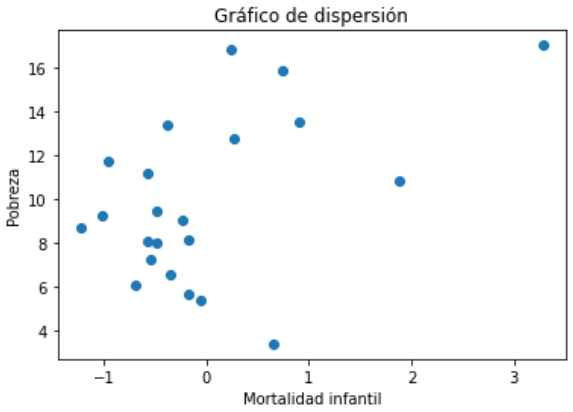
## Clase N° 11. Modelo de Clustering  jerárquico usando Python

A continuación se implementará un modelo de clustering jerárquico con el objetivo de determinar cuántos grupos de provincias se podrían formar en base a características comunes sobre mortalidad infantil e índice de pobreza.

En primer lugar, se representa mediante un gráfico de dispersión la relación entre las variables a analizar:

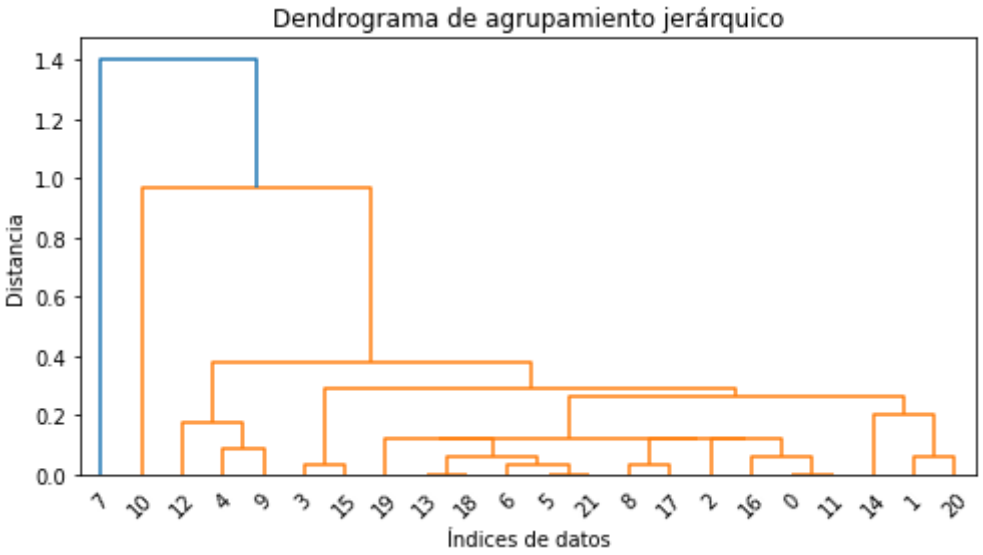



En el gráfico podemos observar un alto grado de dispersión entre la mortalidad infantil y la pobreza.

Luego, se crea una tabla entre las variables y se calculan las distancias entre todos los puntos, de acuerdo a un enlace jerárquico preestablecido según: la *distancia mínima, máxima*  o mediante los *puntos medios* de cada grupo.

```
array([[0.00000000e+00, 1.10000000e+01, 0.00000000e+00, 2.00000000e+00],
       [1.30000000e+01, 1.80000000e+01, 0.00000000e+00, 2.00000000e+00],
       [5.00000000e+00, 2.10000000e+01, 0.00000000e+00, 2.00000000e+00],
       [8.00000000e+00, 1.70000000e+01, 2.92576733e-02, 2.00000000e+00],
       [6.00000000e+00, 2.40000000e+01, 2.92576733e-02, 3.00000000e+00],
       [3.00000000e+00, 1.50000000e+01, 2.92576733e-02, 2.00000000e+00],
       [2.30000000e+01, 2.60000000e+01, 5.85153466e-02, 5.00000000e+00],
       [1.00000000e+00, 2.00000000e+01, 5.85153466e-02, 2.00000000e+00],
       [1.60000000e+01, 2.20000000e+01, 5.85153466e-02, 3.00000000e+00],
       [4.00000000e+00, 9.00000000e+00, 8.77730199e-02, 2.00000000e+00],
       [2.00000000e+00, 3.00000000e+01, 1.17030693e-01, 4.00000000e+00],
       [1.90000000e+01, 2.80000000e+01, 1.17030693e-01, 6.00000000e+00],
       [2.50000000e+01, 3.20000000e+01, 1.17030693e-01, 6.00000000e+00],
       [3.30000000e+01, 3.40000000e+01, 1.17030693e-01, 1.20000000e+01],
       [1.20000000e+01, 3.10000000e+01, 1.75546040e-01, 3.00000000e+00],
       [1.40000000e+01, 2.90000000e+01, 2.04803713e-01, 3.00000000e+00],
       [3.50000000e+01, 3.70000000e+01, 2.63319060e-01, 1.50000000e+01],
       [2.70000000e+01, 3.80000000e+01, 2.92576733e-01, 1.70000000e+01],
       [3.60000000e+01, 3.90000000e+01, 3.80349753e-01, 2.00000000e+01],
       [1.00000000e+01, 4.00000000e+01, 9.65503219e-01, 2.10000000e+01],
       [7.00000000e+00, 4.10000000e+01, 1.40436832e+00, 2.20000000e+01]])
```

