**Trabajo Práctico N° 1**

**Ejercicio 1.**

*El archivo “eurosec.dta” es una base de datos que contiene los porcentajes de empleo de los distintos sectores económicos para un grupo de países europeos. Los sectores son: S1 (Agricultura), S2 (Minería), 53 (Industria), S4 (Energía), S5 (Construcción), S6 (Servicios Industriales), S7 (Finanzas), S8 (Servicios), S9 (Transporte y Telecomunicaciones).*

**(a)** *Obtener la media, varianza y el coeficiente de variación asociados a cada una de las variables.*

Stats | s1 s2 s3 s4 s5 s6 s7 s8 s9

---------+------------------------------------------------------------------------------------------

Mean | 19.13077 1.253846 27.00769 .9076923 8.165385 12.95769 4 20.02308 6.546154

Variance | 241.6958 .9409846 49.10874 .1415385 2.707954 20.93294 7.8768 46.64265 1.936185

CV | .8126474 .7736544 .2594728 .4144752 .201532 .3530916 .7016409 .3410836 .2125628

----------------------------------------------------------------------------------------------------

**(b)** *Obtener la matriz de varianzas y covarianzas.*

| s1 s2 s3 s4 s5 s6 s7 s8 s9

-------------+---------------------------------------------------------------------------------

s1 | 241.696

s2 | .539877 .940985

s3 | -73.1138 3.02637 49.1087

s4 | -2.33985 .147969 1.01594 .141538

s5 | -13.7721 -.040862 5.70228 .037077 2.70795

s6 | -52.421 -1.76003 6.53514 .347538 2.68048 20.9329

s7 | -9.592 -1.2052 -3.0648 .116 .0752 4.694 7.8768

s8 | -79.2911 -1.86169 7.37861 .340215 1.77843 17.8786 2.0632 46.6426

s9 | -12.2207 .211415 3.41963 .196431 .887662 1.19403 -.9604 5.39649 1.93618

**(c)** *Obtener las medidas globales de variabilidad.*

Varianza total= 371,98362.

Varianza media= 41,331513.

Varianza generalizada= 151,86262.

Varianza efectiva= 1,7473642.

**(d)** *Obtener la matriz de correlaciones.*

| s1 s2 s3 s4 s5 s6 s7 s8 s9

-------------+---------------------------------------------------------------------------------

s1 | 1.0000

s2 | 0.0358 1.0000

s3 | -0.6711 0.4452 1.0000

s4 | -0.4001 0.4055 0.3853 1.0000

s5 | -0.5383 -0.0256 0.4945 0.0599 1.0000

s6 | -0.7370 -0.3966 0.2038 0.2019 0.3560 1.0000

s7 | -0.2198 -0.4427 -0.1558 0.1099 0.0163 0.3656 1.0000

s8 | -0.7468 -0.2810 0.1542 0.1324 0.1582 0.5722 0.1076 1.0000

s9 | -0.5649 0.1566 0.3507 0.3752 0.3877 0.1876 -0.2459 0.5679 1.0000

**(e)** *Regresar el porcentaje de empleo en el sector Minería, respecto de las variables restantes. Obtener la varianza de los residuos y el coeficiente de determinación. ¿Es posible obtener las estimaciones a partir de la matriz de varianzas y covarianzas?*

Varianza de los residuos= 0,01730513.

Coeficiente de determinación= 0,98160955.

Sí, es posible obtener las estimaciones a partir de la matriz de varianzas y covarianzas y la matriz de precisión.

**(f)** *Obtener el coeficiente de correlación parcial entre los porcentajes de empleo en el sector agrícola respecto del sector minero. ¿Cómo efectuaría el cómputo a partir de la matriz de varianzas y covarianzas?*

El coeficiente de correlación parcial entre los porcentajes de empleo en el sector agrícola respecto del sector minero es -0,97480157.

**(g)** *Obtener los autovalores asociados a la matriz de correlaciones. ¿Existe alguna relación entre el número de variables y los mismos?*

e1 e2 e3 e4 e5 e6 e7 e8 e9

r1 3.4871512 2.1301732 1.0989576 .99448297 .54321777 .38342764 .22575405 .13678988 .00004563

Sí, existe una relación entre el número de variables y los autovalores, siendo la suma de estos últimos igual al número de variables.

**(h)** *Obtener el coeficiente de dependencia efectiva.*

Coeficiente de dependencia efectiva= 0,80178941.

