



Universidad Autónoma de Yucatán
Facultad de Matemáticas

Conjunto de países con características socio-económicas similares

ANÁLISIS MULTIVARIADO
Reporte Estadístico

Por:

- **Álvarez Samantha**
- **Asencio Alejandra**
- **Sobrino Samantha**

Licenciatura en Actuaría





UADY

UNIVERSIDAD
AUTÓNOMA
DE YUCATÁN

"Luz, Ciencia y Verdad"

FACULTAD DE MATEMÁTICAS

Análisis Multivariado

Análisis de Conglomerados

Reporte Estadístico

Por:

- Álvarez Herrera Samantha
- Asencio González Alejandra
- Sobrino Bermejo Samantha

Profesora: Rocío Acosta Pech

Licenciatura en Actuaría - Especialización en Estadística.

TABLA DE CONTENIDOS

	Página
ABSTRACTO.....	2
INTRODUCCIÓN.....	2
ANTECEDENTES.....	2
MÉTODOLOGÍA.....	2
ANÁLISIS DESCRIPTIVO.....	2
RECOLECCIÓN DE LOS DATOS.....	4
OBJETIVOS.....	5
ANÁLISIS POR CONGLOMERADOS.....	5
RESULTADOS.....	5
CONCLUSIONES.....	5
REFERENCIAS.....	6
ANEXOS	
A. TABLA DE DATOS DEL ESTUDIO.....	i
B. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES DEL ESTUDIO.....	ii
C. MATRIZ DE DISTANCIAS.....	iv
D. ANÁLISIS DE CONGLOMERADOS	vi

Abstracto

En este trabajo se obtienen dos grupos de países con diferencias significativas entre ellos en términos de elementos socioeconómicos. De tal forma que se puedan tomar distintas medidas para mejorar el nivel de vida de las personas de los países considerados. Las variables consideradas son Tasa de mortalidad infantil por cada 1,000 nacidos vivos, porcentaje de mujeres en la población activa, PNB en 1995, Producción de electricidad, Líneas telefónicas por cada 1,000 habitantes, Consumo de agua per cápita, Proporción de superficie del país cubierta por bosques, Proporción de deforestación anual, Consumo de electricidad per cápita y Emisión de CO2 per cápita.

Cabe mencionar que el objetivo inicial del estudio es realizar un análisis de conglomerados de los países para determinar que grupos se forman y así poderlos analizar por separado en estudios posteriores.

Introducción

El mundo está experimentando un cambio demográfico sin precedentes. Desde comienzos del siglo XX hasta la actualidad, la humanidad ha pasado de tener 2 000 millones de habitantes a más de 7 000. Se espera que esa cifra alcance los 10 000 millones de personas a finales del siglo XXI. Otros cambios en la población son la mortalidad, la migración, la urbanización y el envejecimiento. Por lo tanto, es importante estudiar las principales características y tendencias de la población de los países.

Los países que estudiaremos en este proyecto serán divididos en grupos cuyas condiciones son similares y así cada país puede utilizar la información generada para guiar las acciones de los organismos públicos. Para una mejor planificación o, más generalmente, para la toma de decisiones económicas, sociales y culturales, al igual que así los gobiernos podrán elaborar políticas y programas nacionales que mejoren la calidad de vida de las personas.

Las variables que tomamos en cuenta en el estudio son importantes debido a que son esenciales para medir la calidad de vida de todo ser humano.

Antecedentes

Se han hecho investigaciones para analizar las condiciones de vida de las personas según los países a los que pertenecen.

La Comisión Económica para América Latina y el Caribe realizó un estudio donde habla sobre la importancia de promover acciones que dan lugar a lo que suele llamarse "incorporación de las variables demográficas en las políticas, planes o acciones económico-sociales" toman a esa dinámica demográfica como un hecho frente al cual se reacciona buscando los recursos para atender las necesidades sociales y de empleo.

Metodología.

Para este estudio se cuenta con 11 variables, cada una con 22 observaciones. Se realizó un estudio por conglomerados de los 22 países en cuestión con el fin de determinar si es posible formar grupos entre estos que dependan de características sociales y económicas medidas por los 11 indicadores o variables contempladas.

Debido a que los datos se encuentran medidos en diferentes escalas y unidades se optó por estandarizar los datos.

Análisis descriptivo

Procederemos a puntualizar las características de todas las variables involucradas en este estudio, así como la relación existente con el bienestar o estado socioeconómico del país. (*Para mayor detalle, consultar Anexo B*):

- **Tasa anual de crecimiento de la población:** Se refiere a la tasa o índice que expresa el ritmo al que la población aumenta o disminuye durante un período dado, debido al efecto de la natalidad, la mortalidad y las migraciones en un determinado territorio durante un período definido. Expresada como porcentaje de la población al inicio de cada período o año.

Los datos recolectados de esta variable cuentan con una media de 1.47727% y la desviación estándar es igual a 0.741897%. Además, su curtosis estandarizada se encuentra dentro del rango esperado para los datos de una distribución Normal. Lo cual fue corroborado al aplicar una prueba de bondad de ajuste ($Valor - p = 0.75 > 0.05 = \alpha$).