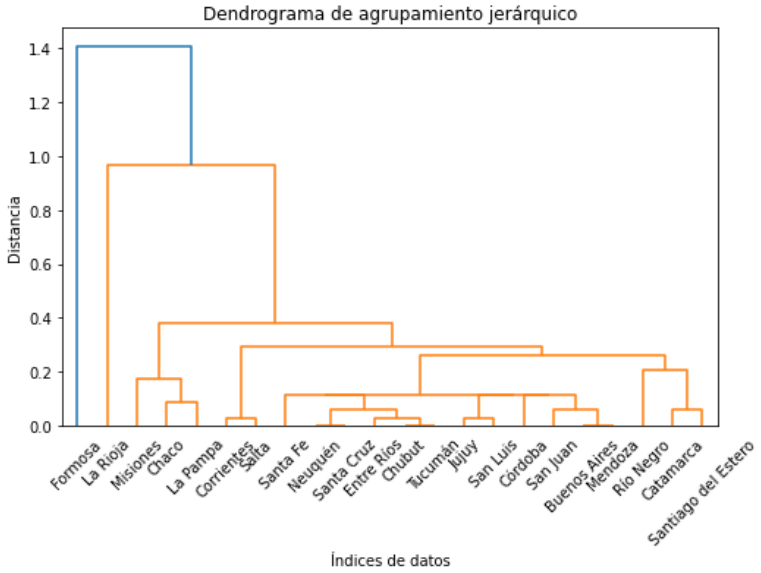
En base a la proximidad entre los puntos, se crea un dendograma a fin de poder visualizar los objetos que presenten mayor similitud o cercanía:



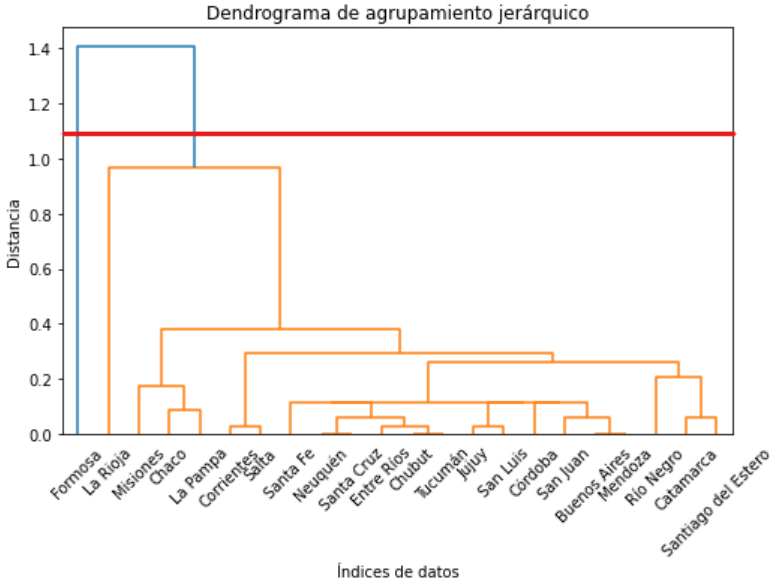

Por último, si reemplazamos los índices de los datos por las provincias,

```
Index(['Buenos Aires', 'Catamarca', 'Córdoba', 'Corrientes', 'Chaco', 'Chubut',
       'Entre Ríos', 'Formosa', 'Jujuy', 'La Pampa', 'La Rioja', 'Mendoza',
       'Misiones', 'Neuquén', 'Río Negro', 'Salta', 'San Juan', 'San Luis',
       'Santa Cruz', 'Santa Fe', 'Santiago del Estero', 'Tucumán'],
      dtype='object', name='province')
```

