

Laboratorio 1: Construcción de la Inercia

KEVIN STEVEN GARCÍA^a, ALEJANDRO VARGAS^b, ALEJANDRO SOTO^c

1. Introducción

En este informe se llevara a cabo un proceso de análisis multivariado con conceptos como la Inercia, centrado y estandarización de los datos, el cálculo de la matriz de varianzas y covarianzas y la matriz de correlaciones, descomposición en valores y vectores propios, construcción de las componentes principales o las proyecciones de la matriz de datos con los vectores propios asociados, y además, se construirán gráficos tanto en segunda como en tercera dimensión donde se representarán las plantas de acuerdo a sus 4 variables. Este proceso se le realizará a la siguiente matriz que contiene datos de 10 plantas acerca de 4 variables que son, X1: altura de la planta (en metros), X2: longitud radicular (en cms), X3: área foliar (en cms^3) y X4: peso en pulpa (en gramos)

Planta	X1	X2	X3	X4
1	1.89	53	4.5	120
2	1.92	62	5.2	136
3	1.95	71	5.4	145
4	2.11	75	6.1	156
5	1.78	58	5.0	127
6	2.12	77	6.5	160
7	2.06	82	7.6	169
8	2.19	78	7.3	164
9	1.93	60	5.5	142
10	1.84	72	5.7	147

2. Punto 1

- Matriz de datos original:

Primero, obtenemos el centro de gravedad o individuo promedio, en este caso, obtendremos la planta promedio de la siguiente manera:

$$g = (\bar{X}1, \bar{X}2, \bar{X}3, \bar{X}4)$$

Entonces:

$$g = (1.979, 68.8, 5.88, 146.6)$$

^aCódigo: 1533173. E-mail: kevin.chica@correounivalle.edu.co

^bCódigo: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co

^cCódigo: 1532457. E-mail: asotomurillo@gmail.com

Posteriormente, calculamos la inercia con la formula:

$$I(N) = \sum p_i \|x_i - g\|_M^2$$

Lo que nos dio:

$$I(N) = 317.9118$$

- Matriz con datos centrados:

Para la matriz con datos centrados (a cada valor de la variable se le resta su media correspondiente) sabemos que el centro de gravedad g está en el origen de coordenadas en R^p , como en nuestro caso tenemos 4 variables el centro de gravedad será

$$g = (0, 0, 0, 0)$$

Entonces, la formula para calcular la inercia se convierte en

$$I(N) = \sum p_i \|x_i\|_M^2$$

Haciendo uso de la formula obtuvimos el mismo resultado anterior:

$$I(N) = 317.9118$$

3. Punto 2

Para la matriz de datos estandarizada, es importante recordar que el proceso de estandarización se debe realizar con la desviación estándar sin la corrección de Bessel, es decir, se debe dividir sobre n y no sobre $n-1$.

Con esta matriz, el centro de gravedad g también es el origen de coordenadas en R^p (R^4 en nuestro caso)

$$g = (0, 0, 0, 0)$$

Aplicando la formulas de la inercia con el centro de gravedad anterior, obtenemos

$$I(N) = 4$$

La diferencia que se logra apreciar entre el calculo de la inercia utilizando la matriz de datos original o centrada y la matriz de datos estandarizada es que en los primeros dos casos, dicha inercia da como resultado la suma de las varianzas de cada variable, mientras que en el segundo caso (matriz de datos estandarizada) la inercia da como resultado el numero de variables involucradas, 4 en nuestro caso.

4. Punto 3

La matriz de varianzas y covarianzas de la matriz de datos es:

	Altura planta	Longitud radicular	Área foliar	Peso pulpa
Altura planta	0.01796556	0.9842222	0.1056444	1.746222
Longitud radicular	0.98422222	96.6222222	8.8733333	151.8
Área foliar	0.10564444	8.8733333	0.9951111	15.246667
Peso pulpa	1.746222	151.8	15.246667	255.6

5. Punto 4

La matriz de correlaciones es

	Altura planta	Longitud radicular	Área foliar	Peso pulpa
Altura planta	1	0.7470236	0.7901154	0.8148896
Longitud radicular	0.7470236	1	0.9049248	0.9659460
Área foliar	0.7901154	0.9049248	1	0.9560018
Peso pulpa	0.8148896	0.9659460	0.9560018	1

6. Punto 5

- Matriz de varianzas y covarianzas:

La descomposición de esta matriz en valores y vectores propios, nos dio

Los valores propios

$$\lambda_1 = 348.3709, \lambda_2 = 4.778776, \lambda_3 = 0.08000646, \lambda_4 = 0.005602631$$

Los vectores propios

λ_1	λ_2	λ_3	λ_4
-0.005759331	-0.01300728	0.001249545	0.9998980345
-0.517026145	0.85544272	-0.028891125	0.0081861926
-0.050711126	-0.06431056	-0.996640617	0.0001167914
-0.854446708	-0.51372435	0.076623899	-0.0117001348

- Matriz de correlaciones:

