo formado por la línea de distribución y los ejes de coordenadas es igual a la mitad del área total del diagrama, es decir $= \frac{100*100}{2} = 5000$

Por lo tanto el área de concentración , es igual al área dentro del triángulo grande, menos el área del triángulo situado en la parte inferior de la curva de Lorenz.

COEFICIENTE DE GINI

ÁREA DE CONCENTRACIÓN = 5000 - 3030,4 = 1969,6

El coeficiente de GINI es:

$$G = \frac{1969,6*2}{100*100} = \frac{3939,2}{10000} = 0,39 \quad \text{De otra manera } \frac{1969,6}{5000} = 0,39$$

EJERCICIO COMPLEMENTARIO

ANEXO DE DATOS TOTALES DE LAS 50 EMPRESAS AGRUPADAS IGUAL QUE EL CUADRO DE DISTRIBUCIÓN DE FRECUENCIAS

GRUPOS %	GRUPO 30 EMPRESAS			GRUPO 15 EMPRESAS		GRUPO 5 E
Total empresas 50	52158	67057	98552	163614	203262	302635
GrupoA 30 = 60%	52919	67750	100900	172451	234426	419674
Grupo B15 = 30%	57379	70763	102734	180219	235493	475442
Grupo C5 = 10%	58272	72040	110487	180651	257169	838655
	59227	75000	128020	185929	274544	1295063
	59545	75978	129278	191003		
	60284	79000	143517	191982		
	63808	86247	143542	194522		
	63867	87665	147477	195040		
	64639	95395	148383	196325		
			T = 2621883		T = 3056630	T = 3331469
			29,10%		33,92%	36,98*
						T = 9009982

LOS VALORES DE LOS % SE APROXIMAN BASTANTE A LOS DE LA DISTRIBUCIÓN DE FRECUENCIAS

ÁREAS DE LOS TRIÁNGULOS $60 \times 29,1/2 = 873 + 30 \times 33,92/2 = 508,8 + 10 \times 36,98/2 = 184,9$

ÁREAS DE LOS CUADRADOS $= 30 \times 29,1 = 873 + 10 \times (29,1 + 33,92) = 630,2$

SUMA $= 873 + 508,8 + 184,9 + 873 + 630,2 = 3069,9$

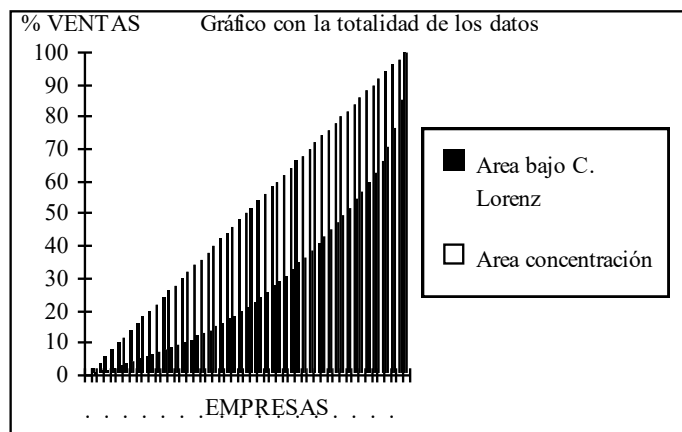
DIFERENCIA $= 5000 - 3069,9 = 1930,1$

$$G = 1930,1/5000 = \mathbf{0,38}$$

La siguiente tabla muestra los % acumulados de las ventas de las cincuenta empresas que reflejaría las líneas de coordenadas en el gráfico.

0,58%	3,77%	7,32%	11,33%	16,29%	22,63%	30,92%	41,02%	51,91%	66,38%
1,17%	4,44%	8,07%	12,21%	17,41%	24,22%	32,83%	43,15%	54,51%	71,04%
1,80%	5,15%	8,85%	13,16%	18,55%	25,82%	34,83%	45,31%	57,12%	76,32%
2,45%	5,85%	9,65%	14,14%	19,77%	27,45%	36,83%	47,47%	59,98%	85,63%
3,11%	6,57%	10,49%	15,19%	21,20%	29,10%	38,90%	49,65%	63,02%	100,0%

El gráfico muestra la curva de Lorenz teniendo en cuenta cada una de las empresas y de las ventas.



- En la línea de abscisas cada empresa representaría $100/50 =$ el 2%.
- La base de los triángulos y cuadrados es siempre igual a 2.
- La altura de los triángulos es igual a la altura del triángulo grande = 100.
- La altura de los cuadrados es igual a la suma de todos los % acumulados menos el último por ocupar ese espacio el triángulo último = 1254,54.
- El área de todos los triángulos es igual $1/2 (2 * 100) = 100$
- El área de los cuadrados el igual $1254 * 2 = 2509,08$
- El área total debajo de la curva de Lorenz = $2509,08 + 100 = 2609,08$
- El área de concentración es igual $5000 - 2609,08 = 2390,52$

$$\text{El coeficiente de GINI} = 2390,52 / 5000 = \mathbf{0,47}$$

EJERCICIO NÚMERO 5

Aplicación del índice de Weaver, a los aprovechamientos agrarios de un municipio medidos en términos de sus necesidades día/ hombre (valores porcentuales), calculando a que distribución de categorías ideal (1 a 6 aprovechamientos) se aproximaría más.

ACTIVIDADES	APROVECHAMIENTOS
Cosecha comercializable	50
Cría ganado porcino	19
Cría ganado vacuno	15
Industria productos lácteos	12
Avicultura	3
Ganado lanar	1

Este índice establece que tipo de combinación ideal se ajusta más a una determinado combinación de aprovechamientos reales.

Esto se mide mediante la fórmula de la varianza, es decir, determinando la diferencia entre el porcentaje ocupado por cada actividad en la realidad y el mismo porcentaje en el modelo ideal.

ÍNDICE DE WEAVER

W =

d = DIFERENCIA ENTRE EL PORCENTAJE DE LOS APROVECHAMIENTOS EN LA REALIDAD Y EL PORCENTAJE IDEAL

N = NÚMERO DE ACTIVIDADES

$$\text{FÓRMULA } W = \sum d^2 / N$$

MÉTODO DE WEAVER	Monoactividad	Combinación 2 actividades	
	A	A	B
% Aprovechamiento. real	50	50	19
% Aprovechamiento ideal	100%	50 %	50%
Diferencia (d)	50	0	31
Diferencia al cuadrado (d2)	2500	0	961
Suma de la diferencia al 2 = ($\sum d^2$)	2500	961	
Suma difers2/nº activids = $\sum d^2/N$	2500	480,5	

	Combinación 3 actividades			Combinación 4 actividades			
	A	B	C	A	B	C	D
% Aprovechamiento. real	50	19	15	50	19	15	12
% Aprovechamiento ideal	33,33 %	33,33%	33,33%	25%	25%	25%	25%
Diferencia (d)	16,7	14,3	18,3	25	6	10	13
Diferencia al cuadrado (d2)	278,89	204,49	334,89	625	36	100	169
Suma de la diferencia al 2 = ($\sum d^2$)	818,27			930			
Suma difers2/nº activids = $\sum d^2/N$	272,75			232,5			

Valor más bajo de la varianza

	Combinación 5 actividades					Combinación 6 actividades					
	A	B	D	C	E	A	B	C	D	E	F
% Aprovechamiento. real	50	19	15	12	3	50	19	15	12	3	1
% Aprovechamiento ideal	20 %	20%	20%	20%	20%	16,6%	16,6 %	16,6%	16,6%	16,6%	16,
Diferencia (d)	30	1	5	8	17	33,4.	2,4.	1,6.	4,6.	13,6.	15,6
Diferencia al cuadrado (d2)	900	1	25	64	289	1115,5	5,5	7,2	5 .21,6	184,96	243,36
Suma de la diferencia al 2 = ($\sum d^2$)	1279					1573,8					
Suma difers2/nº activids = $\sum d^2/N$	255,8					262,3					

EL VALOR MAS BAJO DE LA VARIANZA CORRESPONDE A LA COMBINACIÓN DE CUATRO ACTIVIDADES CON 232,5 LO QUE REPRESENTA LA DIFERENCIA MENOR CON EL VALOR IDEAL .

EJERCICIO NÚMERO 6

Con los datos de las variables correspondientes a los porcentajes de la población activa dedicada a la agricultura y la industria de las comunidades autónomas españolas (EPA, 1986), calcula:

Apartado a)

El coeficiente de correlación de Pearson:

COEFICIENTE DE CORRELACIÓN DE PEARSON

VARIABLE POBLACIÓN ACTIVA AGRARIA = X

VARIABLE POBLACIÓN ACTIVA INDUSTRIAL = Y

Comunidades	Población	Población					
Autónomas	Agrícola X	Industrial Y					
	X	Y	X-X	Y-Y	(X-X)*(Y-Y)	(X-X) ²	(Y-Y) ²
Andalucía	19,40	15,90	2,08	-7,07	-14,72	4,34	49,99
Aragón	16,90	26,90	-0,42	3,93	-1,64	0,17	15,44
Baleares	6,70	9,50	-10,62	-13,47	143,03	112,73	181,46
Canarias	12,90	10,80	-4,42	-12,17	53,76	19,52	148,12
Cantabria	21,70	21,00	4,38	-1,97	-8,64	19,21	3,88
Castilla la M	26,20	21,40	8,88	-1,57	-13,95	78,90	2,47
Castilla Le.	25,50	20,20	8,18	-2,77	-22,67	66,95	7,68
Cataluña	5,60	36,10	-11,72	13,13	-153,85	137,30	172,38
Comunid. V	13,40	28,70	-3,92	5,73	-22,45	15,35	32,83
Extremadura	29,40	10,20	12,08	-12,77	-154,30	145,98	163,09
Galicia	40,50	15,30	23,18	-7,67	-177,82	537,42	58,84
Madrid	1,20	22,60	-16,12	-0,37	5,97	259,78	0,14
Murcia	19,10	23,10	1,78	0,13	0,23	3,18	0,02
Navarra	11,30	34,90	-6,02	11,93	-71,79	36,21	142,31
País Vasco	4,50	38,00	-12,82	15,03	-192,64	164,29	225,88
Asturias	22,20	26,20	4,88	3,23	15,77	23,84	10,43
La Rioja	17,90	29,70	0,58	6,73	3,92	0,34	45,28
	$\Sigma = 294,40$	$\Sigma = 390,50$	$\bar{X} = 17,32$	$\bar{Y} = 22,97$	$\Sigma = -611,78$	$\Sigma = 1625,50$	$\Sigma = 1260,24$

El coeficiente de correlación de Pearson mide la relación entre las dos variables. Tiene unos valores comprendidos entre 0 y ± 1 .

1 es igual a una correlación perfecta, si es positiva será de + 1 si es negativa el valor será e - 1

La relación positiva indica una relación directa, mientras la relación negativa indica una relación inversa.

Existen diferentes grados de correlación, de las variables.

Los valores comprendidos de $\pm 0,7$ a ± 1 indica un elevado grado de asociación

Entre $\pm 0,4$ y $\pm 0,7$ indica una asociación de tipo medio

Entre $\pm 0,2$ y $\pm 0,4$ indica un bajo grado de asociación

La formula se obtiene dividiendo la covarianza entre los productos de las desviaciones típicas.

COEFICIENTE DE CORRELACIÓN DE PEARSON = r

DESVIACIONES TÍPICAS = S_x y S_y

COVARIANZA = $\sum (X - \bar{X}) * (Y - \bar{Y}) / n$

FÓRMULA

$$r = \frac{\sum (X - \bar{X}) * (Y - \bar{Y})}{n \cdot S_x * S_y}$$

En el presente ejemplo el desarrollo de la formula sería el siguiente:

$$\text{DESVIACIÓN TÍPICA DE } X = \sqrt{\frac{\sum (X - \bar{X})^2}{n}} \quad S_x = \sqrt{\frac{1625,5}{17}} = 9,78$$

$$\text{DESVIACIÓN TÍPICA DE } Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n}} \quad S_y = \sqrt{\frac{1260,24}{17}} = 8,61$$

COEFICIENTE DE CORRELACIÓN DE PEARSON =

$$\begin{aligned} r &= \frac{\sum (X - \bar{X}) * (Y - \bar{Y})}{n \cdot S_x * S_y} = \frac{611,78}{17 \cdot 9,78 * 8,61} = -35,99 \text{ Covarianza} \\ &= \frac{-35,99}{84,21} = r = -0,43 \end{aligned}$$

OTRA FORMULACIÓN

$$r = \frac{\sum (X - \bar{X}) * (Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} * \sqrt{\sum (Y - \bar{Y})^2}} = \frac{611,78}{1431,26} = -0,43$$

El resultado indica un grado de relación inversa de tipo medio entre las dos variables, puesto que son dos actividades que están relacionadas cuantitativamente de forma inversa, a mayor % de población activa en un sector menor será en el otro.

Apartado b)

El coeficiente de determinación

COEFICIENTE DE DETERMINACIÓN

El coeficiente de determinación (r^2) indica que la variación de un conjunto de datos explica la variación que aparece en el otro. y se expresa en % . Luego un valor para $r^2 = 1$ indica que se explica el 100 % de la variación de la variable que se relaciona.

En el presente ejemplo el coeficiente de determinación sería $r^2 = -0,43^2 = 0,1849$

Lo que supone que el conjunto de datos de una variable explica el 18,49 % de la variación experimentada por la otra variable.

Apartado c)

Convierte los valores en datos ordinales y calcula el coeficiente de correlación de SPEARMAN.

El coeficiente de correlación de SPEARMAN tiene un valor comprendido entre -1 y +1 , que indica unas correspondencias negativas y positivas. Los valores de 1 y -1 indican unas correspondencias positivas o negativas perfectas. Si el valor del coeficiente es 0 indica que no existe esa correlación.