Se obtiene la siguiente representación:

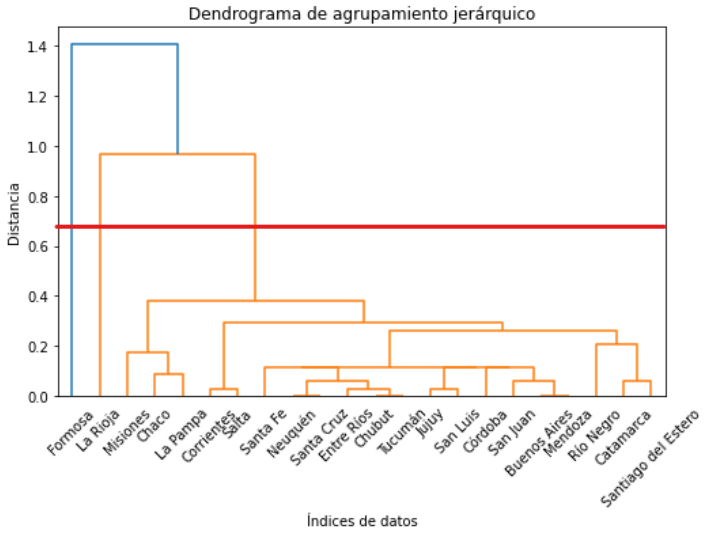


**Caso 1:** formar 2 grupos con características comunes



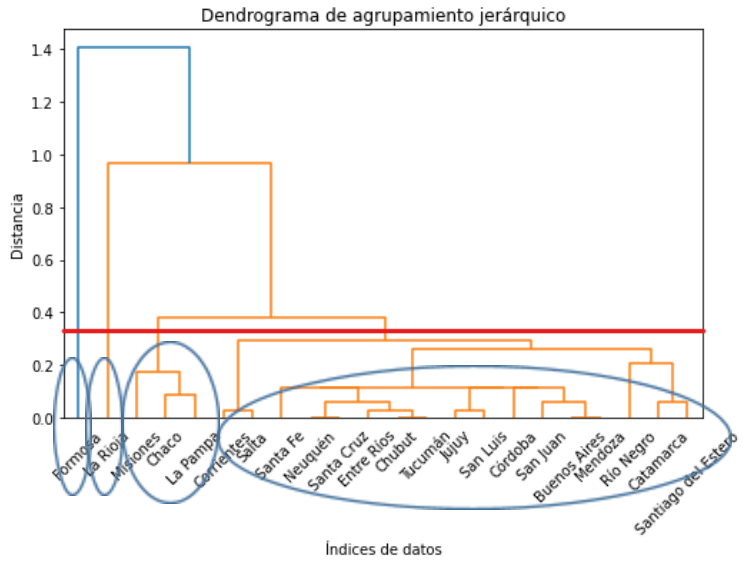

Se puede formar un clúster con la provincia de Formosa y un segundo clúster con el resto de las provincias.

**Caso 2:** formar 3 grupos con características comunes




Se puede formar un grupo con la provincia de Formosa, un segundo grupo con La Rioja y un tercer grupo con el resto de las provincias.

**Caso 3:** formar 4 grupos con características comunes




Se puede formar un grupo con la provincia de Formosa, un segundo grupo con La Rioja, un tercer grupo con: Misiones, Chaco y La Pampa  y un cuarto clúster con el resto de las provincias.

**Actividad:** Preguntas de opción múltiple.

Responder con verdadero o falso a las siguientes afirmaciones:

1. En el dendograma se muestra las distancias entre los índices de los datos.
2. Si aumentamos la cantidad de clústers obtendremos una mejor representación de las variables analizadas.
3. A mayor altura entre los nodos mayor similitud entre los grupos.

**Referencias:**

Para ampliar información sobre dendogramas, pueden visitar los sig. links:

https://plotly.com/python/dendrogram/

https://plotly.com/python-api-reference/generated/plotly.figure_factory.create_dendrogram.html