**Segundo parcial AID Maestría en Data Mining julio de 2017**

**Ejercicio 1**

En el archivo siguiente se encuentran indicadores sociales de 22 países iberoamericanos.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pais** | **Población (miles)** | **Superficie (km2)** | **% menores 15 años** | **Esperanza vida al nacer** | **Tasa de mortalidad infan** | **Teléfonos por 1.000 hab** | **Usuarios Internet por 1000 hab** | **PIB $/hab** | **% PIB Agricultura** | **% PIB Industria** | **% PIB Servicios** |
| México | 95831 | 1958201 | 34 | 72.4 | 30.9 | 96 | 3.7 | 3670 | 5.4 | 26.3 | 68.3 |
| Guatemala | 11621 | 108899 | 44 | 64.2 | 45.9 | 27 | 0.8 | 1470 | 24.1 | 19.6 | 56.3 |
| Honduras | 6147 | 112088 | 43 | 69.8 | 34.9 | 29 | 0.9 | 660 | 21.7 | 30.9 | 47.4 |
| Nicaragua | 4464 | 130000 | 42 | 68.2 | 43.9 | 23 | 1.6 | 380 | 34.2 | 21.9 | 43.9 |
| El Salvador | 6059 | 21041 | 36 | 69.4 | 38.8 | 53 | 0.3 | 1700 | 13.1 | 26.8 | 60.1 |
| Costa Rica | 3650 | 51100 | 34 | 76.8 | 12.1 | 164 | 12.1 | 2640 | 15.5 | 24.1 | 60.4 |
| Panamá | 2767 | 75517 | 32 | 74 | 21.4 | 114 | 1.4 | 3080 | 8.2 | 18.4 | 73.4 |
| Cuba | 11115 | 110861 | 22 | 76 | 9 | 32 | 0.1 | 1983 | 6.9 | 46.3 | 46.8 |
| Rep. Dominicana | 8232 | 48734 | 34 | 71 | 33.5 | 73 | 0 | 1600 | 12.9 | 31.6 | 55.5 |
| Puerto Rico | 3806 | 8875 | 24 | 75.3 | 10.8 | 332 | 0.3 | 12213 | 0.9 | 46.8 | 52.3 |
| Colombia | 37685 | 1138914 | 33 | 70.9 | 24 | 100 | 1.8 | 2140 | 16.3 | 19.9 | 63.8 |
| Venezuela | 23242 | 912050 | 35 | 72.8 | 20.9 | 111 | 2.1 | 3020 | 4.1 | 46.5 | 49.4 |
| Brasil | 165158 | 8547403 | 30 | 67.9 | 42.1 | 75 | 4.2 | 4400 | 14 | 35.9 | 50.1 |
| Ecuador | 12175 | 283561 | 35 | 69.9 | 45.4 | 65 | 0.9 | 1500 | 11.9 | 36.7 | 51.4 |
| Perú | 24797 | 1285216 | 34 | 68.3 | 44.9 | 47 | 2.6 | 2420 | 7.4 | 36.9 | 55.7 |
| Bolivia | 7958 | 1098051 | 40 | 61.4 | 65.5 | 47 | 0.7 | 830 | 17 | 28.5 | 54.5 |
| Chile | 14824 | 756626 | 29 | 75.2 | 12.7 | 132 | 13.1 | 4860 | 8.4 | 35.9 | 62.3 |
| Paraguay | 5223 | 406752 | 40 | 69.7 | 39.1 | 34 | 0.5 | 1850 | 23.7 | 22.1 | 54.2 |
| Argentina | 36123 | 2780400 | 28 | 73.1 | 21.8 | 160 | 5.3 | 8380 | 6 | 30.7 | 63.3 |
| Uruguay | 3239 | 177414 | 24 | 72.8 | 16.5 | 196 | 3.2 | 5760 | 8.9 | 26.4 | 64.7 |
| España | 39754 | 505992 | 16 | 77.9 | 5.5 | 385 | 31 | 14350 | 5.9 | 37.8 | 56.3 |
| Portugal | 9798 | 91982 | 17 | 75 | 6.9 | 361 | 18.3 | 10160 | 1.7 | 53.2 | 45.1 |

1. Realizar un análisis de cluster jerárquico explicando la clasificación obtenida y el método elegido.
2. Idem para un cluster no jerárquico.
3. Indique dos maneras distintas de decidir el número de conglomerados.

**Ejercicio 2**

En la base empleados.xls se encuentran los datos correspondientes a los empleados de diferentes sucursales de una empresa, entre los cuales se ha detectado la presencia de una comunidad minoritaria.

1. Interesa saber si las variables de la base son significativamente distintas para los individuos de la minoría y para los otros. Plantee un test estableciendo las hipótesis necesarias.
2. En caso de rechazar realice un análisis discriminante lineal o cuadrático según corresponda y explique cuál ha elegido y por qué.
3. Analice de dos maneras el poder discriminante de la regla.
4. Excluya la última observación, genere nuevamente la regla y clasifique al último individuo mediante esta regla.

**Ejercicio 3**

Se asignan en forma aleatoria 25 pacientes que padecen hipertensión arterial a cinco grupos en los cuales se administran tres diferentes dosis de un fármaco o dieta hiposódica o bien una dieta hiposódica combinada con una dosis baja del fármaco. En la siguiente tabla están las presiones arteriales de los pacientes correspondientes a cada tratamiento.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dietahipos** | **Dieta+dosisbaja** | **dosis baja** | **dosis media** | **dosis alta** |
| 180 | 172 | 163 | 158 | 147 |
| 173 | 158 | 170 | 146 | 152 |
| 175 | 167 | 158 | 160 | 143 |
| 182 | 160 | 162 | 171 | 155 |
| 181 | 175 | 170 | 155 | 160 |

1. Plantee el modelo y las hipótesis para responder al problema planteado. Explicite los supuestos del modelo para comparar los valores medios de glóbulos rojos de las ratas de los cuatro grupos.
2. Realice la prueba y el análisis diagnostico(supuestos)
3. Si es válida, concluya, si no lo es, utilice otra prueba y concluya.
4. Explique en qué casos realizaría transformaciones de las variables.