Para hallar el coeficiente e ordenado los datos de cada variable en escala de intervalos de 1 a 17, igualando el valor más alto de cada variable a 17, y aplicando a los demás el valor correspondiente.

COEFICIENTE DE CORRELACIÓN DE SPEARMAN

C Autónoma	A	B	A	B	(A - B) = d	d ²
Madrid	1,20	22,60	1	9	-8	64
País Vasca	4,50	38,00	2	17	-15	225
Cataluña	5,60	36,10	3	16	-13	169
Baleares	6,70	9,50	4	1	3	9
Navarra	11,30	34,90	5	15	-10	100
Canarias	12,90	10,80	6	3	3	9
Comunid. V	13,40	28,70	7	13	-6	36
Aragón	16,90	26,90	8	12	-4	16
La Rioja	17,90	29,70	9	14	-5	25
Murcia	19,10	23,10	10	10	0	0
Andalucía	19,40	15,90	11	5	6	36
Cantabria	21,70	21,00	12	7	5	25
Asturias	22,20	26,20	13	11	2	4
Castilla L	25,50	20,20	14	6	8	64
Castilla la M	26,20	21,40	15	8	7	49
Extremadura	29,40	10,20	16	2	14	196
Galicia	40,50	15,30	17	4	13	169
						$\Sigma = 1196$

COEFICIENTE DE CORRELACIÓN DE SPEARMAN = r

DIFERENCIA DE ORDENES = d

SUMATORIO DE LAS DIFERENCIAS AL CUADRADO = $\sum d^2$

NÚMERO DE PARES DE ORDENES = n

COEFICIENTE = ± 1

$$\text{FÓRMULA } r_s = 1 - \frac{6 \cdot \sum d^2}{n^3 - n} \quad r_s = 1 - \frac{6 \cdot 1196}{17^3 - 17} = \frac{7176}{4896} = 1 - 1,46 = -0,46$$

El presente ejemplo presenta una correspondencia media y negativa.

COEFICIENTE DE CORRELACIÓN DE KENDALL

Este coeficiente permite conocer el grado de correlación de las variables.

Se calcula estableciendo el número de inversiones y no inversiones, tomando dos parejas de datos de cada variable. Realizando en el presente ejemplo la combinación de cada valor observado con los otros 16.

$$\text{El total de combinaciones es } = \frac{17 \cdot 16}{2} = 136$$

Para realizar el cálculo e ordenado una variable, en el presente caso la variable X, de menor a mayor, manteniendo la relación con los valores de la variable Y, y procediendo al recuento, relacionando A con B, A con C, A con D,
B con C, B con D,

De esta manera hasta las 136 combinaciones

	Población activa Agrícola	Población activa industrial	INVERSIONES Q	NO INVERS. P	COMBINACIONES
A	1,2	22,6	//// // = 8	//// // = 8	16
B	4,5	38	//// //// //// = 15		15
C	5,6	36,1	//// //// //// = 14		14
D	6,7	9,5		//// //// // = 13	13
E	11,3	34,9	//// //// // = 12		12
F	12,9	10,8	/ = 1	//// //// = 10	11
G	13,4	28,7	//// //// = 9	/ = 1	10
H	16,9	26,9	//// // = 8	/ = 1	9
I	17,9	29,7	//// // = 8		8
J	19,1	23,1	//// / = 6	/ = 1	7
K	19,4	15,9	// = 2	/// = 4	6
L	21,7	21	/// = 3	// = 2	5
M	22,2	26,2	/// = 4		4
N	25,5	20,2	// = 2	/ = 1	3
Ñ	26,2	21,4	// = 2		2
O	29,4	10,2		/ = 1	1
P	40,5	15,3			
			TOTAL 94	42	136

LA FÓRMULA ES = £

P = NÚMERO DE NO INVERSIONES

Q = NÚMERO DE INVERSIONES

$$\text{£} = \frac{P - Q}{P + Q} = \frac{42 - 94}{136} = -0,38$$

En los tres coeficientes de correlación aplicados a las dos variables de población activa agrícola y población activa industrial, se aprecian unos valores parecidos. Ello se produce porque la relación entre las dos variables solo se explica en parte, porque existe otra variable como es el sector de población activa de servicios que está incidiendo en la configuración de los datos de las dos variables presentes. La relación negativa se produce por tener los valores una relación inversa.

Apartado d)

Halla la recta de regresión de una variable sobre la otra

La recta de regresión permite calcular la relación entre dos variables

La variable dependiente o regresor es la variable cuyos valores han de estimarse, a partir de los valores de la variable independiente o predictor.

Si tomamos la población activa agrícola, (variable X) como variable independiente, la variable Y (población activa industrial) se convierte en variable dependiente.

En la situación inversa la variable Y pasa a ser variable independiente y la variable X pasa a ser variable dependiente

En el primer caso estaremos hablando de recta de regresión de Y sobre X

En el segundo caso será una recta de regresión de X sobre Y

En el presente cuadro se desarrolla el cálculo de la recta de regresión de Y sobre X

Comunidades Autónomas	Población agrícola. = X	Población industrial = Y		
	X	Y	X ²	X*Y
Andalucía	19,40	15,90	376,36	308,46
Aragón	16,90	26,90	285,61	454,61
Baleares	6,70	9,50	44,89	63,65
Canarias	12,90	10,80	166,41	139,32
Cantabria	21,70	21,00	470,89	455,70
Castilla la M	26,20	21,40	686,44	560,68
Castilla Le.	25,50	20,20	650,25	515,10
Cataluña	5,60	36,10	31,36	202,16
Comunidad V.	13,40	28,70	179,56	384,58
Extremadura	29,40	10,20	864,36	299,88
Galicia	40,50	15,30	1640,25	619,65
Madrid	1,20	22,60	1,44	27,12
Murcia	19,10	23,10	364,81	441,21
Navarra	11,30	34,90	127,69	394,37
País Vasco	4,50	38,00	20,25	171,00
Asturias	22,20	26,20	492,84	581,64
La Rioja	17,90	29,70	320,41	531,63
	$\sum X = 294,4$	$\sum Y = 390,5$	$\sum X^2 = 6723,82$	$\sum X*Y = 6150,76$
	n = 17	n = 17		

RECTA DE REGRESIÓN

Para hallar la recta de regresión es necesario saber las constantes **a**, y **b**.

La constante **a**, indica el punto de corte con el eje vertical,. La letra **b**, indica el valor de la pendiente de la recta.

FÓRMULA CONSTANTE **b** =

$$\frac{\sum X*Y - n*X*Y}{\sum X^2 - n*X^2} = 6150,76 - (17*17,31*22,97) = 6150,76 - 6759,38 = - 608,62 = - 0,37$$

$$\sum X^2 - n*X^2 = 6723,68 - (17*17,21*17,21) = 6723,68 - 5093,81 = 1629,87$$

FÓRMULA CONSTANTE **a** =

$$Y - b*X = 22,97 - - 0,37*17,31 = 29,43$$

$$\text{Recta de regresión de Y sobre X} = Y = b * x + a = Y = - 0,37.x + 29,43$$

La variable **X** representa un valor conocido

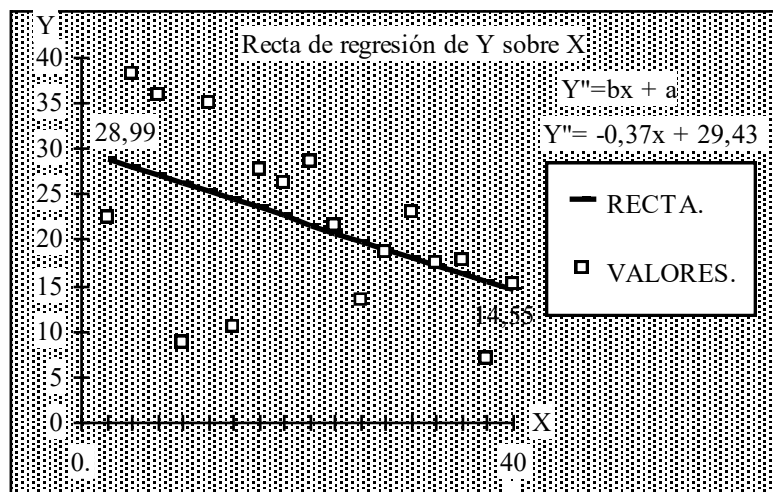
La aplicación de de la recta regresión de Y sobre X nos dará un valor estimado para Y que será el más aproximado valor real de la variable Y.

Para marcar la recta se hallan dos valores estimados para Y, y se unen los mismos, en el presente caso he hallado el valor mínimo y máximo estimados.

$$Y = -0,37*1,2 + 29,43 = 28,99$$

$$Y = -0,37*40,5 + 29,43 = 14,45$$

La representación gráfica aproximada es la siguiente.



Se puede apreciar que los valores estimados de la variable **Y** tienden a disminuir cuando aumentan los valores de **X**, por ello la correlación es inversa, a más empleo en el sector agrícola menor en el industrial

Cálculo de la recta de regresión de la variable **X** sobre la variable **Y** convirtiendo la variable **X** en variable dependiente, y la variable **Y** en variable independiente.

CUADRO DE DATOS

$\sum Y = 390,5$	$\sum X = 294,4$	$\sum Y^2 = 10230,25$	$\sum Y*X = 6150,76$
$n = 17$	$n = 17$		
$\bar{Y} = 22,97$	$\bar{X} = 17,31$		

FÓRMULA CONSTANTE b =

$$\frac{\sum Y*X - n*\bar{Y}*\bar{X}}{\sum Y^2 - n*\bar{Y}^2} = \frac{6150,76 - (17*22,97*22,97)}{10230,25 - (17*22,97*22,97)} = \frac{6150,76 - 6759,38}{10230,25 - 8969,56} = \frac{-608,62}{1260,69} = -0,48$$

$$\sum Y^2 - n*\bar{Y}^2 = 10230,25 - (17*22,97*22,97) = 10230,25 - 8969,56 = 1260,69$$

FÓRMULA CONSTANTE a =

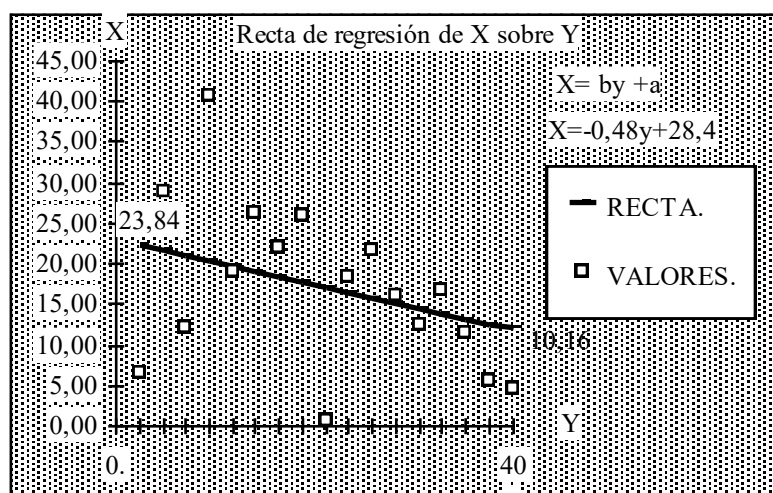
$$\bar{X} - b*\bar{Y} = 17,31 - (-0,48*22,97) = 17,31 + 11,89 = 29,20$$

$$\text{Recta de regresión de X sobre Y} = X = b*y + a = X = -0,48.y + 29,20$$

El valor mínimo y máximo para marcar la recta son los siguientes

$$X = -0,48*9,5 + 29,20 = 23,84$$

$$X = -0,48*38 + 29,20 = 10,16$$



La recta también tiene una relación inversa pero con unos valores diferentes al cambiar la nube de puntos.

EJERCICIO COMPLEMENTARIO

El análisis de regresión, permite calcular la forma de relación entre dos variables, pero esta no siempre se produce de forma exacta, produciéndose diferencias entre los valores observados y los valores estimados de la variable dependiente.

Estas desviaciones (valores observados menos valores estimados), reciben el nombre de residuales. Cuanto menor sea el valor del residual, más ajustada será la recta de regresión y más precisas serán las estimaciones.

En el siguiente ejemplo donde tomo la variable Y como variable dependiente es $(Y - Y'')$

Se toman los valores conocidos de la variable X , y se les aplica la formula de las constantes **a** y **b** de la recta de regresión de Y sobre X.

Por lo tanto los valores residuales se calculan aplicando a cada valor observado de la variable X la fórmula $Y'' = b \cdot x + a$, de esta manera sabemos el valor estimado en cada caso para Y. El siguiente paso es saber el valor residual, que se halla restando el valor estimado para Y al valor observado de Y.

