

Ejercicios de Evaluación:
Análisis Discriminante y Cluster

El archivo “DatosPaíses.xlsx” (que podeis descargar del campus) contiene información sobre 7 variables socioeconómicas de 133 países. Seleccionar aleatoriamente una muestra de 100 países con el procedimiento *surveyselect* de la siguiente forma:

```
proc surveyselect data=países method=srs n=100 out=sample_países;  
run;
```

Para la muestra obtenida, realizar un Análisis Cluster incluyendo SOLO las variables demográficas (*Pobl Natalidad Esperanza Vida Mortalidad*), que debe incluir como mínimo:

1. Análisis Cluster jerárquico con al menos dos métodos de agrupamiento. A partir de los procedimientos estudiados en clase, determinar el número (o números) adecuado de grupos.
2. Análisis Cluster no jerárquico con al menos dos métodos de selección de centroides iniciales. Teniendo en cuenta los resultados del apartado anterior, determinar a través del contraste F de Beale y otros procedimientos estudiados en clase, el número adecuado de grupos.
3. Caracterización de los clusters resultantes a partir de todas las variables del conjunto de datos.

Dados los resultados del análisis cluster, realizar un Análisis Discriminante tomando como variable de clasificación los clusters obtenidos, que debe incluir como mínimo:

1. Comprobación de la hipótesis de normalidad y la no existencia de datos atípicos, transformación en su caso de las variables.
2. Evaluar la hipótesis de igualdad de varianzas.
3. Selección de las variables a incluir en el análisis. Recuerda que el tamaño muestral de cada grupo debe ser mayor que el número de variables independientes.
4. Evaluar las funciones de clasificación resultantes (aplicando todos los métodos estudiados) y escribirlas explícitamente. En el caso de funciones cuadráticas, puedes omitir las interacciones aparentemente no significativas.
5. A partir de los coeficientes de las funciones, caracterizar los clusters.