

MAESTRÍA EN ECONOMETRÍA

ANÁLISIS ESTADÍSTICO MULTIVARIADO

Docentes:

- Clemente, Alejandra;
- Mendoza Greco, Maximiliano.

Alumnos:

- Dignani, Franco;
- Guzzi, David.

Ciclo lectivo: Segundo Trimestre 2024.



UNIVERSIDAD
TORCUATO DI TELLA

INTRODUCCIÓN

En este estudio se han seleccionado los países miembros de la **Organización para la Cooperación y el Desarrollo Económicos (OCDE)** junto con **Argentina** para analizar diversos indicadores económicos y sociales relacionados, de forma general, con la **Agricultura**. La elección de los países de la OCDE se basa en su diversidad económica y geográfica, así como en su relevancia en políticas internacionales y estándares de desarrollo. Argentina fue incluida como un caso de estudio adicional, proporcionando una perspectiva contrastante respecto a los países desarrollados. Se considera que este enfoque permite **comparar y contrastar una amplia gama de datos y naciones**, al tiempo de identificar patrones y tendencias clave.

Para el presente análisis, se ha seleccionado el **año 2019** como punto de referencia clave (dado que se considera representa el último año con condiciones económicas y sociales estables antes del impacto global de la pandemia de COVID-19) en conjunto con las **siguientes variables** que se resumen y definen del siguiente modo:

Variables	Definición
Agricultural land (% of land area)	Mide la proporción de tierras agrícolas, que incluye tierras arables, cultivos permanentes y pastos permanentes.
Arable land (% of land area)	Mide la proporción de tierra que se utiliza para cultivos temporales, pastos temporales y jardines, excluyendo tierras abandonadas por cultivo itinerante.
Agricultural raw materials exports (% of merchandise exports)	Mide proporción de exportaciones de materias primas agrícolas dentro del total de exportaciones de mercancías.
Agricultural raw materials imports (% of merchandise imports)	Mide la proporción de las importaciones de materias primas agrícolas respecto al total de importaciones de mercancías.
Agriculture, forestry, and fishing, value added (% of GDP)	Representa el valor agregado neto de los sectores de agricultura, silvicultura y pesca como porcentaje del PIB. Incluye la producción agrícola, ganadera, caza, pesca y silvicultura, y excluye la depreciación de activos y la degradación de recursos naturales.
Employment in agriculture (% of total employment) (modeled ILO estimate)	Mide la proporción de personas empleadas en el sector agrícola, que incluye actividades de agricultura, caza, silvicultura y pesca, en comparación con el total de empleo.
Rural population (% of total population)	Mide la proporción de personas que viven en áreas rurales como porcentaje de la población.

ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

Medidas descriptivas:

Indicador / Variable	Agricultural land	Arable land	Agricultural exports	Agricultural imports	Agriculture, forestry, and fishing, value added	Employment in agriculture	Rural population
Cantidad obs.	38	38	38	38	38	38	38
Promedio	0,3825	0,1836	0,0245	0,0138	0,0233	0,0501	0,2160
Desvío estándar	0,1752	0,1581	0,0287	0,0064	0,0157	0,0444	0,1119
Mínimo	0,0270	0,0120	0,0014	0,0055	0,0022	0,0068	0,0196
Percentil 25%	0,2973	0,0667	0,0069	0,0089	0,0128	0,0211	0,1351
Mediana	0,4374	0,1634	0,0134	0,0115	0,0188	0,0372	0,1936
Percentil 75%	0,4868	0,2740	0,0274	0,0178	0,0326	0,0574	0,2904
Máximo	0,7242	0,5990	0,1168	0,0291	0,0641	0,1811	0,4627
Coef. de Variación	0,4582	0,7521	1,1742	0,4652	0,6723	0,8860	0,5179
Kurtosis	-0,4182	0,7749	4,1573	0,0087	0,9200	1,9896	-0,3474
Asimetría	-0,4175	0,8868	2,1202	0,9792	1,1671	1,6272	0,5370
ARG	0,4259	0,1491	0,0088	0,0105	0,0532	0,0733	0,0801

La **matriz de datos** presenta **38 observaciones** pertenecientes a países de la **OCDE** y, sumando a **Argentina**, da **39 en total**. Si bien el **análisis considerará el conjunto de 39 observaciones**, aquí se dividen ambos grupos para **enriquecer la descripción** de cada una de las variables, que se pasa a comentar de forma general brevemente.