- **Tasa de mortalidad infantil por cada mil nacidos vivos:**

Indicador demográfico que señala el número de defunciones de niños en una población de cada mil nacimientos vivos registrados, durante el primer año de su vida. La consideración del primer año de vida se debe a que el primer año de vida es el más crítico en la supervivencia del ser humano. Indicador relacionado directamente con los niveles de pobreza y de calidad de la sanidad gratuita (a mayor pobreza o menor calidad sanitaria, mayor índice de mortalidad infantil) y constituye el objeto de uno de los 8 Objetivos del Milenio de las Naciones Unidas.

Su media y su desviación estándar son de 25.95 y 21.122, respectivamente. La curtosis y el sesgo estandarizados están dentro del intervalo $(-2, 2)$, sin embargo, los datos se ajustan a una distribución normal ($Valor - p = 0.0078 < 0.05 = \alpha$).

- **Porcentaje de mujeres en la población activa:**

Número de mujeres que forman parte de la población activa entre el total de esta población. Se define como población activa de un país a la compuesta por todos los habitantes en edad laboral que trabajan en un empleo remunerado (población ocupada) o bien se halla en búsqueda de empleo (población en paro). Por tanto, la población activa se divide en dos grupos, los empleados y los desempleados.

Esta variable tiene una media de 37 y desviación estándar igual a 6.45. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2 , y se ajustan a una distribución normal ($Valor - p = 0.46 > 0.05 = \alpha$).

- **PNB en 1995 (en MDD):**

Producto Nacional Bruto, es el conjunto de los bienes y servicios generados por los factores productivos de un país durante un espacio de tiempo, en esta casi en el año 1995, en el propio país o fuera de él, deduciéndole la parte que se debe a los servicios prestados por factores productivos extranjeros. Está expresada en millones de dólares americanos.

Su media y su desviación estándar son de 343995 MDD y 374128 MDD, respectivamente. Su sesgo y curtosis estandarizados se encuentran fuera del intervalo $(-2, 2)$ de una distribución normal y según la prueba de bondad de ajuste estos datos no se ajustan a una normal ($Valor - p = 0.0008 < 0.05 = \alpha$).

- **Producción de electricidad (en millones kW/h):**

Esta variable hace referencia al total de energía eléctrica producida en el país durante un año, dependiente de variables como los tipos de industrias existentes en la zona, climatologías extremas, tipo de electrodomésticos que se utilizan más frecuentemente, tipo de calentador de agua que haya instalado en los hogares. Es medida en millones de kW/hr.

Esta variable tiene una media de 191671 millones de kW/hr y una desviación estándar igual a 227477 millones de kW/hr. Además, el sesgo y la curtosis no se encuentran dentro del intervalo de -2 a 2 , mientras que la prueba de Shapiro Wilk determinó que los datos no se ajustan a una normal. ($Valor - p = 0.00016 < 0.05 = \alpha$).

- **Líneas telefónicas por cada 1000 habitantes:**

Una “línea telefónica” es un circuito eléctrico de un sistema de telecomunicaciones por teléfono. Se considera, el disponer de línea telefónica, un equipamiento del edificio. Cuando se habla de viviendas con línea telefónica se debe entender que están ubicadas en un edificio con instalación de cableado o tendido telefónico. Esta variable hace referencia a la proporción de líneas telefónicas disponibles por cada 1000 personas en el país, mide el nivel de la comunicación.

Tiene una media de 10.1045 y una desviación estándar de 1.47977. Su sesgo y curtosis se encuentran dentro del intervalo de -2 a 2 , por lo que los datos pueden haber provenido de una distribución normal, sin embargo, esto se refutó al realizar una prueba de Shapiro Wilk ($Valor - p = 0.00047 > 0.05 = \alpha$).

- **Consumo de agua per cápita:**

Se entiende por consumo doméstico de agua por habitante a la cantidad de agua que dispone una persona para sus necesidades diarias de consumo, aseo, limpieza, riego, etc. y se mide en litros por habitante y día (l/hab-día). Es un valor muy representativo de las necesidades y/o consumo real de agua dentro de una comunidad o población y, por consiguiente, refleja también de manera indirecta su nivel de desarrollo económico y social.

Este indicador social se obtiene a partir del suministro medido por contadores, estudios locales, encuestas o la cantidad total suministrada a una comunidad dividida por el número de habitantes.

Esta variable tiene una media de 592.32 y una desviación estándar igual a 439.41. Además, el sesgo estandarizado es mayor al límite 2, mientras que la curtosis sí se encuentra dentro del intervalo de -2 a 2, la prueba de Shapiro Wilk arrojó resultados significativos, por lo que los datos no se ajustan a una distribución normal. ($Valor - p = 0.0315 < 0.05 = \alpha$).

- **Proporción de superficie del país cubierta por bosques:** Este indicador señala el porcentaje de la superficie cubierta por bosques con relación a la superficie total del territorio del país. Los bosques se definen como las tierras que se extienden por más de 0,5 hectáreas, dotadas de árboles de una altura superior a los 5m y una cubierta de copas superior al 10%, o de árboles capaces de alcanzar una altura mínima de 5m in situ. La definición comprende los bosques naturales y las plantaciones forestales, así como los bosques utilizados fundamentalmente con fines de producción, conservación o usos múltiples. No incluye la tierra sometida a un uso predominantemente agrícola o urbano. La superficie total del territorio del país corresponde a la superficie total de las tierras del país, con exclusión de la superficie ocupada por masas de agua interiores (principales ríos y lagos). Medida en porcentaje.