VALORES OBSERVADOS-VALORES ESTIMADOS-RESIDUALES

Comunidades Autónomas	Empleo Agríc. = X	Empleo Indust. = Y	Y ²	Y'' = b*x+a -0,37x+29,4	Y'' ²	r = (Y - Y'')	r ²
Andalucía	19,40	15,90	252,81	22,25	495,15	-6,35	40,35
Aragón	16,90	26,90	723,61	23,18	537,17	3,72	13,86
Baleares	6,70	9,50	90,25	26,95	726,36	-17,45	304,54
Canarias	12,90	10,80	116,64	24,66	607,97	-13,86	192,02
Cantabria	21,70	21,00	441,00	21,40	458,00	-0,40	0,16
Castilla la M	26,20	21,40	457,96	19,74	389,51	1,66	2,77
Castilla Le.	25,50	20,20	408,04	20,00	399,80	0,21	0,04
Cataluña	5,60	36,10	1303,21	27,36	748,46	8,74	76,42
Comunidad V	13,40	28,70	823,69	24,47	598,88	4,23	17,88
Extremadura	29,40	10,20	104,04	18,55	344,18	-8,35	69,76
Galicia	40,50	15,30	234,09	14,45	208,66	0,86	0,73
Madrid	1,20	22,60	510,76	28,99	840,19	-6,39	40,78
Murcia	19,10	23,10	533,61	22,36	500,10	0,74	0,54
Navarra	11,30	34,90	1218,01	25,25	637,51	9,65	93,14
País Vasco	4,50	38,00	1444,00	27,77	770,90	10,24	104,76
Asturias	22,20	26,20	686,44	21,22	450,12	4,98	24,84
La Rioja	17,90	29,70	882,09	22,81	520,16	6,89	47,51
$\Sigma X = 294,4$ $\Sigma Y = 390,5$ $\Sigma Y^2 = 10230,25$ $\Sigma Y'' 391,38$ $\Sigma Y''^2 = 9233,11$							

Cálculo de la Varianza de los valores observados de la población activa industrial =

$$S^2 = \frac{\sum Y^2}{n} - \frac{Y^2}{n} = \frac{10230,25}{17} - \frac{22,972}{17} = 601,7 - 527,62 = 74,16$$

Cálculo de la Varianza de los valores estimados

$$S^2 = \frac{\sum Y''^2}{n} - \frac{Y^2}{n} = \frac{9233,11}{17} - \frac{22,972}{17} = 527,62 - 527,62 = 15,5$$

Para el cálculo de del porcentaje de variación total (S^2) que se explica por la recta de regresión, se relaciona la varianza de los valores observados y la varianza de los valores estimados. Esta relación establece el coeficiente de determinación.

r^2 = Coeficiente de determinación

S^2 = Varianza de los valores estimados de la variable dependiente
 Y''

S^2 Varianza de los valores observados
 Y

$$r^2 = \frac{S^2_{Y''}}{S^2_Y} = \frac{15,5}{74,15} = 0,21$$

Transformada esta razón en porcentajes, se puede decir que el 21 % de la variación de la variable dependiente se explica mediante la recta de regresión.

El análisis de los residuales de la variable Y demuestra un escaso ajuste entre los valores observados y los valores estimados, lo que demuestra que el 79 % no se explica mediante la recta de regresión. Por la tanto es de suponer que ello se produce porque en la configuración de lo valores de las dos variables están incidiendo otros factores que impiden una recta de regresión más ajustada, siendo en este caso las diferencias que provoca otra variable no contemplada como es el sector servicios. Luego, sería necesaria una nueva hipótesis de trabajo para configurar una recta de regresión más ajustada.

Los valores residuales indican el grado de ajuste entre los valores observados y los valores estimados de tal manera que aquellos que se sitúan más cercanos a 0 son los que los más ajustados. Para tener un índice de referencia de cada residual, se estandarizan el valor de los mismos.

Ello se hace dividiendo el valor absoluto del residual (r a) entre la desviación típica de la variable dependiente, ya que su valor depende de las unidades de medida de la misma.

RESIDUALES

FÓRMULA = r.e

(r.a) = VALOR DEL RESIDUAL

S y = DESVIACIÓN TÍPICA DE LA VARIABLE DEPENDIENTE = 8,61

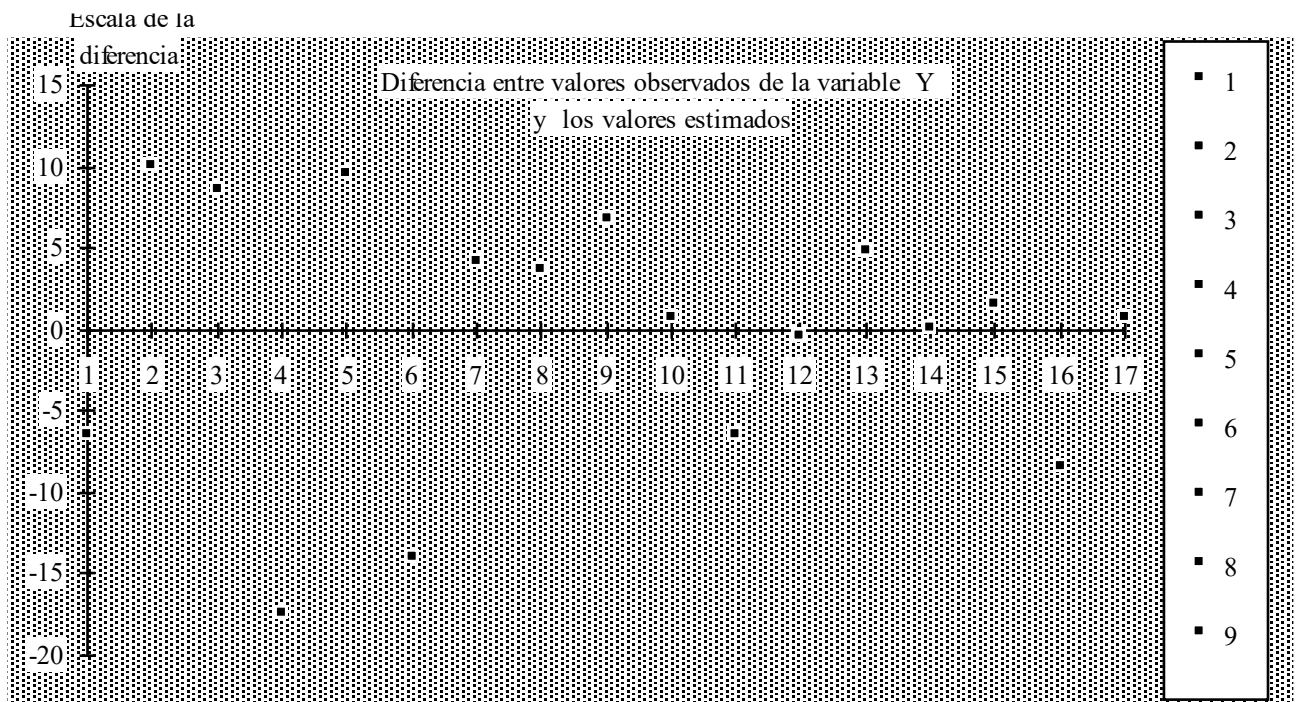
$r.e = \frac{r.a}{S_y}$

Sy

Vistos estos residuales espacialmente podemos determinar que Comunidades Autónomas se acercan más entre el valor observado y el estimado. Aquellas marcadas con una señal son en las que más se corresponden con ambos valores.

Comunidades Autónomas	Empleo agrícola = X	Empleo	r = (Y - Y")	FÓRMULA
		industrial = Y		re = ra/Sy
Andalucía	19,40	15,90	-6,35	-0,74
Aragón *	16,90	26,90	3,72	0,43
Baleares	6,70	9,50	-17,45	-2,03
Canarias	12,90	10,80	-13,86	-1,61
Cantabria *	21,70	21,00	-0,40	-0,05
Castilla la M *	26,20	21,40	1,66	0,19
Castilla León. *	25,50	20,20	0,21	0,02
Cataluña	5,60	36,10	8,74	1,02
Comunidad Val.*	13,40	28,70	4,23	0,49
Extremadura	29,40	10,20	-8,35	-0,97
Galicia *	40,50	15,30	0,86	0,10
Madrid	1,20	22,60	-6,39	-0,74
Murcia *	19,10	23,10	0,74	0,09
Navarra	11,30	34,90	9,65	1,12
País Vasco	4,50	38,00	10,24	1,19
Asturias*	22,20	26,20	4,98	0,58
La Rioja	17,90	29,70	6,89	0,80

En el presente gráfico la línea de abscisas refleja los valores estimados, los puntos son los valores observados, la línea de coordenadas es una escala que permite apreciar la diferencia entre los valores estimados y los valores observados.



SEGUNDA PARTE

EJERCICIO 1

Halla el centro de gravedad de la distribución de ciudades de Nottinghamshire y contrástalo con el centro de gravedad de la superficie delimitada por sus límites administrativos. ¿Cual es la distancia existente entre ambos? Calcular el valor del radio dinámico.

En la mayoría de los estadísticos espaciales el cálculo se realiza en función de las distancias hacia el este y hacia el norte. En la presente figura tiene un sistema de coordenadas con origen en el ángulo inferior izquierdo. El eje horizontal, que mide las distancias hacia el este, se ha denominado X, y el eje vertical que mide las distancias hacia el norte, se ha denominado Y .

CENTRO MEDIO Y CENTRO MEDIO PONDERADO

El centro medio se determina calculando la media de las coordenadas X (hacia el este) y la media de las coordenadas Y (hacia el norte) Estas dos coordenadas medias marcan la localización del centro medio de la distribución o centro de gravedad.

La formula: $\bar{X} = \frac{\sum X}{n}$ $\bar{Y} = \frac{\sum Y}{n}$

En el cálculo del centro medio se da a cada punto un peso idéntico. Sin embargo, lo normal es que los puntos que se miden no tengan el mismo peso, entonces se calcula el centro medio ponderado.

La formula = $\bar{X}_w = \frac{\sum X_w}{\sum W}$ $\bar{Y}_w = \frac{\sum Y_w}{\sum W}$

La **X** e **Y** en negrita indican las medias.

El mapa de la distribución de las ciudades de Nottinghamshire presenta los datos de las coordenadas hacia el este **X** seguidas de las **Y** hacia el norte y entre paréntesis las poblaciones totales en millares.

CIUDAD	COORDENADAS		PESO	COORDENADAS	
				PONDERADAS	
	(Estes) X	(Nortes) Y	W	Xw	Yw
East Retford	17	34	18	306	612
Worksop	9	33	21	189	693
Warsop	8	27	36	288	972
Ollerton	14	26	30	420	780
Mansfield Woodh.	7	23	25	175	575
Mansfield	6	22	58	348	1276
Sutton in Ashfield	4	21	41	164	861
Kirkbi in Ashfield	3	19	24	72	456
Southwell	16	17	27	432	459
Newark	23	17	25	575	425
Hucknall	6	15	26	156	390
Eastwood	2	13	11	22	143
Arnold	10	12	33	330	396
Carlton	10	10	45	450	450
Nottingham	7	9	300	2100	2700
bingham	17	8	39	663	312
Beeston	4	7	63	252	441
West bridgford	9	6	28	252	168
n = 18	$\sum X = 172$	$\sum Y = 319$	$\sum W = 850$	$\sum Xw = 7194$	$\sum Yw = 12109$

CENTRO DE GRAVEDAD

$$\bar{X} = \frac{\sum X}{n} = \frac{172}{18} = 9,56 \quad \bar{Y} = \frac{\sum Y}{n} = \frac{319}{18} = 17,72$$

CENTRO DE GRAVEDAD PONDERADO

$$\bar{X}_w = \frac{\sum Xw}{\sum W} = \frac{7194}{850} = 8,46 \quad \bar{Y}_w = \frac{\sum Yw}{\sum W} = \frac{12109}{850} = 14,25$$

CENTRO DE GRAVEDAD DE LA SUPERFICIE

Se calcula de una manera semejante a la indicada en el caso del centro de gravedad para puntos ponderados, siguiendo el siguiente procedimiento.

Se construye una red con una retícula cuadrangular que cubra toda la superficie, en el presente caso el mapa de Nottinghamshire.

Se seleccionan dos ejes (X e Y) y las líneas de las retículas dan unos valores numéricos. El eje vertical se ajusta al punto más occidental del mapa, y el eje horizontal al punto más meridional. Se calcula la media de los valores de Y (filas) y luego los valores de las X (columnas).

A cada cuadrícula se le da un valor de 1 y a cada cuadrícula parcial se le da un valor de 0,5.

Para la aplicación de la cuadrícula se ha tenido en cuenta la escala del mapa que viene determinada en millas. El valor de cada cuadro de la retícula lo represento en este cuadro-

En el presente caso he aplicado una retícula cuadrangular, elaborada con el ordenador donde cada cuadro tiene un valor equivalente a dos 2 millas de lado.