- En la gran mayoría de las variables se observa una disparidad importante entre los **valores mínimos y máximos**. Esto nos da cuenta de la **heterogeneidad de economías** presentes en matriz de datos;
- A pesar de ello, considerando la matriz de **forma agrupada** puede observarse que esta heterogeneidad parecería desaparecer dado que los guarismos para los **promedios** y las **medianas** no son muy disímiles, lo que indicaría que los **casos extremos** dentro de la matriz no serían significativos. Por otro lado, no obstante, algunas variables presentan altos **desvíos estándar**, como el caso de Agricultural land, Arable land, y Rural population.
- Con relación a la **distribución de las variables**, la gran mayoría presentan distribuciones positivas sesgadas hacia la derecha al tiempo de evidenciar colas de distribución alargadas (dados los valores extremos presentes).

MATRICES: COVARIANZAS Y CORRELACIONES

Matriz de Varianzas y Covarianzas:

Variables	Agricultural land (1)	Arable land (2)	Agricultural exports (3)	Agricultural imports (4)	Agriculture, forestry, and fishing, value added (5)	Employment in agriculture (6)	Rural population (7)
1	0,0299						
2	0,0145	0,0186					
3	-0,0015	-0,0009	0,0008				
4	0,0000	0,0005	0,0001	0,0000			
5	-0,0002	-0,0006	0,0002	0,0000	0,0005		
6	0,0005	-0,0012	0,0001	0,0000	0,0006	0,0019	
7	0,0029	0,0022	-0,0001	0,0002	-0,0001	0,0007	0,0127

Comentarios:

- Se observa **variancia máxima** en *Agricultural land* y **varianza mínima** en *Agricultural imports*.
- Predominan **relaciones positivas** entre las variables, siendo la más significativa entre *Agricultural land* y *Arable land*. Por su parte, la **relación negativa** más significativa se produce entre *Agricultural land* y *Agricultural exports* (contraintuitivo, en principio).
- Dado el rango que pueden tomar los valores (0 a 1), las medidas de variabilidad global indicarían un **bajo nivel de varianza global** (considerando o no los valores no presentes en la diagonal principal).
- El análisis de la **matriz de correlación**, que presenta guarismos no influidos por las unidades de medida de las distintas variables, permitirá confirmar los anteriores comentarios.

Matriz de Correlaciones:

Variables	Agricultural land (1)	Arable land (2)	Agricultural exports (3)	Agricultural imports (4)	Agriculture, forestry, and fishing, value added (5)	Employment in agriculture (6)	Rural population (7)
1	1						
2	0,6069	1					
3	-0,5079	-0,2549	1				
4	-0,0218	0,5737	0,5101	1			
5	-0,0596	-0,2523	0,5555	0,0419	1		
6	0,0674	-0,1942	0,0926	0,0654	0,8028	1	
7	0,1476	0,1410	-0,0522	0,2899	-0,0485	0,1450	1

Comentarios:

- Las variables, al ser estandarizadas, confirmarían los **grados y direcciones de relación** comentados anteriormente. Incluso, los guarismos de relación serían más altos en términos absolutos.
- Las medidas globales indicarían un bajo nivel de dependencia conjunta. Sin embargo, esto puede explicarse no porque las variables estén estrictamente correlacionadas sino más bien a que las mismas están expresadas en relación a otra medida de comparación común (como el PBI, por ejemplo).