Esta variable tiene media de 33.32% y 21.24% como desviación estándar. Por otro lado, el sesgo y la curtosis sí se encuentran dentro del intervalo de -2 a 2, además la prueba de Shapiro Wilk no resultó significativa, por lo que los datos se ajustan a una distribución normal. ($Valor - p = 0.13 > 0.05 = \alpha$).

- **Proporción de deforestación anual:** Hace referencia a la superficie de bosque, que en términos promedio, se ha perdido anualmente entre dos instantes de tiempo determinados. La unidad de medida es hectáreas por año (ha/año).

Esta variable tiene una media de 0.53 ha/año y una desviación estándar de 1.31 ha/año.

Por otro lado, el sesgo y la curtosis no se encuentran dentro del intervalo de -2 a 2, además la prueba de Shapiro Wilk resultó significativa, por lo que los datos se ajustan a una distribución normal. ($Valor - p = 0.00072 < 0.05 = \alpha$).

- **Consumo de energía per cápita:** Tradicionalmente este indicador es utilizado como una medida de progreso económico, al relacionar el consumo de energía con el grado de industrialización de un país y con la calidad de vida de sus habitantes. La relación entre la cantidad de energía consumida por habitante del país, en un momento determinado, informa acerca de cómo la energía es utilizada, directa e indirectamente, por la población.

Tiene media de 2205.27 y desviación estándar de 2095.4. Además, el sesgo y la curtosis no se encuentran dentro del intervalo de -2 a 2, mientras que la prueba de Shapiro Wilk determinó que los datos no se ajustan a una distribución normal. ($Valor - p = 0.0062 < 0.05 = \alpha$).

- **Emisión de CO2 per cápita:** Las emisiones de dióxido de carbono son las que provienen de la quema de combustibles fósiles y de la fabricación del cemento. Incluyen el dióxido de carbono producido durante el consumo de combustibles sólidos, líquidos, gaseosos y de la quema de gas. Se mide en toneladas métricas per cápita.

Esta variable posee una media de 4.85 y una desviación estándar de 4.17. Por otro lado, el sesgo estandarizado no se encuentra dentro del intervalo referencia, mientras que la curtosis estandarizada sí. Una prueba de Shapiro Wilk determinó que los datos no se ajustan a una distribución normal. ($Valor - p = 0.00599 < 0.05 = \alpha$).

Recolección de los datos

Los datos del proyecto provienen de “100 Problemas Resueltos de Estadística Multivariante (Implementados En Matlab)” por Baillo Moreno, Amparo y Grané Chávez.

Las autoras son provenientes de la Facultad de Ciencias de la Universidad Autónoma de Madrid y la Facultad de Ciencias Jurídicas y Sociales de la Universidad Carlos III de Madrid.

La información contenida en este libro es fruto de la experiencia docente de las autoras y fue recuperada de distintas plataformas de información oficiales de organizaciones nacionales e internacionales.

Todos los datos fueron validados por las autoridades de las Instituciones correspondiente. Lo anterior nos da la certeza de la veracidad y aleatoriedad de la muestra.

Objetivos del estudio

General: Realizar un análisis exploratorio de las variables o indicadores socioeconómicos y su influencia sobre 22 países estudiados con el fin de determinar si es posible la formación de grupos de países con características similares que permitan una aplicación de técnicas y programas de mejora basados en dichas características.

Específicos: Formar grupos de países homogéneos que puedan compartir y complementar planes de mejora común, estudios acertados y demás técnicas cuyo objetivo sea el desarrollo económico, social y sustentable.

Análisis por Conglomerados

El estudio desea formar conglomerados con 27 países, a los que se midieron once variables.

Se hará uso del método de Ward, en conjunto con la distancia euclídea al cuadrado, para ir formando los conglomerados, es decir, en cada etapa del proceso se unirán aquella pareja de clusters cuya unión proporcione el menor incremento en la suma total de errores, E .

El incremento potencial en E que se producirá con la unión de los clusters C_t y C_r es:

$$\Delta E_{rt} = \frac{n_r n_t}{n_r + n_t} \sum_{j=1}^n (m_j^r - m_j^t)^2$$

Donde:

n_k es el número de individuos que posee el cluster k .

m_i^k es el componente correspondiente a la i -ésima variable del cluster k .

Para determinar el número de clústers adecuado se analizará el gráfico de distancia de aglomeración, deteniendo el proceso cuando la distancia entre los conglomerados sea grande.

Resultados

Tras analizar el gráfico de distancia de aglomeración, obtenido con el método de Ward, se decidió trabajar con dos clusters.

El primer grupo lo forman China, India, Brasil, Suráfrica, México, Venezuela, Argentina, Colombia, Chile, Perú, Cuba, Ecuador, Guatemala y Haití.

El segundo conglomerado está conformado por Canadá, Francia, Reino Unido, Australia, Corea del Sur, España, Suecia y Suiza.