**(i)** *Comentar, brevemente, la información que brinda la matriz de precisión.*

La matriz de precisión brinda información sobre la relación multivariada entre cada una de las variables y el resto. Contiene información sobre:

* por filas y por fuera de la diagonal principal, los coeficientes de regresión múltiple de la variable correspondiente a esa fila, explicada por todas las demás;
* en la diagonal, las inversas de las varianzas residuales de la regresión de cada variable con el resto;
* estandarizando los elementos de esta matriz, los elementos fuera de la diagonal principal son los coeficientes de correlación parcial entre estas variables.

Por lo tanto, contiene toda la información sobre las regresiones de cada variable sobre las restantes.

**Ejercicio 2.**

*En el archivo “individual\_t410.dta”, se encontrará el corte por personas de la Encuesta Permanente de Hogares correspondiente al cuarto trimestre del año 2010. Se propone la construcción de una base para analizar la estructura de la muestra ocupada mayor de 15 años de edad, por aglomerado y por rama de actividad, de acuerdo con la clasificación CAES-Mercosur, considerando las grandes ramas: Agricultura, Ganadería, Caza y Pesca; Minería; Industria; Energía; Construcción; Comercio; Correo y Telecomunicaciones; Ss. Financieros; Otros Ss.; y Administración Pública.*

**(a)** *Analizar la variabilidad de la proporción de ocupados por grandes ramas de actividad.*

media

Agricultur~a .05130951

Minería .02980964

Industria .34399797

Energía .01516862

Construcción .31695826

Comercio .61320805

Correo\_Tel~s .03139245

SS.\_Financ~s .04629727

Otros Ss. 1.3059524

Administra~a .37090579

var

Agricultur~a .00222073

Minería .00701461

Industria .11365215

Energía .00013202

Construcción .01454471

Comercio .11632129

Correo\_Tel~s .00062979

SS.\_Financ~s .00216025

Otros Ss. .49536335

Administra~a .0216334

cv

Agricultur~a .91843911

Minería 2.8096042

Industria .98001552

Energía .75747711

Construcción .3804963

Comercio .5561882

Correo\_Tel~s .79941574

SS.\_Financ~s 1.0039137

Otros Ss. .53893275

Administra~a .39655079

**(b)** *Analizar la estructura de correlaciones entre las proporciones de ocupados de las ramas de actividad consideradas.*

Varianza total= 0,77367232.

Varianza media= 0,07736723.

Varianza generalizada= 5,414e-25.

Varianza efectiva= 0,00374413.

**(c)** *Regresar la proporción de ocupados de la rama servicios, respecto de las proporciones observadas en las ramas restantes. Obtener la varianza de los residuos y el coeficiente de determinación. ¿De qué otra forma se hubiera podido obtener estas estimaciones?*

Varianza de los residuos= 0,01650179.

Coeficiente de determinación= 0,96668751.

Estas estimaciones se podrían haber obtenido mediante la matriz de varianzas y covarianzas y la matriz de precisión.

**(d)** *Obtener el coeficiente de correlación parcial entre los porcentajes de ocupados en el*

*sector comercial respecto del sector Otros Ss.*

El coeficiente de correlación parcial entre los porcentajes de ocupados en el sector Comercio respecto del sector Otros Ss. es 0,53293597.

**(e)** *Obtener los autovalores asociados a la matriz de correlaciones.*

e1 e2 e3 e4 e5 e6 e7 e8 e9 e10

r1 5.1644303 1.5298711 .97845644 .87863852 .54266892 .41764761 .26798865 .12086755 .07397417 .0254568

**(f)** *Proponer una métrica que resuma la dependencia entre las proporciones de ocupados*

*entre los distintos sectores.*

Coeficiente de dependencia efectiva= 0,64312145.