23°	Y						0,5	0,5	0,5						= 1,5
22°						0,5	0,5	0,5	0,5	0,5	0,5				= 3
21°						0,5	0,5	0,5	1	1	0,5				= 4
20°				0,5	0,5	1	1	1	1	1	0,5				= 5,5
19°				0,5	1	1	1	1	1	1	0,5				= 6
18°			0,5	0,5	1	1	1	1	1	1	0,5	0,5			= 7
17°			0,5	1	1	1	1	1	1	1	1	0,5			= 8
16°			0,5	1	1	1	1	1	1	1	0,5	0,5			= 7,5
15°		0,5	0,5	1	1	1	1	1	1	1	0,5	0,5	0,5	0,5	= 9
14°		0,5	0,5	1	1	1	1	1	1	1	1	0,5	0,5		= 9
13°	0,5	0,5	0,5	1	1	1	1	1	1	1	1	0,5			= 9
12°	0,5	0,5	1	1	1	1	1	1	1	1	1	0,5			= 10,5
11°	0,5	1	1	1	1	1	1	1	1	1	1	0,5			= 11
10°	0,5	1	1	1	1	1	1	1	1	1	1	0,5	0,5		= 11,5
9°	0,5	1	1	1	1	1	1	1	1	1	1	0,5	0,5		= 11,5
8°	0,5	1	1	1	1	1	1	1	1	1	0,5	0,5			= 10,5
7°	0,5	0,5	1	1	1	1	1	1	1	1	0,5				= 9,5
6°		0,5	1	1	1	1	1	1	1	0,5	0,5				= 8,5
5°		0,5	1	1	1	1	1	1	1	0,5					= 8
4°		0,5	0,5	1	1	1	1	1	0,5	0,5					= 7
3°		0,5	1	1	1	1	1	0,5							= 6
2°		0,5	1	1	0,5	0,5	0,5	0,5							= 4,5
1°		0,5	0,5	0,5	0,5	0,5	0,5								= 3
	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	X

$$(3)(8,5)(12,5)(14,5)(17,5)(19,5) (20,5)(19,5)18,5)(17) (12) (5,5) (2) (0,5) = 171$$

CÁLCULO DEL CENTRO DE GRAVEDAD DE LA SUPERFICIE DE (Nottinghamshire)

FILAS (Y)	Ponderación		COLUMNAS (X)	Ponderación	
Y	(w)	Y*w	X	(w)	X*w
0,5	3	1,5	0,5	3	1,5
1,5	4,5	6,75	1,5	8,5	12,75
2,5	6	15	2,5	12,5	31,25
3,5	7	24,5	3,5	14,5	50,75
4,5	8	36	4,5	17,5	78,75
5,5	8,5	46,75	5,5	19,5	107,25
6,5	9,5	61,75	6,5	20,5	133,25
7,5	10,5	78,75	7,5	19,5	146,25
8,5	11,5	97,75	8,5	18,5	157,25
9,5	11,5	109,25	9,5	17	161,5
10,5	11	115,5	10,5	12	126
11,5	10,5	120,75	11,5	5,5	63,25
12,5	9	112,5	12,5	2	25,00
13,5	9	121,5	13,5	0,5	6,75
14,5	9	130,5		$\sum w = 171$	$\sum X*w = 1101,5$
15,5	7,5	116,25			
16,5	8	132		10,81	6,44
17,5	7	122,5		21,61	12,88
18,5	6	111,00			
19,5	5,5	107,25			
20,5	4	82			
21,5	3	64,5			
22,5	1,5	33,75			
	$\sum w = 171$	$\sum Y*w = 1848$			

$$Y_w = \frac{\sum Y*w}{\sum w} = \frac{1848}{171} = 10,81 \quad X_w = \frac{\sum X*w}{\sum w} = \frac{1101,5}{171} = 6,44$$

El centro de gravedad se encuentra en el punto de la cuadrícula 10,81 hacia el norte en intersección con el punto de la cuadrícula 6,44 hacia el este.

Puesto que el lado de cada cuadrícula tiene un valor de 2 millas.

Tomando como referencia los ejes de la retícula que coinciden con los extremos occidental y meridional del mapa del condado.

El valor de Y = 10,81*2 = **21,61** millas dirección norte

El valor de X = 6,44*2 = **12,88** millas dirección este

CÁLCULO PARA HALLAR LAS DISTANCIAS DE CADA CIUDAD AL CENTRO DE GRAVEDAD APLICANDO EL TEOREMA DE PITÁGORAS

COORDENADAS DEL CENTRO DE GRAVEDAD

$$X = 9,56 \quad Y = 17,72$$

	X	Y	a = X-X	b = Y-Y	(X-X) ²	(Y-Y) ²	d ²	d = √ d ²
East Retford	17,00	34,00	7,44	16,28	55,35	265,04	320,39	17,90
Worksop	9,00	33,00	-0,56	15,28	0,31	233,48	233,79	15,29
Warsop	8,00	27,00	-1,56	9,28	2,43	86,12	88,55	9,41
Ollerton	14,00	26,00	4,44	8,28	19,71	68,56	88,27	9,40
Mansfield Woodh.	7,00	23,00	-2,56	5,28	6,55	27,88	34,43	5,87
Mansfield	6,00	22,00	-3,56	4,28	12,67	18,32	30,99	5,57
Sutton in Ashfield	4,00	21,00	-5,56	3,28	30,91	10,76	41,67	6,46
Kirkbi in Ashfield	3,00	19,00	-6,56	1,28	43,03	1,64	44,67	6,68
Southwell	16,00	17,00	6,44	-0,72	41,47	0,52	41,99	6,48
Newark	23,00	17,00	13,44	-0,72	180,63	0,52	181,15	13,46
Hucknall	6,00	15,00	-3,56	-2,72	12,67	7,40	20,07	4,48
Eastwood	2,00	13,00	-7,56	-4,72	57,15	22,28	79,43	8,91
Arnold	10,00	12,00	0,44	-5,72	0,19	32,72	32,91	5,74
Carlton	10,00	10,00	0,44	-7,72	0,19	59,60	59,79	7,73
Nottingham	7,00	9,00	-2,56	-8,72	6,55	76,04	82,59	9,09
bingham	17,00	8,00	7,44	-9,72	55,35	94,48	149,83	12,24
Beeston	4,00	7,00	-5,56	-10,72	30,91	114,92	145,83	12,08
West bridgford	9,00	6,00	-0,56	-11,72	0,31	137,36	137,67	11,73
							Σd ² = 1814,06	

Una vez conocidas las distancias respecto del centro de gravedad (en el presente caso aplicando a cada punto observado el teorema de Pitágoras), se puede hallar también la desviación típica de las distancias con la siguiente fórmula

S =

$$\sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{1814,06}{18}} = \sqrt{100,78} = \mathbf{10,04}$$

RADIO DINÁMICO

El radio dinámico mide la relación de dos fuerzas, la expansión y contracción de un hecho geográfico sobre el espacio a partir de un núcleo con respecto a otros núcleos en una región. En el presente caso el núcleo es el centro medio de gravedad, y el resto de núcleos las ciudades con sus poblaciones.

p = a la población de los restante núcleos considerados,

d = a la distancia desde el núcleo de partida a los restantes núcleos

P = a la población total de la región considerada.

$$\text{RADIO DINÁMICO } r = \sqrt{\frac{\sum p \cdot d^2}{P}}$$

El radio dinámico en el espacio geográfico de Nottinghamshire es el siguiente.

TABLA PARA EL CÁLCULO DEL RADIO DINÁMICO

CIUDAD	Distancia al centro de gravedad (d)	d ²	Ponderación Población en miles (p)	p*d ²
East Retford	17,90	320,39	18	5767,056
Worksop	15,29	233,79	21	4909,63
Warsop	9,41	88,55	36	3187,87
Ollerton	9,40	88,27	30	2648,16
Mansfield Woodh.	5,87	34,43	25	860,80
Mansfield	5,57	30,99	58	1797,54
Sutton in Ashfield	6,46	41,67	41	1708,55
Kirkbi in Ashfield	6,68	44,67	24	1072,13
Southwell	6,48	41,99	27	1133,78
Newark	13,46	181,15	25	4528,80
Hucknall	4,48	20,07	26	521,87
Eastwood	8,91	79,43	11	873,75
Arnold	5,74	32,91	33	1086,10
Carlton	7,73	59,79	45	2690,64
Nottingham	9,09	82,59	300	24777,60
bingham	12,24	149,83	39	5843,45
Beeston	12,08	145,83	63	9187,42
West bridgford	11,73	137,67	28	3854,82
			$\Sigma p = P = 850$	$\Sigma p*d = 76449,96$

$$\text{RADIO DINÁMICO} = r = \sqrt{\frac{\sum p*d^2}{P}} = \sqrt{\frac{76449,96}{850}} = \sqrt{89,94} = \mathbf{9,48}$$

El radio dinámico valorado en distintos periodos de tiempo puede indicar los cambios ocurridos en una distribución espacial a través del tiempo.

El contraste y la distancia entre el centro de gravedad de la superficie y el centro de gravedad de la distribución de las ciudades y el centro ponderado, lo planteo de dos maneras, a través de comparar los valores numéricos de las coordenadas y través de su representación visual en el mapa.

LOS EJES PARA MEDIR LAS DISTANCIA DE LAS CIUDADES

Los ejes desde los que teóricamente están dadas las coordenadas de las ciudades se pueden hallar tomando las referencias de una o varias ciudades y midiendo en sentido contrario.

Al realizar este cálculo he encontrado en el mapa algunas inexactitudes.

Los valores de las coordenadas que se ofrecen en millas para determinar la posición de las ciudades, están redondeados a la unidad. Ello se puede apreciar observando por ejemplo la ciudad de Mansfield

W. que está a siete millas dirección este al igual que Nottingham, pero los valores no están en línea con la vertical del mapa. Lo mismo ocurre con los valores en dirección norte de las ciudades de Southwell y Newark, ambas están a 17 millas dirección norte, pero los valores no están exactamente con la línea horizontal del mapa.

En el caso de otras ciudades no se ajustan exactamente las coordenadas con la escala del mapa, como es el caso de Neewark a 23 millas dirección este, 16 millas más que Nottingham sin que guarde la proporción debida, o el caso de East Retford a 34 millas dirección norte, 17 millas más que Newark sin que tampoco guarde la proporción debida, coincidiendo estos desajustes con las ciudades más alejadas de los ejes teóricos de referencia.

Esta falta de exactitud impide trazar unos ejes coincidentes con todas las coordenadas de las ciudades, ello me ha llevado a adoptar de manera convencional los ejes meridional y occidental de la retícula (por ser los más aproximados) como los ejes de referencia, para desde de ellos situar los distintos centros de gravedad en el mapa y comparar las distancias.

POSICIÓN DE LOS DISTINTOS VALORES DE LOS CENTROS DE GRAVEDAD

El siguiente cuadro muestra los valores de los distintos centros de gravedad

	X (Estes) Millas	Y (Norte) Millas
Centro de gravedad	9,56	17,72
Centro de gravedad ponderado	8,46	14,25
Centro de gravedad de la superficie	12,88	21,61

Las distancias del centro de gravedad y del centro de gravedad ponderado respecto del centro de gravedad de la superficie, aplicando el Teorema de Pitágoras es el siguiente:

Centro gravedad de la superficie $X = 12,88$ millas $Y = 21,61$ millas

	X	Y	a = X-X	b = Y-Y	X-X ²	Y-Y ²	d ² = (x-X) ² +(y-Y) ²	d = $\sqrt{d^2}$
Centro gravedad	9,56	17,72	-3,32	-3,89	11,02	15,13	26,15	5,11
Centro grav. P	8,46	14,25	-4,42	-7,36	19,54	54,17	73,71	8,59

El centro de gravedad se sitúa 5,11 millas del centro de la superficie.

El centro de gravedad ponderado se sitúa a 8,59 millas del centro de la superficie.

Los valores demuestran que el mayor peso de las ciudades y de su población, se encuentran situados ligeramente hacia el suroeste del centro de gravedad de la superficie.

Este desplazamiento es debido a que en esa dirección se encuentran situadas el mayor número de ciudades, que a su vez concentran más población, como es el caso de Nottingham.

Una vez conocidos los distintos valores, la posición en el mapa queda marcada de la siguiente manera:

CP = Centro ponderado Y = 14,25
 X = 8,46

RD = Radio dinámico = 9,48

Los ejes desde los que se miden las distancias son los de la retícula

EJERCICIO COMPLEMENTARIO

La desviación típica de las distancias mide la dispersión de los puntos en torno al valor central, y su valor puede considerarse como el radio de círculo que mide la dispersión en torno al centro de gravedad.

Otra fórmula para calcular la desviación típica de las distancias si no se conocen las distancias de las ciudades al centro medio.

COORDENADAS	X	Y	X ²	Y ²
East Retford	17	34	289	1156
Worksop	9	33	81	1089
Warsop	8	27	64	729
Ollerton	14	26	196	676
Mansfield Woodh.	7	23	49	529
Mansfield	6	22	36	484
Sutton in Ashfield	4	21	16	441
Kirkbi in Ashfield	3	19	9	361
Southwell	16	17	256	289
Newark	23	17	529	289
Hucknall	6	15	36	225
Eastwood	2	13	4	169
Arnold	10	12	100	144
Carlton	10	10	100	100
Nottingham	7	9	49	81
bingham	17	8	289	64
Beeston	4	7	16	49
West bridgford	9	6	81	36
n = 18	$\sum X = 172$	$\sum Y = 319$	$\sum X^2 = 2200$	$\sum Y^2 = 6911$
	$\bar{X} = 9,56$	$\bar{Y} = 17,72$		
	$\sum X^2 = 91,3936$	$\sum Y^2 = 313,9984$	$2200/18 = 122,22$	$6911/18 = 383,94$

DESVIACIÓN TÍPICA DE LAS DISTANCIAS S =

$$\sqrt{\left(\frac{\sum X^2}{n} - \bar{X}^2\right) + \left(\frac{\sum Y^2}{n} - \bar{Y}^2\right)} = \sqrt{\left(\frac{\sum 220}{18} - 9,56^2\right) + \left(\frac{\sum 691}{18} - 17,72^2\right)} =$$

$$\sqrt{(122,22 - 91,39) + (383,94 - 313,99)} = \sqrt{30,83 + 69,95} = \sqrt{100,78} = \mathbf{10,04}$$

LA DESVIACIÓN TÍPICA PONDERADA DE LAS DISTANCIAS

La desviación típica ponderada de las distancias mide la dispersión de los puntos en torno al valor central, y su valor puede considerarse como el radio de círculo que mide la dispersión en torno al centro de gravedad ponderado

TABLA PAR EL CÁLCULO DE LA DESVIACIÓN TÍPICA PONDERADA DE LAS DISTANCIAS

	X	Y	X ²	Y ²	PESO	X ² *w	Y ² *w
East Retford	17	34	289	1156	18	5202	20808
Worksop	9	33	81	1089	21	1701	22869
Warsop	8	27	64	729	36	2304	26244
Ollerton	14	26	196	676	30	5880	20280
Mansfield Woodh.	7	23	49	529	25	1225	13225
Mansfield	6	22	36	484	58	2088	28072
Sutton in Ashfield	4	21	16	441	41	656	18081
Kirkbi in Ashfield	3	19	9	361	24	216	8664
Southwell	16	17	256	289	27	6912	7803
Newark	23	17	529	289	25	13225	7225
Hucknall	6	15	36	225	26	936	5850
Eastwood	2	13	4	169	11	44	1859
Arnold	10	12	100	144	33	3300	4752
Carlton	10	10	100	100	45	4500	4500
Nottingham	7	9	49	81	300	14700	24300
bingham	17	8	289	64	39	11271	2496
Beeston	4	7	16	49	63	1008	3087
West bridgford	9	6	81	36	28	2268	1008
	X_w = 8,46	Y_w = 14,25	X_{2w} = 71,57	Y_{2w} = 203,06	Σw = 850	ΣX_{2w} = 77436	ΣY_{2w} = 221123