Revisando los valores medios de las variables, vemos que, el primer grupo presenta valores significativamente menores en aspectos como el PNB, el número de líneas telefónicas, el consumo de agua y de energía per cápita. También destacan por su alto índice de mortalidad infantil y proporción de deforestación anual. Por otro lado, el segundo grupo cuenta con tasas bajas de crecimiento de la población y mortalidad infantil, y valores altos en emisión de CO₂, consumo de energía por habitante y un mayor porcentaje de participación de las mujeres en la población activa.

Conclusiones

El primer grupo está formado por países con altas tasas de crecimiento de la población y mortalidad infantil y una menor participación de las mujeres en la población activa, a comparación del otro grupo. El Producto Nacional Bruto, el número de líneas telefónicas por cada 1000 habitantes y el consumo de energía per cápita, de estos países son notablemente menores que los valores obtenidos de estas variables de los países pertenecientes al otro conglomerado, y se encuentran incluso debajo de la media.

Dado que las variables analizadas nos dan una indicación de la calidad de vida de los habitantes pertenecientes a los países estudiados, podemos nombrar al primer grupo como “países con baja calidad de vida de sus habitantes” y al segundo como “países con baja calidad de vida de sus habitantes”.

Referencias

- Johnson, R. y Wichern, D. (2007) *Applied Multivariate Statistical Analysis*, sexta edición, Pearson, Nueva Jersey.
- Ambientum El consumo de agua en porcentajes. *Recuperado de:*
https://www.ambientum.com/enciclopedia_me dioambiental/aguas/el-consumo-de-agua-en-porcentajes.asp
- Celade: Tasa de mortalidad Infantil. *Recuperado de:*
https://celade.cepal.org/redatam/PRYESP/SISPPI/Webhelp/tasa_de_mortalidad_infantil.htm
- DANE. Hoja metodológica de indicadores cuenta satélite ambiental, dirección de síntesis y cuentas nacionales. Consumo de energía per cápita. *Recuperado de:*
https://www.dane.gov.co/files/investigaciones/pib/ambientales/cuentas_ambientales/indicadores/cuenta-ambiental-y-economica-flujos-de-energia/consumo-energia-percapita/hoja-metodologica-consumo-de-energia-percapita.pdf
- Economipedia: Población activa. *Recuperado de:*
<https://economipedia.com/definiciones/poblacion-activa.html>
- Economipedia: Producto nacional bruto (PNB). *Recuperado de:*
<https://economipedia.com/definiciones/producto-nacional-bruto-pnb.html>
- Endesa: Generación. *Recuperado de:*
<https://www.endesa.com/es/sobre-endesa/a201610-generacion-produccion-energia-electrica.html>
- Estadísticas ambiente: Tasa media anual de crecimiento poblacional. *Recuperado de:*
http://estadisticas.ambiente.gob.ar/archivos/web/Indicadores/file/multisitio/pdf/13_%20Tasa%20media%20anual%20de%20crecimiento%20poblacional.pdf
- Eustat. Instituto Vasco de Estadística. Línea telefónica. *Recuperado de:*
http://www.eustat.eus/documentos/opt_0/tema_443/elem_11381/definicion.html
- FAO: 4 Estimados de cobertura forestal y tasas de deforestación. *Recuperado de:*
<http://www.fao.org/3/ad680s/ad680s05.htm>
- Indexmundi. Emisiones de CO2. *Recuperado de:*
<https://www.indexmundi.com/es/datos/indicadores/EN.ATM.CO2E.PC>
- Sinia: Sistema de protección ambiental. *Recuperado de:*
<https://sinia.minam.gob.pe/indicadores/proporcion-superficie-cubierta-bosques>

ANEXO A: TABLA DE DATOS DEL ESTUDIO.

Tabla 1

Indicadores económicos y sociales sobre países del mundo

País	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
25. China	1.3	34	45	744890	928083	34	461	13	0.7	664	2.3
16. Canadá	1.3	6	45	573695	554227	590	1602	49	-1.1	7854	14.4
37. Francia	0.5	6	44	1451051	476200	558	665	25	-0.1	4042	6.3
46. India	1.9	68	32	319660	386500	13	612	17	0.6	248	0.9
75. Reino Unido	0.3	6	43	1094734	325383	502	205	10	-1.1	3732	9.8
13. Brasil	1.6	44	35	579787	260682	75	246	66	0.6	718	1.4
85. Suráfrica	2.3	50	37	130918	189316	95	359	4	-0.8	2146	7.5
6. Australia	1.4	6	43	337909	167155	510	933	19	0	5341	15.3
20. Corea del Sur	0.9	10	40	435137	164993	415	632	66	0.1	2982	6.6
33. España	0.2	7	36	532347	161654	385	781	51	0	2458	5.7
60. México	2.1	33	31	304596	147926	96	899	25	1.3	1561	3.8
83. Suecia	0.6	4	48	209720	142895	681	341	68	0	5723	6.6
92. Venezuela	2.4	23	33	65382	73116	111	382	52	1.2	2186	5.7
5. Argentina	1.3	22	31	278431	65962	160	1043	22	0.1	1504	3.5
84. Suiza	0.8	6	40	286014	65724	613	173	30	-0.6	3629	6.4
17. Colombia	1.8	26	37	70263	43354	100	174	52	0.7	622	1.8
24. Chile	1.6	12	32	59151	25276	132	1626	12	-0.1	1012	2.6
72. Perú	2.1	47	29	55019	15563	47	300	53	0.4	367	1
23. Cuba	0.9	9	38	7150	10982	32	870	16	1	923	2.6
28. Ecuador	2.3	36	26	15997	8256	61	581	43	1.8	565	1.8
41. Guatemala	2.9	44	26	14255	3161	27	139	39	1.8	210	0.6
42. Haití	2	72	43	1777	362	8	7	1	5.1	29	0.1

Donde:

X_1 : Tasa anual de crecimiento de la población.