**Ejercicio 3.**

*El archivo “records.dta” contiene información sobre récords obtenidos por atletas de diferentes nacionalidades en varias especialidades. La siguiente tabla muestra la descripción del contenido de cada variable:*

|  |  |
| --- | --- |
| *m\_100* | *100 metros* |
| *m\_200* | *200 metros* |
| *m\_400* | *400 metros* |
| *m\_800* | *800 metros* |
| *m\_1000* | *1000 metros* |
| *m\_1500* | *1500 metros* |
| *km\_5* | *5 kilómetros* |
| *km\_10* | *10 kilómetros* |
| *maratón* | *42 kilómetros* |

**(a)** *Realizar un análisis descriptivo de los datos. Señalar si, en el caso de querer efectuar un análisis de componentes principales, se recomendaría la estandarizaci6n de las variables.*

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

m\_100 | 55 10.47109 .3514292 9.93 12.18

m\_200 | 55 20.94036 .6446476 19.72 23.2

m\_400 | 55 46.43873 1.457018 43.86 52.94

m\_800 | 55 1.793273 .0636848 1.7 2.02

m\_1500 | 55 3.698182 .1559094 3.51 4.24

-------------+---------------------------------------------------------

km\_5 | 55 13.84582 .8011605 13.01 16.7

km\_10 | 55 28.98964 1.80785 27.38 35.38

maraton | 55 136.624 9.227033 128.22 164.7

Stats | m\_100 m\_200 m\_400 m\_800 m\_1500 km\_5 km\_10 maraton

---------+--------------------------------------------------------------------------------

Mean | 10.47109 20.94036 46.43873 1.793273 3.698182 13.84582 28.98964 136.624

Variance | .1235025 .4155705 2.1229 .0040558 .0243077 .6418582 3.268322 85.13814

CV | .0335618 .0307849 .0313751 .0355132 .0421584 .057863 .0623619 .067536

------------------------------------------------------------------------------------------

| m\_100 m\_200 m\_400 m\_800 m\_1500 km\_5 km\_10 maraton

-------------+------------------------------------------------------------------------

m\_100 | 1.0000

m\_200 | 0.9226 1.0000

m\_400 | 0.8411 0.8507 1.0000

m\_800 | 0.7560 0.8066 0.8702 1.0000

m\_1500 | 0.7002 0.7750 0.8353 0.9180 1.0000

km\_5 | 0.6195 0.6954 0.7786 0.8636 0.9281 1.0000

km\_10 | 0.6324 0.6964 0.7872 0.8691 0.9346 0.9746 1.0000

maraton | 0.5199 0.5962 0.7050 0.8065 0.8655 0.9322 0.9432 1.0000

| m\_100 m\_200 m\_400 m\_800 m\_1500 km\_5 km\_10 maraton

-------------+------------------------------------------------------------------------

m\_100 | .123502

m\_200 | .209022 .41557

m\_400 | .4307 .799056 2.1229

m\_800 | .01692 .033115 .080743 .004056

m\_1500 | .038367 .077888 .189742 .009115 .024308

km\_5 | .17441 .359139 .90888 .044062 .115929 .641858

km\_10 | .4018 .811639 2.07364 .100059 .263438 1.41166 3.26832

maraton | 1.68601 3.54621 9.47786 .473903 1.24516 6.89105 15.7328 85.1381

Partial Semipartial Partial Semipartial Significance

Variable | corr. corr. corr.^2 corr.^2 value

------------+-----------------------------------------------------------------

m\_200 | 0.7288 0.3730 0.5312 0.1391 0.0000

m\_400 | 0.3330 0.1238 0.1109 0.0153 0.0194

m\_800 | 0.0658 0.0231 0.0043 0.0005 0.6534

m\_1500 | -0.1327 -0.0469 0.0176 0.0022 0.3633

km\_5 | -0.1080 -0.0381 0.0117 0.0014 0.4601

km\_10 | 0.1578 0.0560 0.0249 0.0031 0.2787

maraton | -0.1489 -0.0528 0.0222 0.0028 0.3071

En el caso de querer realizar un análisis de componentes principales, se recomendaría la estandarización de las variables, ya que, de lo contrario, se estarían considerando varianzas de unidades de medida diferentes.

**(b)** *Realizar un análisis de componentes principales.*

Basado en matriz de correlaciones:

Principal components/correlation Number of obs = 55

Number of comp. = 8

Trace = 8

Rotation: (unrotated = principal) Rho = 1.0000

--------------------------------------------------------------------------

Component | Eigenvalue Difference Proportion Cumulative

-------------+------------------------------------------------------------

Comp1 | 6.62209 5.74442 0.8278 0.8278

Comp2 | .877673 .718389 0.1097 0.9375

Comp3 | .159284 .035228 0.0199 0.9574

Comp4 | .124056 .044177 0.0155 0.9729

Comp5 | .0798788 .0119138 0.0100 0.9829

Comp6 | .067965 .021524 0.0085 0.9914

Comp7 | .046441 .0238273 0.0058 0.9972

Comp8 | .0226137 . 0.0028 1.0000

--------------------------------------------------------------------------

Principal components (eigenvectors)