Desviación típica ponderada de las distancias (Sp) =

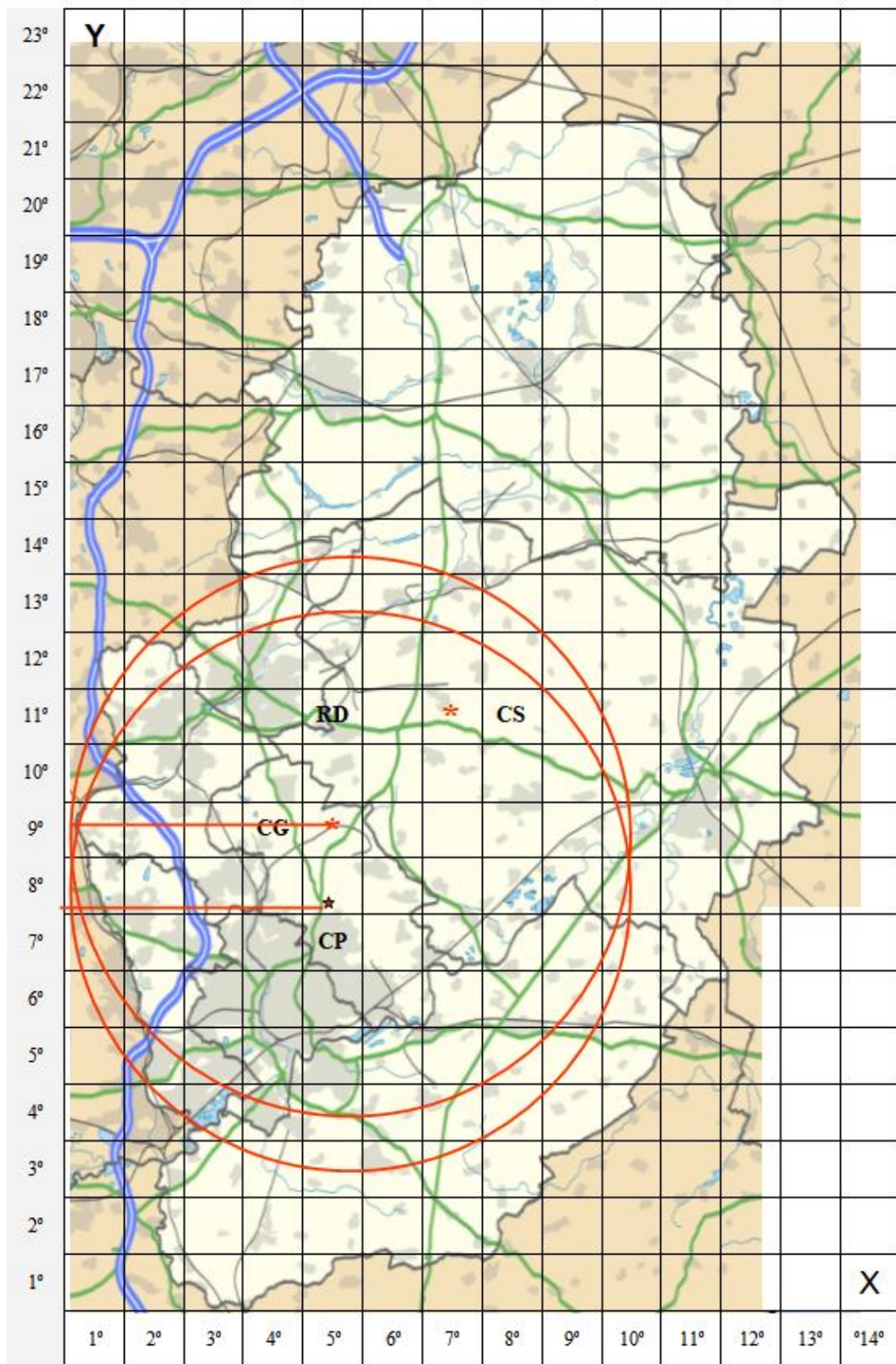
$$\sqrt{\left(\frac{\sum X_{2w}}{\sum w} - X_{2w}\right) + \left(\frac{\sum Y_{2w}}{\sum w} - Y_{2w}\right)} = \sqrt{\left(\frac{77436}{850} - 8,46^2\right) + \left(\frac{221123}{850} - 14,25^2\right)} =$$

$$\sqrt{(91,10 - 71,57) + (260,14 - 203,06)} = \sqrt{19,52 + 57,08} = \sqrt{76,61} = \mathbf{8,75}$$

DESVIACIÓN TÍPICA DE LAS DISTANCIAS = 10,04

DESVIACIÓN TÍPICA PONDERADA = 8,75

El radio del círculo que mide la dispersión en torno al centro de gravedad, queda reflejado en el presente gráfico. De igual manera se representa el radio que mide la dispersión en torno al centro de gravedad ponderado



EJERCICIO NÚMERO 2

Aplicando el análisis de vecindad al plano de la figura, decidir si su estructura espacial se adecua a una distribución dispersa o concentrada.

ANÁLISIS DE VECINDAD

El análisis de vecindad es una técnica desarrollada por ecologistas botánicos (Clark y Evans en 1954) destinada a medir la estructura en función de la disposición de un conjunto de puntos.

El método seguido es, en primer lugar se mide la distancia observada (d) de cada punto con respecto al punto más próximo. Calculando el promedio de todas estas distancias se obtiene un valor (d_o). El número de distancias a los vecinos más próximos es el mismo que el número de puntos.

$$d_o = \frac{\sum d}{n}$$

En segundo lugar se calcula cuál sería el promedio de las distancias esperadas de los puntos, con respecto a los puntos más próximos si estuviesen distribuidos al azar (aleatoriamente).

d_a = la distancia media esperada al vecino más próximo de una disposición aleatoria de puntos

A = al área de la región estudiada

n = el número de puntos

$$d_a = \frac{1}{2} \sqrt{\frac{n}{A}}$$

De esta manera, para cualquier distribución de puntos se puede calcular el promedio de las distancias observadas (d_o) y relacionarlo con la distancia media esperada si los puntos estuviesen dispuestos aleatoriamente (d_a).

El índice de vecindad proporciona una medida en un solo valor. Mide la divergencia entre estos dos conjuntos de valores y se calcula dividiendo la distancia media real observada, por la distancia media esperada al vecino más próximo si los puntos estuvieran distribuidos aleatoriamente.

$$R = \frac{d_o}{d_a}$$

El índice R , indica el valor de la diferencia y el grado de separación o concentración de los puntos con respecto a la distribución aleatoria.

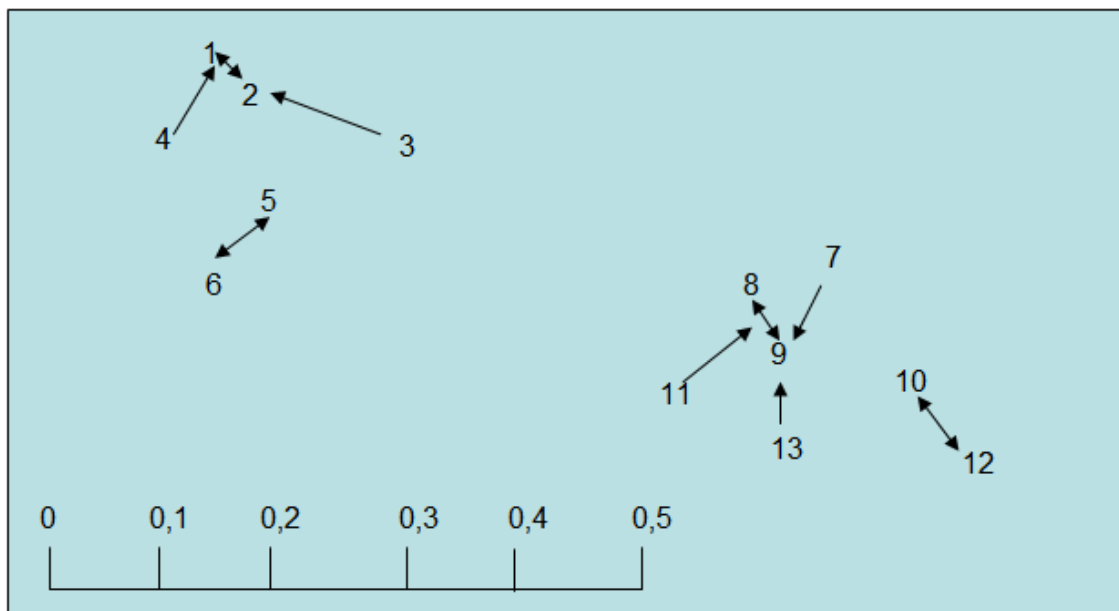
Este índice R (índice del elemento más próximo) está comprendido entre 0 y 2,149.

El valor 0 representa una situación en la que d_o es igual a 0, o todas las distancias entre los puntos, son 0, por lo tanto todos los puntos tienen la misma localización y se da una concentración absoluta.

Cuando R es igual a 1, y d_a es igual d_o , significa que los puntos se distribuyen aleatoriamente, si R es igual a 2, existe una disposición regular de los puntos. formando cuadrados iguales . Cuando R es igual a 2,149, los puntos se distribuyen formando triángulos equiláteros tendiendo los valores mayores a la dispersión.

< Aumento a la concentración <				> Aumento a la dispersión >			
0	0,333	0,667	1	1,333	1,667	2	2,149

En el presente ejercicio los puntos son las casas en venta, la distancia está medida en kilómetros, el número de observaciones es de 13, y el área es de 0,29 Km².



Para calcular la distancia entre puntos se ha hallado el valor de un milímetro en la escala.

La escala tiene 500 metros en una línea de 75 milímetros, luego un milímetro representa = $500 / 75 = 6,666$ metros.

CÁLCULO DEL VECINO MÁS PRÓXIMO

CASAS	Casa más próxima	Distancia en milímetros	Distancia en metros = mm * 6,666
1	2	5	33,33
2	1	5	33,33
3	2	20	133,33
4	1	15,5	103,33
5	6	4	26,67
6	5	4	26,67
7	9	9	60,00
8	9	9	60,00
9	8	9	60,00
10	12	13	86,67
11	8	16	106,67
12	10	13	86,67
13	9	15	100,00
		137,5	$\Sigma d = 916,67$

Puesto que los valores hay que calcularlos en las mismas unidades de medida transformo los metros en kilómetros = $916,67 / 1000 = 0,91667$

El promedio de las distancias de los valores observados =

$$d_o = \frac{\sum d}{n} = \frac{0,91667}{13} = 0,0705$$

El promedio de las distancias de los valores esperados, en el caso de una distribución aleatoria =

$$d_a = \frac{1}{2 \sqrt{\frac{n}{A}}} = \frac{1}{2 \sqrt{\frac{13}{0,29}}} = \frac{1}{2 \sqrt{44,82}} = \frac{1}{2 * 6,69} = \frac{1}{13,39} = 0,0746$$

El índice de vecindad =

$$R = \frac{d_o}{d_a} = \frac{0,0705}{0,0746} = 0,945$$

El índice de vecindad presenta una ligera tendencia a la concentración pero cerca del valor 1 de aleatoriedad.

EJERCICIO NÚMERO 3

El mapa de la figura represente las relaciones por carretera existentes entre diversos núcleos urbanos del Sur de Gales y Suroeste de Inglaterra. A partir del mismo dibujar la red topológica, Calculando las principales medidas de conexión y accesibilidad indicando la importancia relativa de sus arcos.

ANÁLISIS TOPOLÓGICO

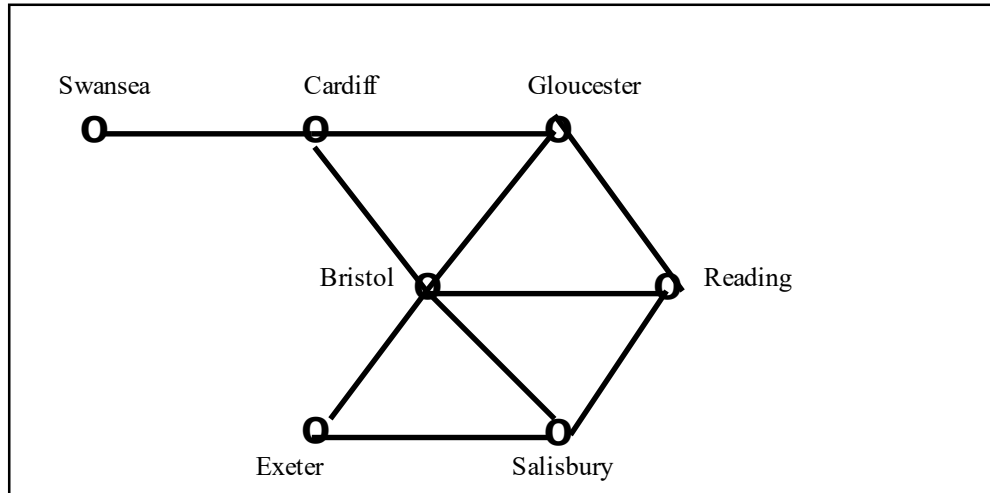
La topología estudia la posición y las relaciones entre puntos, líneas y superficies pero no tiene en cuenta la forma ni el tamaño de las áreas.