X_2 : Tasa de mortalidad infantil por cada 1000 nacidos vivos.

X_3 : Porcentaje de mujeres en la población activa.

X_4 : PNB en 1995 (en millones de dólares).

X_5 : Producción de electricidad (en millones kW/h).

X_6 : Líneas telefónicas por cada 1000 habitantes.

X_7 : Consumo de agua per cápita.

X_8 : Proporción de la superficie del país cubierta por bosques.

X_9 : Proporción de deforestación anual.

X_{10} : Consumo de energía per cápita.

X_{11} : Emisión de CO2 per cápita.

ANEXO B: ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

Análisis individual de las variables.

El estudio cuenta con once variables de interés, con 22 observaciones. A cada una se le determinó su media y su desviación estándar y se evaluó si su sesgo y su curtosis estandarizados se encuentran dentro del rango esperado para los datos de una distribución Normal. También se aplicó la prueba de Shapiro-Wilk para normalidad con un nivel de significancia de $\alpha = 5\%$. El no rechazar la hipótesis nula indicaría que la muestra se ajusta a la distribución Normal.

Tabla 2. Resumen estadístico y prueba de normalidad

Variable	Recuento	Media	Desviación Estándar	Mínimo	Máximo	Rango	Sesgo Estandarizado	Curtosis Estandarizada	Prueba de normalidad (SW)	
									Estadístico W	P-valor
Tasa anual de crecimiento de la población	22	1.47727	0.741897	0.2	2.9	2.7	-0.127396	-0.792493	0.972723	0.759412
Tasa de mortalidad infantil por cada mil nacidos vivos	22	25.9545	21.122	4	72	68	1.48966	-0.368821	0.874026	0.00787575
Porcentaje de mujeres en la población activa	22	37	6.45128	26	48	22	-0.197244	-0.992891	0.958533	0.455633
PNB en 1995 (en MDD)	22	343995	374128	1777	1.45E+06	1.45E+06	3.11348	2.64805	0.825275	0.000864902
Producción de electricidad (en millones kW/h)	22	191671	227477	362	928083	927721	3.70796	4.09033	0.785428	0.000162794
Líneas telefónicas por cada 1000 habitantes	22	238.409	236.983	8	681	673	1.37686	-1.19467	0.811167	0.000472479
Consumo de agua per cápita	22	592.318	439.408	7	1626	1619	2.02985	0.780735	0.902805	0.0315576
Proporción de superficie del país cubierta por bosques	22	33.3182	21.2387	1	68	67	0.396247	-1.23842	0.931679	0.13118
Proporción de deforestación anual	22	0.527273	1.3061	-1.1	5.1	6.2	3.92287	6.31384	0.821187	0.000724791
Consumo de energía per cápita	22	2205.27	2095.4	29	7854	7825	2.35499	1.02121	0.870012	0.0061975
Emisión de CO2 per cápita	22	4.85	4.1697	0.1	15.3	15.2	2.38986	1.21228	0.858229	0.00599721

En el resumen estadístico individual podemos observar la media y la desviación estándar correspondiente a cada una de las once variables del estudio. Además, podemos ver el sesgo y la curtosis estandarizados, las cuales pueden usarse para determinar si la muestra proviene de una distribución normal. Valores de estos estadísticos fuera del rango de -2 a +2 indican desviaciones significativas de la normalidad, en este caso, las variables cuyos valores de sesgo estandarizado y de curtosis estandarizada se presentan en color rojo corresponden a valores fuera del rango esperado.

Efectuando una prueba de Shapiro-Wilk vemos que los datos de las variables con valores-p en rojo no se ajustan a la distribución Normal con un nivel de confianza del 95%.

Mediante el paquete estadístico podemos obtener también la matriz de correlaciones de las variables:

Tabla 3. Correlaciones

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X1		0.7116	-0.6547	-0.629	-0.3004	-0.7181	-0.1718	-0.0612	0.4901	-0.5246	-0.4143
		0.0002	0.0009	0.0017	0.1743	0.0002	0.4445	0.7868	0.0206	0.0122	0.0552
X2	0.7116		-0.4214	-0.372	-0.0699	-0.7606	-0.4239	-0.2403	0.6252	-0.6918	-0.6313
	0.0002		0.0508	0.0883	0.7574	<0.01	0.0493	0.2814	0.0019	0.0004	0.0016
X3	-0.6547	-0.4214		0.5065	0.5145	0.6554	-0.0151	-0.0838	-0.2266	0.6551	0.5596
	0.0009	0.0508		0.0162	0.0143	0.0009	0.9469	0.711	0.3105	0.0009	0.0068
X4	-0.629	-0.372	0.5065		0.6951	0.5129	0.0617	-0.0719	-0.4389	0.4171	0.3788
	0.0017	0.0883	0.0162		0.0003	0.0146	0.7849	0.7505	0.041	0.0534	0.0822
X5	-0.3004	-0.0699	0.5145	0.6951		0.1949	0.1498	-0.1438	-0.3239	0.2898	0.2677
	0.1743	0.7574	0.0143	0.0003		0.3848	0.5058	0.523	0.1414	0.1908	0.2285
X6	-0.7181	-0.7606	0.6554	0.5129	0.1949		0.1713	0.2431	-0.5796	0.9039	0.7658
	0.0002	<0.01	0.0009	0.0146	0.3848		0.446	0.2757	0.0047	<0.01	<0.01
X7	-0.1718	-0.4239	-0.0151	0.0617	0.1498	0.1713		-0.1082	-0.3653	0.3569	0.3755
	0.4445	0.0493	0.9469	0.7849	0.5058	0.446		0.6318	0.0945	0.103	0.0851
X8	-0.0612	-0.2403	-0.0838	-0.0719	-0.1438	0.2431	-0.1082		-0.15	0.2079	0.0108
	0.7868	0.2814	0.711	0.7505	0.523	0.2757	0.6318		0.5052	0.3533	0.9618
X9	0.4901	0.6252	-0.2266	-0.4389	-0.3239	-0.5796	-0.3653	-0.15		-0.5834	-0.601
	0.0206	0.0019	0.3105	0.041	0.1414	0.0047	0.0945	0.5052		0.0044	0.0031
X10	-0.5246	-0.6918	0.6551	0.4171	0.2898	0.9039	0.3569	0.2079	-0.5834		0.907
	0.0122	0.0004	0.0009	0.0534	0.1908	<0.01	0.103	0.3533	0.0044		<0.01
X11	-0.4143	-0.6313	0.5596	0.3788	0.2677	0.7658	0.3755	0.0108	-0.601	0.907	
	0.0552	0.0016	0.0068	0.0822	0.2285	<0.01	0.0851	0.9618	0.0031	<0.01	

Correlación

Valor-P

En esta tabla se pueden observar las correlaciones de Pearson entre cada par de variables. El rango de estos coeficientes de correlación va de -1 a +1, y miden la fuerza de la relación lineal entre las variables. El segundo número en cada bloque de la tabla es un valor-P que prueba la significancia estadística de las correlaciones estimadas. Los pares de variables con valores-P en rojo presentan correlaciones *significativamente diferentes de cero* (es decir, no independencia entre las variables), con un nivel de confianza del 95%.

ANEXO C: Matriz de distancias

	CHINA	CANADÁ	FRANCIA	INDIA	REINO UNIDO	BRASIL	SURÁFRICA	AUSTRALIA	COREA DEL SUR	ESPAÑA	MÉXICO
CHINA	0	41.8822	19.771	14.5854	23.1021	18.1427	20.8429	34.5019	25.5828	25.0636	20.9266
CANADÁ	41.8822	0	20.3156	52.9622	23.5689	45.367	38.7339	10.0282	19.8008	23.0796	38.2541
FRANCIA	19.771	20.3156	0	35.6256	4.40655	26.6765	31.913	17.7356	14.337	12.3447	28.813
INDIA	14.5854	52.9622	35.6256	0	34.7774	8.60984	7.92423	35.8958	23.8582	23.0989	5.79878
REINO UNIDO	23.1021	23.5689	4.40655	34.7774	0	27.849	23.5483	12.7387	14.104	11.7099	28.4983
BRASIL	18.1427	45.367	26.6765	8.60984	27.849	0	14.9534	32.3363	10.1077	12.5156	8.61988
SURÁFRICA	20.8429	38.7339	31.913	7.92423	23.5483	14.9534	0	18.4629	19.4478	21.2473	7.77248
AUSTRALIA	34.5019	10.0282	17.7356	35.8958	12.7387	32.3363	18.4629	0	11.9277	13.9282	20.9884
COREA DEL SUR	25.5828	19.8008	14.337	23.8582	14.104	10.1077	19.4478	11.9277	0	2.10738	13.5383
ESPAÑA	25.0636	23.0796	12.3447	23.0989	11.7099	12.5156	21.2473	13.9282	2.10738	0	13.4883
MÉXICO	20.9266	38.2541	28.813	5.79878	28.4983	8.61988	7.77248	20.9884	13.5383	13.4883	0
SUECIA	38.4985	19.7578	19.1338	42.6343	17.3375	24.7764	31.5192	13.9463	5.57875	10.1992	30.1738
VENEZUELA	28.2032	39.2106	34.356	12.9293	31.2697	7.13641	9.97477	20.8021	10.0879	14.3174	4.30767
ARGENTINA	23.6213	33.9013	24.3952	9.71563	22.2083	11.2657	9.59421	17.9065	10.2995	7.55382	2.61904
SUIZA	29.6873	25.5369	15.0501	27.3193	8.57129	18.6177	16.1601	9.97145	5.45053	6.02146	17.8074
COLOMBIA	24.4045	47.542	33.1898	11.221	29.4611	4.15219	11.1853	27.0926	9.59119	12.5668	6.7226
CHILE	32.0303	37.6328	34.7773	16.4184	33.6182	22.2943	15.7259	22.6958	18.6595	15.2051	6.62335
PERÚ	31.039	57.0898	43.7649	7.86186	40.1258	4.93233	11.632	36.3284	16.8267	18.9641	6.20856
CUBA	23.9709	41.7527	29.2892	14.6243	25.3032	15.1779	13.4401	20.7453	12.6735	10.0557	6.56429
ECUADOR	33.4548	57.114	45.8206	9.33276	43.9904	9.09831	14.123	34.7462	19.2828	20.4681	3.53194
GUATEMALA	36.8543	70.8084	54.7767	10.5639	51.3371	11.2393	15.0124	44.1665	27.4097	30.4013	7.70393
HAITÍ	37.9187	91.571	63.8925	20.9468	61.3865	29.1003	28.295	56.3715	46.3065	48.6182	23.2858