------------------------------------------------------------------------------------------------------------

Variable | Comp1 Comp2 Comp3 Comp4 Comp5 Comp6 Comp7 Comp8 | Unexplained

-------------+--------------------------------------------------------------------------------+-------------

m\_100 | 0.3176 0.5669 0.3322 0.1277 0.2626 -0.5937 0.1367 0.1051 | 0

m\_200 | 0.3370 0.4616 0.3607 -0.2591 -0.1541 0.6559 -0.1133 -0.0962 | 0

m\_400 | 0.3557 0.2482 -0.5605 0.6523 -0.2182 0.1568 -0.0028 0.0004 | 0

m\_800 | 0.3687 0.0124 -0.5324 -0.4800 0.5401 -0.0147 -0.2382 -0.0375 | 0

m\_1500 | 0.3728 -0.1398 -0.1534 -0.4046 -0.4876 -0.1575 0.6105 0.1380 | 0

km\_5 | 0.3644 -0.3120 0.1900 0.0296 -0.2541 -0.1417 -0.5900 0.5478 | 0

km\_10 | 0.3668 -0.3069 0.1812 0.0804 -0.1332 -0.2192 -0.1784 -0.7965 | 0

maraton | 0.3419 -0.4389 0.2635 0.2993 0.4980 0.3156 0.3989 0.1573 | 0

------------------------------------------------------------------------------------------------------------

Basado en matriz de varianzas y covarianzas:

Principal components/covariance Number of obs = 55

Number of comp. = 8

Trace = 91.73866

Rotation: (unrotated = principal) Rho = 1.0000

--------------------------------------------------------------------------

Component | Eigenvalue Difference Proportion Cumulative

-------------+------------------------------------------------------------

Comp1 | 89.9139 88.5012 0.9801 0.9801

Comp2 | 1.4127 1.15277 0.0154 0.9955

Comp3 | .25993 .150462 0.0028 0.9983

Comp4 | .109468 .0821851 0.0012 0.9995

Comp5 | .0272831 .0145485 0.0003 0.9998

Comp6 | .0127347 .0104898 0.0001 1.0000

Comp7 | .00224482 .00179918 0.0000 1.0000

Comp8 | .000445635 . 0.0000 1.0000

--------------------------------------------------------------------------

Principal components (eigenvectors)

------------------------------------------------------------------------------------------------------------

Variable | Comp1 Comp2 Comp3 Comp4 Comp5 Comp6 Comp7 Comp8 | Unexplained

-------------+--------------------------------------------------------------------------------+-------------

m\_100 | 0.0199 0.2107 -0.0295 0.3588 -0.1902 0.8869 0.0523 -0.0139 | 0

m\_200 | 0.0416 0.3589 -0.0193 0.8335 0.0479 -0.4101 -0.0623 -0.0038 | 0

m\_400 | 0.1106 0.8278 -0.3773 -0.3964 0.0123 -0.0476 -0.0204 -0.0095 | 0

m\_800 | 0.0055 0.0232 0.0053 0.0096 0.0110 -0.0072 0.2610 0.9649 | 0

m\_1500 | 0.0144 0.0446 0.0499 0.0162 0.0434 -0.0672 0.9592 -0.2620 | 0

km\_5 | 0.0793 0.1300 0.3365 -0.0178 0.9092 0.1839 -0.0527 -0.0000 | 0

km\_10 | 0.1811 0.2990 0.8488 -0.1340 -0.3642 -0.0680 -0.0456 0.0044 | 0

maraton | 0.9728 -0.1808 -0.1419 0.0283 -0.0066 0.0035 0.0010 -0.0009 | 0

------------------------------------------------------------------------------------------------------------

**(c)** *Efectuar la selección de los componentes principales de acuerdo con los siguientes modos:*

**(i)** *Búsqueda del “codo”.*



Mediante este modo, se seleccionarían los dos primeros componentes principales.

**(ii)** *Graficando el porcentaje de varianza explicada por cada componente.*





Mediante este modo, se seleccionarían los dos primeros componentes principales.

**(d)** *Graficar las componentes en pares, en función de lo determinado en el inciso anterior. Interpretar.*