Una red topológica tiene rasgos comunes con una red de transporte, está constituida por arcos que equivalen a rutas y nodos que representan núcleos o intersecciones de caminos.

La red tiene un número finito de lugares

Cada ruta une dos lugares diferentes

RED TOPOLÓGICA DEL SUR DE GALES Y SUROESTE DE INGLATERRA



En una red topológica, el segmento que une dos puntos a nodos se llama arco.

El tramo es un conjunto de arcos que une varios nodos.

La longitud topológica de un tramo, es el número de arcos que tiene.

La distancia topológica más corta entre dos nodos viene dada por el tramo más corto que los une.

Un circuito es un tramo que empieza y acaba en el mismo nodo.

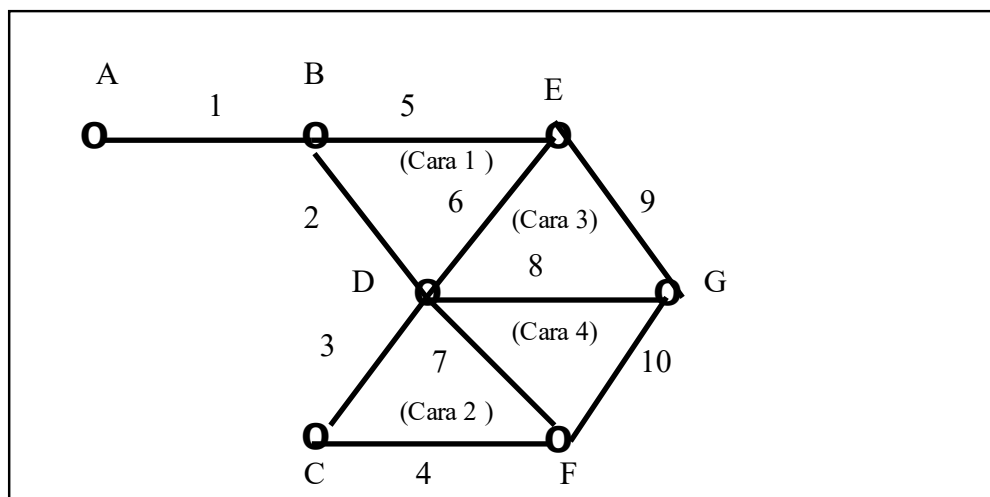
Las regiones o caras, son superficies completamente cerradas por un circuito.

Una red topológica es nula, cuando sólo consta de nodos.

Se considera relacionada cuando sus nodos están unidos por arcos.

Una red topológica relacionada sin ningún circuito se llama árbol

Se considera a una red relacionada cuando sus nodos están unidos por arcos



En el presente gráfico los círculos representan a los nodos y las líneas a los arcos.

MEDIDAS DE CONEXIÓN

a = arcos

n = nodos

El índice beta mide el nivel de conexión entre los arcos y los nodos. Es el cociente de dividir el número de arcos (a) entre el número de nodos (n).

Las redes simples tienen un valor inferior a 1. Una red conectada que forma un circuito tiene valor 1.

Las redes más complejas tienen valores superiores a 1

ÍNDICE BETA

$$B = \frac{a}{n} = \frac{10}{7} = 1,42 = \text{Red compleja}$$

El número ciclomático indica el número de circuitos existentes en una red, y se calcula restando del número total de arcos (a), el número de nodos menos 1

$$\text{Número ciclomático} = a - (n - 1) = 10 - (7 - 1) = 10 - 6 = 4$$

El índice alfa resulta de dividir el número de circuitos entre el mayor número de circuitos posibles en una red. Mide por lo tanto el índice de circuitos existentes en relación con los posibles

$$\alpha = \frac{\text{Nº ciclomático}}{2n - 5} = \frac{4}{2 \cdot 7 - 5} = \frac{4}{9} = 0,44$$

Es decir tiene el 44 % de los circuitos posibles

MEDIDAS DE ACCESIBILIDAD

Topológicamente, un nodo es tanto más accesible, cuanto mayor es el número de arcos que lo unen a la red.

En una red compleja, es necesario confeccionar una matriz para determinar la accesibilidad relativa de los nodos.

Los números de las casillas, indican el número de arcos del tramo más corto que une todos los posibles pares de nodos.

Matriz de datos de la red topológica del Sur de Gales y del Suroeste de Inglaterra

	Swansea	Cardiff	Bristol	Exeter	Gloucester	Salisbury	Reading	Nº Asociado	Índice de Shimbel
Swansea		1	2	3	2	3	3	3	14
Cardiff	1		1	2	1	2	2	2	9
Bristol	2	1		1	1	1	1	2	7
Exeter	3	2	1		2	1	2	3	11
Gloucester	2	1	1	2		2	1	2	9
Salisbury	3	2	1	1	2		1	3	10
Reading	3	2	1	2	1	1		3	10

Accesibilidad total de la red. Índice de dispersión = 70

Confeccionada la matriz se puede determinar las siguientes medidas de accesibilidad.

Número asociado de un nodo. Es el número mínimo de arcos necesario para unirlo con el nodo más distante topológicamente de él (penúltima columna de la matriz)

Índice Shimbel. Indica el número de arcos necesarios para unir cualquier nodo, con los demás de la red (última columna de la matriz).

La accesibilidad total de la red. Se obtiene mediante el llamado índice de dispersión, que resulta de sumar los índices de Shimbel de cada uno de los nodos de la red.

La matriz permite señalar los núcleos menos accesibles, Swansea, Salisbury, Reading, Exeter, que son aquellos que se encuentran más en la periferia siendo su distancia topológica de 3 . Los más accesibles son Cardiff, Bristol y Gloucester con una distancia topológica de 2 como se desprende del número asociado respectivo. El índice Shimbel, da una accesibilidad mayor a Bristol pues necesita el menor número de arcos para unir cualquier nodo.

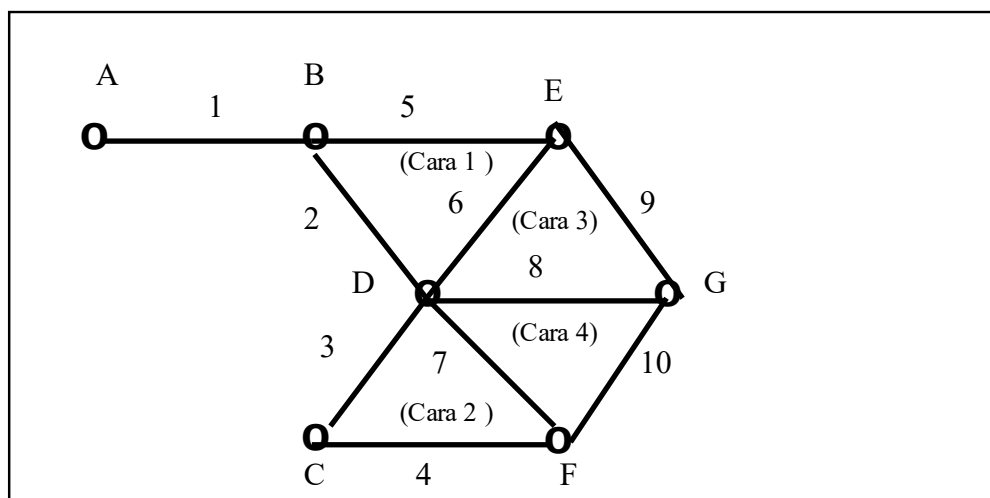
La accesibilidad total de la red de la red topológica es de 70.

LOS TRAMOS MÁS FRECUENTADOS

La necesidad de usar los tramos para unir los distintos nodos determina una mayor o menor frecuencia de uso de unos u otros tramos.

En primer lugar se consideran los tramos más cortos entre los diferentes nodos de la red.

Se numeran los diferentes arcos y se construye una tabla de conteo.



PARES POSIBLES DE NODOS

	B	C	D	E	F	G
A	/	/	/	/	/	/
	B	/	/	/	/	/
		C	/	/	/	/
			D	/	/	/
				E	/	/
					F	/

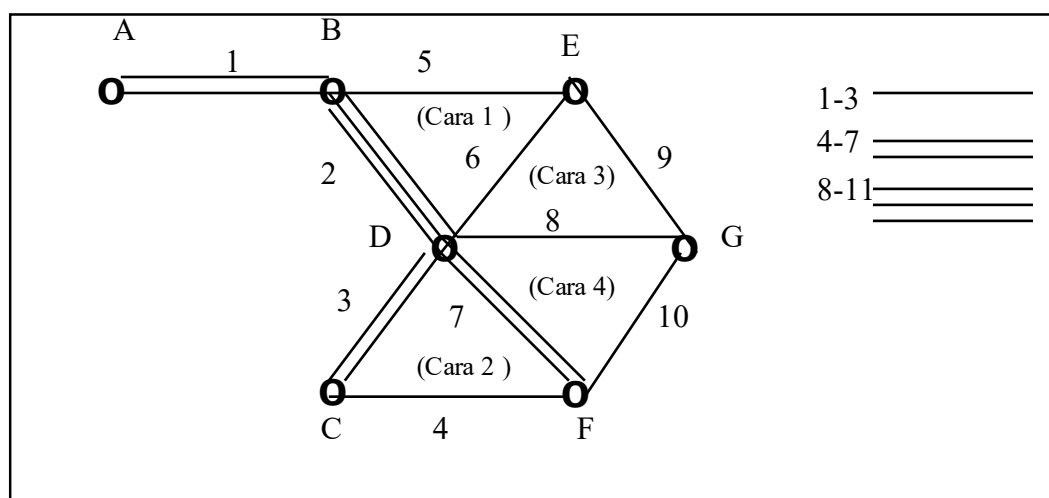
Cada trazo en las casillas indica que se han considerado todos los tramos más cortos que unen los diferentes nodos (Tabla de conteo)

PARES de	A	A	A	A	A	A	B	B	B	B	B	C	C	C	C	D	D	D	E	E	F	
NODOS	B	C	D	E	F	G	C	D	E	F	G	D	E	F	G	E	F	G	F	G	G	TOTAL
Arco 1	1	1	1	1	1	1																6
Arco 2		1	1		1	1	1	1		1	1											8
Arco 3		1					1					1	1									4
Arco 4														1	1							2
Arco 5				1					1													2
Arco 6													1			1			1			3
Arco 7					1					1							1		1			4
Arco 8						1					1							1				3
Arco 9																				1		1
Arco10															1						1	2

En los tramos de igual recorrido en número de arcos he utilizado los arcos de menor recorrido en kilómetros en el mapa real. Casos (A G por el interior) (B G por el interior) (C G por el exterior) (E F por el interior)

El siguiente gráfico muestra la frecuencia de uso de los arcos. La escala refleja la frecuencia de uso de cada arco

ACCESIBILIDAD RELATIVA DE LOS ARCOS



EJERCICIO COMPLEMENTARIO

TRAZADO REAL Y TRAZADO IDEAL

El grado de simplicidad y de inadecuación de una red, se pone también de manifiesto cuando se compara el trazado real, con el ideal. Los trazados ideales son los trazados más cortos. Los trazados reales son el resultado entre las distancias más cortas que median entre los lugares y la necesidad de vencer los obstáculos.

El presente cuadro refleja las distancias reales entre ciudades y las distancias ideales, y la desviación del trazado entre ambos modelos.

La distancia ideal la he hallado tomando como referencia la escala de 100 Kilómetros del mapa con una longitud de 38 milímetros. Cada milímetro tiene un valor en la escala de 2,63 Kilómetros. (Los números están redondeados)

Tramo entre ciudades	Número del arco	Distancia real Km	mm	mm*2,63 = Distancia ideal Km	Desviación del trazado = (Dist. real / Dist. ideal) %
Swansea - Cardiff	1	72	22	58	24 %
Cardiff - Bristol	2	71	15	40	77 %
Bristol- Exeter	3	122	41	108	12 %
Exeter- Salisbury	4	145	52	137	5 %
Cardiff- Gloucester	5	90	31	82	9 %
Bristol - Gloucester	6	57	21	55	3 %
Bristol - Salisbury	7	82	30	79	3 %
Bristol - Reading	8	124	45	118	5 %
Gloucester - Reading	9	115	40	105	9 %
Salisbury - Reading	10	86	27	71	21 %

El arco número 2 que se corresponde con la distancia entre Swansea y Cardiff tiene la máxima desviación. La desviación mínima corresponde a los arcos números 6 y 7.

EJERCICIO NÚMERO 4

Aplicar el análisis de formas según los índices descritos al condado de Gloucestershire antes y después de la reforma de la administración local en 1974.

La forma de las unidades espaciales, objeto primordial de la Geografía puede definirse como, la cualidad de un objeto que depende de relaciones constantes de posición y distancia de todos los puntos que constituyen su contorno. El concepto de forma es difícil de evaluar con precisión, pero puede comprenderse de un modo indirecto a través de algunas de sus características más destacadas. Compacidad, alargamiento, dirección....

El índice que suele emplearse para describir formas o contornos de las unidades espaciales, es el llamado índice de compacidad.

Se trata en realidad de una medida de desviación de una forma, en relación a la forma más compacta posible, un círculo. Un círculo es la forma más compacta en el sentido que tiene el menor perímetro posible en relación al área contenida en su interior.

El índice de compacidad en su versión más elemental, es el que relaciona el eje mayor (L1) y el menor (L2) de la unidad territorial considerada. El eje mayor ha de ser el eje más largo que pueda trazarse en la forma que se estudia. El eje menor es el eje más largo que puede trazarse perpendicularmente al eje mayor.

El índice de compacidad = $(I_c) = L_2 ./ L_1$

Si los ejes coinciden en su longitud el valor es igual a 1, si el eje menor, es menor que el mayor el índice será menor de uno 1.