	SUECIA	VENEZUELA	ARGENTINA	SUIZA	COLOMBIA	CHILE	PERÚ	CUBA	ECUADOR	GUATEMALA	HAITÍ
CHINA	38.4985	28.2032	23.6213	29.6873	24.4045	32.0303	31.039	23.9709	33.4548	36.8543	37.9187
CANADÁ	19.7578	39.2106	33.9013	25.5369	47.542	37.6328	57.0898	41.7527	57.114	70.8084	91.571
FRANCIA	19.1338	34.356	24.3952	15.0501	33.1898	34.7773	43.7649	29.2892	45.8206	54.7767	63.8925
INDIA	42.6343	12.9293	9.71563	27.3193	11.221	16.4184	7.86186	14.6243	9.33276	10.5639	20.9468
REINO UNIDO	17.3375	31.2697	22.2083	8.57129	29.4611	33.6182	40.1258	25.3032	43.9904	51.3371	61.3865
BRASIL	24.7764	7.13641	11.2657	18.6177	4.15219	22.2943	4.93233	15.1779	9.09831	11.2393	29.1003
SURÁFRICA	31.5192	9.97477	9.59421	16.1601	11.1853	15.7259	11.632	13.4401	14.123	15.0124	28.295
AUSTRALIA	13.9463	20.8021	17.9065	9.97145	27.0926	22.6958	36.3284	20.7453	34.7462	44.1665	56.3715
COREA DEL SUR	5.57875	10.0879	10.2995	5.45053	9.59119	18.6595	16.8267	12.6735	19.2828	27.4097	46.3065
ESPAÑA	10.1992	14.3174	7.55382	6.02146	12.5668	15.2051	18.9641	10.0557	20.4681	30.4013	48.6182
MÉXICO	30.1738	4.30767	2.61904	17.8074	6.7226	6.62335	6.20856	6.56429	3.53194	7.70393	23.2858
SUECIA	0	22.4462	25.3977	6.41775	21.2	35.3924	33.4863	24.9494	37.6059	46.0606	58.8832
VENEZUELA	22.4462	0	8.01645	15.0131	2.88066	14.9297	4.41306	10.3838	3.78653	6.1357	26.6692
ARGENTINA	25.3977	8.01645	0	12.49	8.1994	2.90838	8.9891	3.55588	7.73203	14.6588	33.3472
SUIZA	6.41775	15.0131	12.49	0	13.2979	21.4897	21.8228	13.7081	24.7934	30.8353	46.0608
COLOMBIA	21.2	2.88066	8.1994	13.2979	0	16.0511	2.94517	7.76666	5.40515	7.19221	23.4063
CHILE	35.3924	14.9297	2.90838	21.4897	16.0511	0	16.7684	5.68343	13.1398	22.0118	41.8509
PERÚ	33.4863	4.41306	8.9891	21.8228	2.94517	16.7684	0	12.9662	2.40088	3.15369	25.6412
CUBA	24.9494	10.3838	3.55588	13.7081	7.76666	5.68343	12.9662	0	11.1605	18.1363	26.4597
ECUADOR	37.6059	3.78653	7.73203	24.7934	5.40515	13.1398	2.40088	11.1605	0	1.97745	22.2975
GUATEMALA	46.0606	6.1357	14.6588	30.8353	7.19221	22.0118	3.15369	18.1363	1.97745	0	19.8776
HAITÍ	58.8832	26.6692	33.3472	46.0608	23.4063	41.8509	25.6412	26.4597	22.2975	19.8776	0

La distancia fue medida como la distancia euclídea al cuadrado : $d_{ij}^2 = \sum_{K=1}^p (X_{iK} - X_{jK})^2$

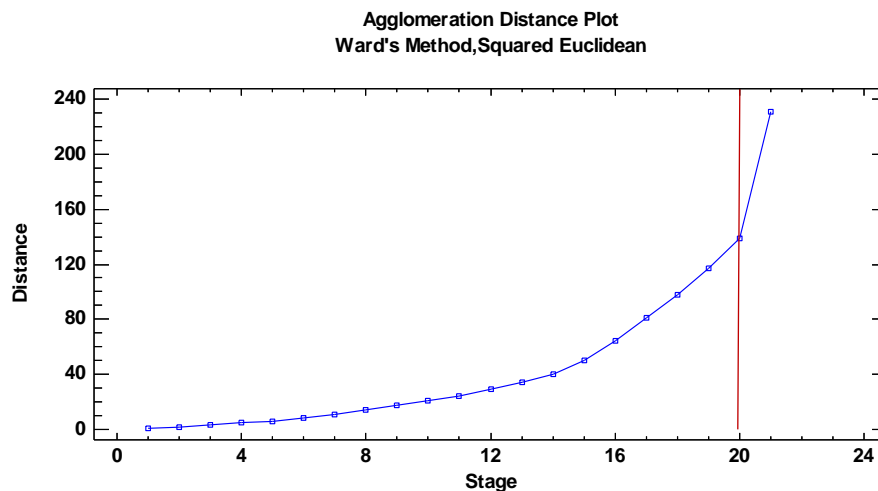
Donde: d_{ij}^2 representa la distancia entre las observaciones i y j ; X_{iK} es el valor de la variable X_K para la observación i