Medidas de variabilidad global:

Medida	Valor
Varianza total	0,0642
Varianza media	0,0092
Varianza generalizada	0,0000
Varianza efectiva	0,0015

Medidas de dependencia global:

Medida	Valor
Dependencia conjunta	0,0642
Dependencia efectiva	0,0092

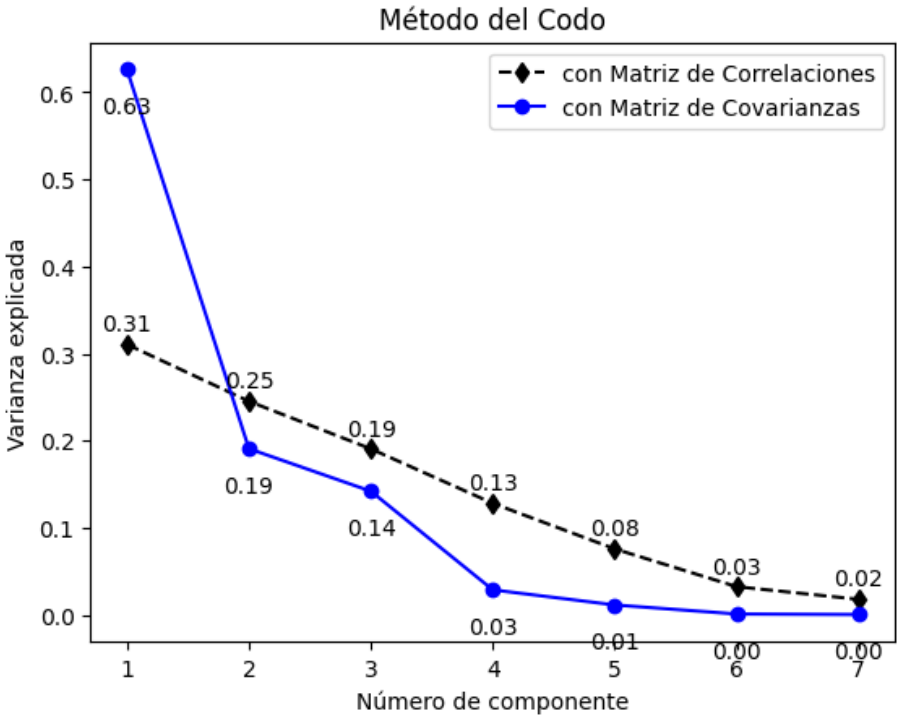
ANÁLISIS DE COMPONENTES PRINCIPALES

Se presentan a continuación los resultados del **análisis de componentes principales** para una **matriz de datos estandarizados (Matriz de Correlaciones)**:

Nº de Componente	Varianza	Desviación Estándar	Varianza Explicada	Varianza Exp. Acum.
Componente 1	2,25	1,49	0,51	0,51
Componente 2	1,76	1,33	0,25	0,56
Componente 3	1,57	1,17	0,19	0,75
Componente 4	0,92	0,96	0,15	0,87
Componente 5	0,54	0,74	0,08	0,95
Componente 6	0,25	0,48	0,05	0,98
Componente 7	0,15	0,36	0,02	1

Para una **matriz de datos no estandarizados (Matriz de Covarianzas)**:

Nº de Componente	Varianza	Desviación Estándar	Varianza Explicada	Varianza Exp. Acum.
Componente 1	0,04	0,20	0,63	0,63
Componente 2	0,01	0,11	0,19	0,82
Componente 3	0,01	0,10	0,14	0,96
Componente 4	0,00	0,04	0,05	0,99
Componente 5	0,00	0,03	0,01	1,00
Componente 6	0,00	0,01	0,00	1,00
Componente 7	0,00	0,00	0,00	1,00



Comentarios:

- Dado el objetivo de reducción de cantidad de variables con la menor pérdida de varianza posible para la matriz de datos analizada, se ha realizado un análisis de componente principales (PCA) de dos formas distintas, cuyos resultados son sensiblemente distintos.
- Dada la matriz de datos estandarizada, PCA nos indica que el primer componente explica en un 31% la varianza total de los datos. Sin embargo, dada la matriz de datos no estandarizada, PCA nos indica que este mismo guarismo es del 63%.
- Al presentar la información de varianzas explicadas de forma gráfica y conjunta para ambos análisis de PCA, puede concluirse que las diferencias entre ambos resultados se mantienen: mientras que el primer análisis indicaría que para contar con un nivel de varianza explicada del 95% se requerirían 5 componentes, el segundo análisis concluiría que mismo guarismo podría conseguirse con tan solo 3 componentes.