Este índice tiene el problema que puede ofrecer los mismos valores para formas diferentes.

Otros índices que miden la compacidad son.

El índice de J. Cole, define la forma de las unidades geográficas a través del coeficiente de compacidad y para ello divide la superficie geográfica de estudio (A) entre el círculo mínimo circunscrito a dicha unidad geográfica, (Ax) $I_c = A / A_x * 100$.

El índice de Cole se define también $I_c = 4 A / \pi D^2$

Siendo D el diámetro del círculo menor circunscrito

El índice de Gibbs $I_G = 4A / \pi L^2$

Siendo A, el área de la unidad geográfica y L, la longitud del eje más largo.

El índice $I(s)$ Es la razón entre el radio del círculo mínimo circunscrito (Rc) y el radio que tiene la misma superficie que la unidad territorial considerada RA $I(s) = R_A / R_c$

El índice $I(T)$ es la relación entre el radio del círculo mínimo circunscrito (Rc) y el radio del círculo máximo inscrito (Ri) $I(T) = R_i / R_c$

Cada uno de estos índices oscilan entre 1, forma más compacta (círculo) a 0.

APLICACIÓN DE LOS PRESENTES ÍNDICES AL CONDADO DE GLOUCESTERSHIRE ANTES Y DESPUÉS DE LA REFORMA DE LA ADMINISTRACIÓN LOCAL EN 1974.

El cálculo de las distancias de los datos de aplicación para la resolución de los índices, teniendo en cuenta que la escala que se ofrece está en millas, lo he realizado de la siguiente manera.

El valor de diez millas según la medición efectuada con el calibre (Pie de Rey) equivale a una de las divisiones estándar de la pulgada $3 / 8$ de pulgada = (9,52 milímetros)

El valor de un milímetro en millas en esta escala = $10 * 1 / 9,52 = 1,05$.millas

Este valor se puede hallar también dividiendo la distancia de la escala por el número de divisiones de la misma.

Distancia de la escala en milímetros = 47,5

Número de divisiones de la escala = 50

Valor de un milímetro en millas = $50 * 1 / 47,5 = 1,05$

ÍNDICE DE COMPACIDAD

DATOS DE APLICACIÓN (Los datos están redondeados)

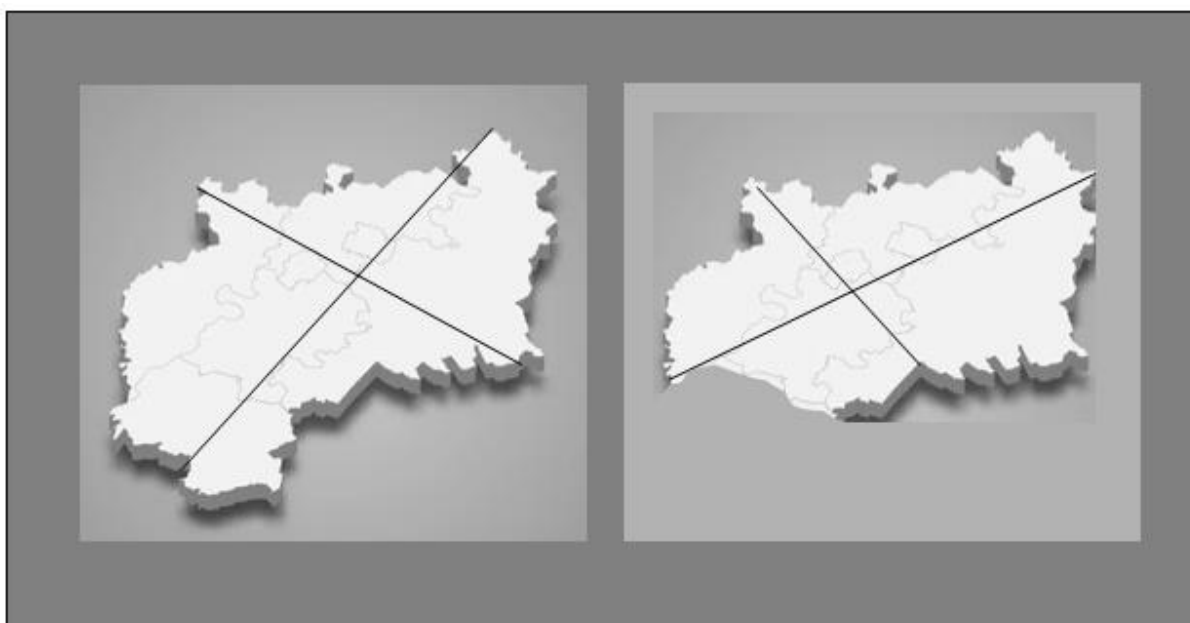
CONDADO DE	ANTES DE LA REFORMA		DESPUÉS DE LA REFORMA	
GLOUCESTERSHIRE	Distancia milímetros	mm * 1,05 = Millas	Distancia milímetros	mm * 1,05 = Millas
Longitud en millas del eje mayor = L1	57	60	49,5	52
Longitud en millas del eje menor = L2	38	40	28,5	30

Antes de la reforma

Después de la reforma

El índice de compacidad = $(Ic) = \frac{L2}{L1}$ $\frac{40}{60} = 0,66$

$\frac{30}{52} = 0,57$



DATOS DE APLICACIÓN PARA EL ÍNDICE DE COLE Y DE GIBBS

(El diámetro del círculo mínimo circunscrito coincide en las dos situaciones del Condado de Gloucestershire con la longitud del eje más largo). Los datos se presentan en millas

CONDADO DE GLOUCESTERSHIRE	Antes de la reforma	Después de la reforma
Área de la superficie (en millas 2) = A	1250	840
Diámetro del círculo mínimo circunscrito = D	60	52
Radio del círculo mínimo circunscrito = Rc	30	26
Círculo mínimo circunscrito = Ax	$\pi R_c^2 =$ $30*30*3,1416 = 2827,44$	$\pi R_c^2 =$ $26*26*3,1416 = 2123,72$

ÍNDICE DE COLE

$$I_c = \frac{A}{A_x} * 100 \quad \frac{1250}{2827,44} * 100 = \mathbf{44,2} \quad \frac{840}{2123,72} * 100 = \mathbf{39,55}$$

El índice de Cole se puede plantear de otra manera

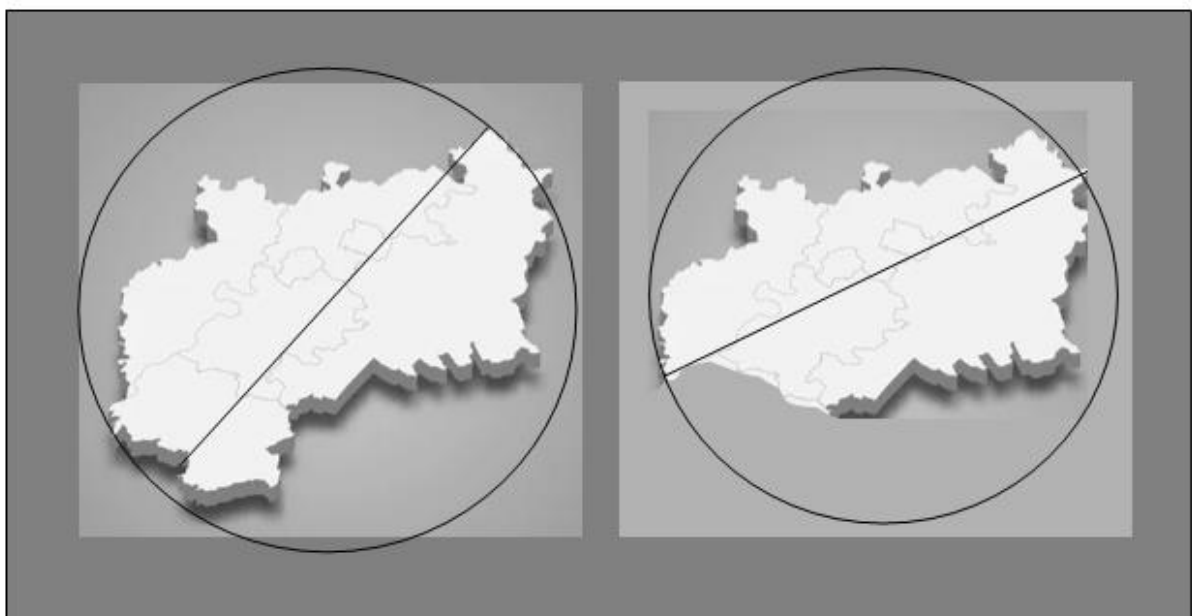
$$I_c = \frac{4A}{\pi D^2} \quad \frac{4*1250}{60*60*3,1416} = \frac{5000}{11309,76} = \mathbf{0,442} \quad \frac{4*840}{52*52*3,1416} = \frac{3360}{8494,88} = \mathbf{0,395}$$

De esta forma el valor del índice se sitúa entre 0 y 1

ÍNDICE DE GIBBS

$$I(G) = \frac{4A}{\pi L^2} \quad \frac{4*1250}{60*60*3,1416} = \frac{5000}{11309,76} = \mathbf{0,442} \quad \frac{4*840}{52*52*3,1416} = \frac{3360}{8494,88} = \mathbf{0,395}$$

El índice de Gibbs y el índice de Cole coinciden, porque coinciden el diámetro del círculo menor circunscrito y la longitud del eje más largo en las dos situaciones estudiadas.



ÍNDICE: I (s) -- ÍNDICE: I (T)

DATOS DE APLICACIÓN

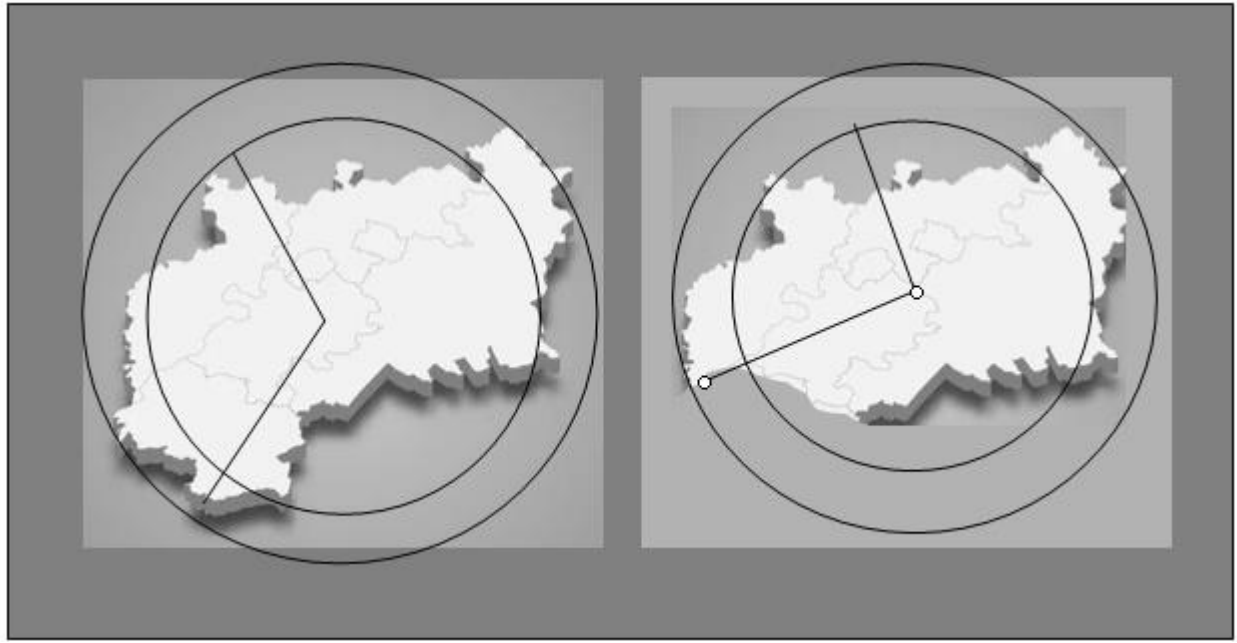
Condado de Gloucesterhire	Antes de la reforma		Después de la reforma	
Radio del círculo de igual área = RA	$(\frac{\sqrt{A}}{\pi}) = \frac{\sqrt{1250}}{3,1416} = \sqrt{397,88} = \mathbf{19,94}$		$(\frac{\sqrt{A}}{\pi}) = \frac{\sqrt{840}}{3,1416} = \sqrt{267,37} = \mathbf{16,44}$	
Radio del círculo mayor inscrito = Ri	Distancia milímetros	mm * 1,05 = Millas	Distancia milímetros	mm * 1,05 = Millas
	11,5	12	11,5	12

Antes de la reforma

$$I(s) = \frac{RA}{Rc} = \frac{19,94}{30} = \mathbf{0,664}$$

Después de la reforma

$$\frac{16,44}{26} = \mathbf{0,632}$$

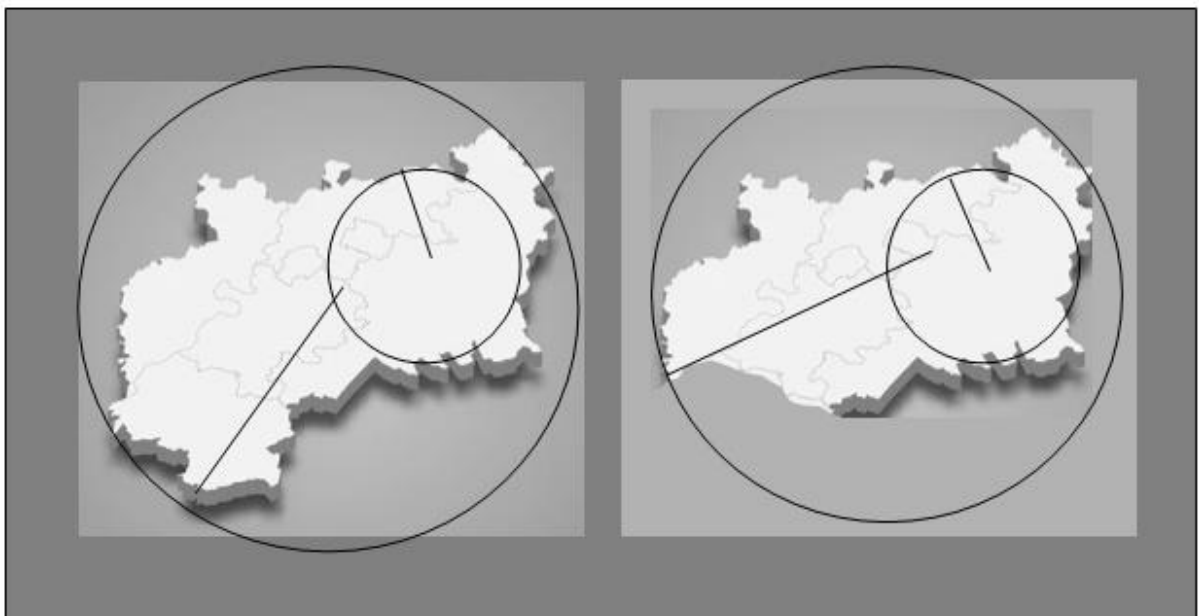


Antes de la reforma

$$I(T) = \frac{Ri}{Rc} = \frac{12}{30} = \mathbf{0,4}$$

Después de la reforma

$$\frac{12}{26} = \mathbf{0,461}$$



EJERCICIO NÚMERO 5

A partir de la matriz de la estructura de los factores, (cuadro 3) aplicado a algunos aspectos geográficos de México sobre una tabla de datos de 18 variables 32 lugares.

a) Calcular el % de la varianza

b) Deducir el sentido de cada factor

a) Aclaración sobre el cálculo del % de la varianza.