X_{jK} es el valor de la variable X_K para la observación j

Las casillas resaltadas para cada país son los 4 países que se encuentran a menor distancia con el país respectivo. Se espera que sean parte del mismo conglomerado.

ANEXO D: Análisis de conglomerados

El gráfico de distancia de aglomeración es útil para saber hasta que etapa conviene detener el proceso de conglomeración.

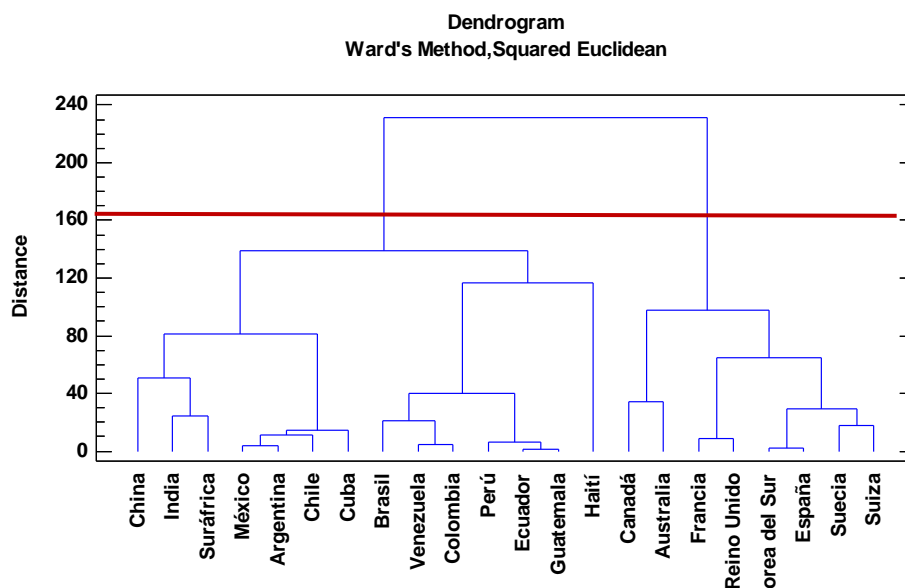


Analizando la gráfica vemos que en la etapa 20 la distancia entre los conglomerados es demasiado grande. Si se observa la tabla de las etapas de aglomeración, se tiene que en la etapa 20 los últimos objetos en unirse son el 1 y 6.

Agglomeration Schedule
Clustering Method: Ward's
Distance Metric: Squared Euclidean

	Combined	Combined		Previous Stage	Previous Stage	Next
Stage	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Stage
1	20	21	0.988723	0	0	5
2	9	10	2.04241	0	0	12
3	11	14	3.35193	0	0	7
4	13	16	4.79226	0	0	10
5	18	20	6.31421	0	1	14
6	3	5	8.51749	0	0	16
7	11	17	11.2582	3	0	8
8	11	19	14.1966	7	0	17
9	12	15	17.4054	0	0	12
10	6	13	20.6882	0	4	14
11	4	7	24.6503	0	0	15
12	9	12	29.3315	2	9	16
13	2	8	34.3456	0	0	18
14	6	18	39.9201	10	5	19
15	1	4	50.4088	0	11	17
16	3	9	64.7234	6	12	18
17	1	11	81.1334	15	8	20
18	2	3	97.5458	13	16	21
19	6	22	116.715	14	0	20
20	1	6	138.826	17	19	21
21	1	2	231.0	20	18	0

Observando el dendrograma vemos que se obtendría dos conglomerados.



Entonces los conglomerados estarán formados de la siguiente manera:

Row	Label	Cluster
1	China	1
2	Canadá	2
3	Francia	2
4	India	1
5	Reino Unido	2
6	Brasil	1
7	Suráfrica	1
8	Australia	2
9	Corea del Sur	2
10	España	2
11	México	1
12	Suecia	2
13	Venezuela	1
14	Argentina	1
15	Suiza	2
16	Colombia	1
17	Chile	1
18	Perú	1
19	Cuba	1
20	Ecuador	1
21	Guatemala	1
22	Haití	1

Se analizan los centroides para determinar características de los países pertenecientes a cada grupo.

Centroids

Cluster	tac	tmi	mpa	PNB	pe	lt	ca	pcb	pda	ce	CO2
1	1.89286	37.1429	33.9286	189091.	154181.	70.7857	549.929	29.6429	1.02857	911.071	2.54286
2	0.75	6.375	42.375	615076.	257279.	531.75	666.5	39.75	-0.35	4470.13	8.8875