(Este ejercicio queda sin realizar, debido a que el cálculo factorial ha sido realizado aplicando una rotación oblicua y por tanto **no se cumple** una de las relaciones fundamentales del análisis factorial, que indica que la suma de los cuadrados de los coeficientes factoriales pertenecientes a cada variable es igual a 1, y el total de la varianza es igual a 1 por el número de variables)

El ejemplo de la tercera variable de la matriz de la estructura de los factores, muestra ese incumplimiento.

Tercera variable	Factor 1	Coefici. al cuadrado	Factor 2	Coefici. al cuadrado	Factor 3	Coefici. al cuadrado	Factor 4	Coefici. al cuadrado	Total Coefic.al cuadrado
Servicio	0,97	0,94	0,77	0,59	0,6	0,36	0,41	0,16	2,05

b) Deducir el sentido de cada factor

Las asociaciones o correlaciones principales de las variables con los factores se indican en el cuadro número 4 del ejercicio, pudiendo deducirse del mismo el siguiente sentido para cada factor.

FACTOR 1		FACTOR 2		FACTOR 3		FACTOR 4	
1 Agricultura	-0,9	12 Instrucción	0,99	9 Población 50-70	0,94	7 Hijos por mujer	-0,9
3 Servicio	0,97	13 Alfabetismo	0,99	8 Migración	0,93	17 comen carne	0,78
4 Profesionales	0,94	3 Servicios	0,77	18 ing.de más 499	0,68		
16 Energía	0,92	18 Ing.r más 499	0,77				
15 Agua Entubad	0,9	4 Profesionales	0,76				
18 Ingr. más 499	0,82	16 Energ. eléctrica	0,74				
2 Industrias	0,77	2 Agricultura	-0,7				
12 Instrucción	0,74	15 Agua entubada	0,71				
13 Alfabetismo	0,74	14 V.más de un C	0,7				
5 Local. más 499	0,69						

FACTOR 1

El primer factor representa el estatus socioprofesional mejor cualificado y pagado que se sitúa en el sector servicios principalmente y en la industria en menor medida y goza de las mejores condiciones de habitabilidad y nivel de instrucción cultural y se localiza en las localidades mayores de 499

habitantes. En contraposición a la población del sector agrícola

Ello se aprecia en las variables con una alta correlación positiva como son las de, Servicios, industria, profesionales, personal con ingresos de más de 499 pesos mensuales, personas con instrucción, personas alfabetas, residencia en ciudades de más de 499 habitantes, y viviendas que disponen de agua entubada y energía eléctrica.

La alta correlación negativa del sector agrícola indica que la población de ese sector carece de las ventajas del estamento anterior y que se sitúa principalmente en poblaciones por debajo de 499 habitantes.

FACTOR 2

El segundo factor representa el nivel de instrucción de la población y la calidad de vida que se corresponde con los profesionales, técnicos, y funcionarios superiores, que trabajan en el sector servicios, con buenos ingresos y residencias bien acondicionadas.

Las correlaciones positivas importantes con alguna variación en los coeficientes son las mismas que en el primer factor, excepto en la variable de industria, (que indica un nivel de instrucción menor en la población que trabaja en este sector), y la variable de población con residencia en poblaciones de más de 499 habitantes, (lo que demuestra que la población del sector servicios también reside en localidades por debajo de ese número de habitantes.) Se añade la variable de viviendas con más de un cuarto con un coeficiente algo mayor que en el primer factor, (que indica un mejor nivel de vida en la población de nivel cultural alto y profesionales que trabajan en el sector servicios).

De nuevo la variable de la población que trabaja en el sector agrícola tiene una correlación negativa con el estamento social anterior, poniéndose de relieve, que es la población menos instruida y con peores condiciones de vida.

FACTOR 3

Este factor representa el crecimiento de la población de los lugares de estudio entre los años 1950-1970, que es debido a la población que no nació en el lugar donde reside, es decir debido los movimientos inmigratorios durante esos años y que tiene una correspondencia media con la población con ingresos de más de 499 pesos mensuales.

FACTOR 4

El cuarto factor contrapone dos tipos de variables con una correlación negativa o inversa entre el promedio de hijos por cada 10 mujeres y el porcentaje de viviendas que se come carne por lo menos una vez por semana, o visto de otra manera existe una relación entre el mayor número de hijos y una menor alimentación de carne.

EJERCICIO NÚMERO 6

A partir de la matriz de datos iniciales relativos a los 18 distritos de Madrid y a 9 variables o indicadores de la diferenciación social de Madrid, cuadro 1, se ha aplicado un análisis factorial, obteniéndose los resultados en la matriz factorial, cuadro 2, los valores de las notaciones factoriales, cuadro 3, y la representación gráfica del primer factor, figura 1.

- a) - Comenta el porcentaje de la varianza para cada uno de los factores.
- b) - El significado y representatividad de cada factor (dos primeros).
- c)- Las regularidades espaciales del primer factor.

ANÁLISIS FACTORIAL

a) Comentario del porcentaje de la varianza en cada factor

La matriz factorial, correspondiente a una matriz de datos de 9 variables, 18 lugares, presenta una acumulación de la varianza del 94,21 % en los dos primeros factores

CÁLCULO DEL PORCENTAJE DE LA VARIANZA EN LOS FACTORES

VARIABLES	FACTOR 1	(Factor 1) ²	FACTOR 2	(Factor 2) ²	% de varianza de cada variable en los dos factores	
% Pobl.0 a 14 años	-0,21921	0,05	-0,94118	0,89	0,05 + 0,89 =	0,93
%Pobl.65 y+ años	0,26616	0,07	0,95408	0,91	0,07 + 0,91 =	0,98
%Pobl. Analfabeta	-0,89336	0,80	-0,25221	0,06	0,80 + 0,06 =	0,86
%Pobl. Ed. Superior.	0,9154	0,84	0,34362	0,12	0,84 + 0,12 =	0,96
%Pobl. Sector Indust.	-0,7877	0,62	-0,52927	0,28	0,62 + 0,28 =	0,90
%Pobl. Sector Servic.	0,78029	0,61	0,57666	0,33	0,61 + 0,33 =	0,94
% Técnicos	0,94107	0,89	0,30561	0,09	0,89 + 0,09 =	0,98
% Directivos	0,96337	0,93	0,05186	0,00	0,93 + 0,0 =	0,93
% Trabaj. Manuales	-0,89121	0,79	-0,4471	0,20	0,79 + 0,20 =	0,99
Varianza total = 9						
Varianza explicada =		5,59		2,89		
Va. expl. acumulada =		5,59		8,48	TOTAL =	8,48
% de varianza =	$\frac{5,59 \cdot 100}{9} =$	62,14%	$\frac{2,89 \cdot 100}{9} =$	32,07%		
% acumulado =		62,14%		94,21%		94,21%

La varianza máxima de cada variable no puede ser superior a 1, siendo la varianza total igual a 9.

El valor de la varianza de cada variable en los factores, se halla elevando al cuadrado el coeficiente factorial respectivo.

La suma de los cuadrados de los coeficientes factoriales pertenecientes a cada variable, representa el valor de la varianza de ese factor.

El primer factor supone el 62,14% de la varianza que integra las siete variables que explican la correlación socioprofesional y cultural, contraponiendo a dos grupos sociales.

El segundo factor acumula el 32,07% de la varianza e integra las dos variables que explican la correlación de los dos grupos de edad 0-14 años y 65 y más años con las variables de población industrial y población del sector servicios.

b) - Comentario del significado y representatividad de cada factor

PRIMER FACTOR

El primer factor representa el estatus social, profesional y cultural, cuestión que se aprecia en la polarización de dos tipos de variables entre sí.

De una parte, la población de nivel educativo superior, (variable N° 4) que tiene un alto grado de correlación positiva con los técnicos y con los directivos, (variables N° 7, 8,) que a su vez tienen un grado de correlación importante con la variable de personal empleado en servicios, (variable N° 6).

De otra parte en correlación negativa, la población analfabeta, (variable N° 3) que tiene un alto grado de correlación con la variable (N° 9) de trabajadores manuales y la empleada en el sector industrial, (variable N° 5).

Las otras dos variables (N° 1, 2,) apenas si tienen significación en el primer factor.

SEGUNDO FACTOR

El segundo factor refleja la correlación de los grupos de edad en las distintas variables.

La población de edad de 65 años y más, (variable N° 2) tiene un grado de correlación positiva media con la variable de la población empleada en los servicios y en menor medida con la población de titulación superior y de técnicos, (variables 4, 7) ello tal vez debido a que es un sector que permite realizar trabajos después de los 65 años,

La variable de población de edad comprendida entre 0 y 14 años, (N° 1) tiene una correlación de tipo medio con la población empleada en la industria, y los trabajadores manuales, (variables N° 5, 9) que representan con toda probabilidad a trabajadores jóvenes con hijos menores de 14 años. Estas variables tienen lógicamente correlación negativa o inversa con la variable de población de 65 y más años.

El resto de las variables (N° 3, 8) tienen escasa significación en el segundo factor

c) - Las regularidades espaciales del primer factor

Establecidos los factores que explican la mayor parte de la varianza, a partir de los datos de la composición de la población en cada lugar, (en el presente caso los distritos de Madrid) se obtiene los valores de los factores en cada uno de estos lugares de referencia dando lugar a la matriz de las notaciones factoriales.

Desde el punto de vista del significado fundamental del primer factor , teniendo en cuenta la matriz de notaciones factoriales, el mapa de la zona de estudio y los valores de clasificación, se pueden apreciar las siguientes cuestiones.

Existe una clara diferenciación entre el centro y la periferia, pudiendo señalarse cuatro núcleos.

1º La población de mayor nivel educativo y estatus socioprofesional se sitúa en los distritos del Retiro, Salamanca, Chamartín y Chamberí, con notaciones superiores a 1 y que ocupan el centro de la zona de estudio.

2º Los distritos de Centro, Arganzuela, Tetuán y Moncloa con notaciones entre 0 y 1, tienen un incidencia del estatus social anterior algo menor, pero que sigue siendo importante y se extiende geográficamente a partir del núcleo anterior. Ello refleja un cierto grado de expansión hacia estos distritos de las funciones que demandan una mano de obra cualificada.

3º La zona geográfica más amplia que compone la mayor parte de la periferia de la zona de estudio con notaciones de - 1 a 0, refleja una incidencia de trabajadores manuales industriales con un nivel cultural bajo y se corresponde con los distritos de Fuencarral, Latina, Carabanchel, Moratalaz, Ciudad Lineal, y Hortaleza.

4º Los barrios de Villaverde, Mediodía, Vallecas y San Blas con notaciones de ≤ -1 , componen otra zona periférica, siendo donde se asienta la población de menor nivel cultural y el mayor porcentaje de oficios manuales industriales respecto de la población ocupada.

En una valoración global se puede considerar que el área metropolitana de Madrid es un espacio desequilibrado, con gran concentración del empleo del sector terciario en el centro con profesiones cualificadas y alto nivel cultural en oposición a una periferia con un bajo nivel cultural y mano de obra dedicada a trabajos manuales en la industria.

Este fenómeno es común a las grandes urbes, donde debido al desarrollo económico se produce una concentración del sector terciario en el centro de la ciudad, a partir de la zona administrativa y comercial más antigua. Relacionado con el tipo de cualificación profesional de técnicos y directivos que demanda este sector se explica la localización laboral de esta población en el centro y zonas cercanas.

La localización de la industria se realiza en los municipios periféricos que reúnen mejores condiciones de espacios libres para su ubicación. Este sector económico demanda principalmente trabajadores manuales sin necesidad de conocimientos especiales, lo que implica a su vez, debido a la época en que se realiza la industrialización de esas zonas periféricas de Madrid y la procedencia rural de muchos de los trabajadores, un nivel de analfabetismo importante.