La descomposición de esta matriz en valores y vectores propios, nos dio

Los valores propios

$$\lambda_1 = 3.59564465, \lambda_2 = 0.29512043, \lambda_3 = 0.09190433, \lambda_4 = 0.01733060$$

Los vectores propios

λ_1	λ_2	λ_3	λ_4
-0.4636832	0.8742229	0.11689440	0.08407071
-0.5046999	-0.3941588	0.61315298	0.46255844
-0.5089282	-0.1956428	-0.77874615	0.31027468
-0.5208339	-0.20517575	0.06271807	-0.82625748

7. Punto 6

8. Punto 7

9. Punto 8

- Proyección de la matriz de datos originales asociada a los vectores propios de la matriz de varianzas y covarianzas:

Esta proyección la podemos encontrar haciendo el producto $A \cdot u$ donde A es la matriz de datos originales y u es la matriz compuesta por los vectores propios correspondientes a la matriz de varianzas y covarianzas. En nuestro caso:

$$\begin{pmatrix} 1.89 & 53 & 4.5 & 120 \\ 1.92 & 62 & 5.2 & 136 \\ 1.95 & 71 & 5.4 & 145 \\ 2.11 & 75 & 6.1 & 156 \\ 1.78 & 58 & 5.0 & 127 \\ 2.12 & 77 & 6.5 & 160 \\ 2.06 & 82 & 7.6 & 169 \\ 2.19 & 78 & 7.3 & 164 \\ 1.93 & 60 & 5.5 & 142 \\ 1.84 & 72 & 5.7 & 147 \end{pmatrix} \cdot \begin{pmatrix} -0.005759331 & -0.01300728 & 0.001249545 & 0.9998980345 \\ -0.517026145 & 0.85544272 & -0.028891125 & 0.0081861926 \\ -0.050711126 & -0.06431056 & -0.996640617 & 0.0001167914 \\ -0.854446708 & -0.51372435 & 0.076623899 & -0.0117001348 \end{pmatrix}$$

$$= \begin{pmatrix} -130.1751 & -16.62244 & 3.181117 & 0.9201849 \\ -148.5351 & -17.18845 & 3.449468 & 0.8367372 \\ -160.8887 & -14.12624 & 3.679773 & 0.8351320 \\ -172.3921 & -16.40253 & 3.709623 & 0.8992407 \\ -138.7661 & -15.97202 & 3.074571 & 0.7692845 \\ -176.8643 & -16.77240 & 3.559692 & 0.8788583 \\ -187.1949 & -17.18867 & 3.008472 & 0.7546226 \\ -180.8401 & -18.02421 & 3.040072 & 0.9103302 \\ -152.6430 & -22.00111 & 3.668014 & 0.7601980 \\ -163.1292 & -14.31611 & 3.505000 & 0.7099642 \end{pmatrix}$$

- Proyección de la matriz de datos centrados asociada a los vectores propios de la matriz de varianzas y covarianzas:

Esta proyección la podemos encontrar haciendo el producto $A \cdot u$ donde A es la matriz de datos centrados y u es la matriz compuesta por los vectores propios correspondientes a la matriz de varianzas y covarianzas. En nuestro caso:

$$\begin{pmatrix} 1.89 & 53 & 4.5 & 120 \\ 1.92 & 62 & 5.2 & 136 \\ 1.95 & 71 & 5.4 & 145 \\ 2.11 & 75 & 6.1 & 156 \\ 1.78 & 58 & 5.0 & 127 \\ 2.12 & 77 & 6.5 & 160 \\ 2.06 & 82 & 7.6 & 169 \\ 2.19 & 78 & 7.3 & 164 \\ 1.93 & 60 & 5.5 & 142 \\ 1.84 & 72 & 5.7 & 147 \end{pmatrix} \cdot \begin{pmatrix} -0.005759331 & -0.01300728 & 0.001249545 & 0.9998980345 \\ -0.517026145 & 0.85544272 & -0.028891125 & 0.0081861926 \\ -0.050711126 & -0.06431056 & -0.996640617 & 0.0001167914 \\ -0.854446708 & -0.51372435 & 0.076623899 & -0.0117001348 \end{pmatrix}$$

$$= \begin{pmatrix} -130.1751 & -16.62244 & 3.181117 & 0.9201849 \\ -148.5351 & -17.18845 & 3.449468 & 0.8367372 \\ -160.8887 & -14.12624 & 3.679773 & 0.8351320 \\ -172.3921 & -16.40253 & 3.709623 & 0.8992407 \\ -138.7661 & -15.97202 & 3.074571 & 0.7692845 \\ -176.8643 & -16.77240 & 3.559692 & 0.8788583 \\ -187.1949 & -17.18867 & 3.008472 & 0.7546226 \\ -180.8401 & -18.02421 & 3.040072 & 0.9103302 \\ -152.6430 & -22.00111 & 3.668014 & 0.7601980 \\ -163.1292 & -14.31611 & 3.505000 & 0.7099642 \end{pmatrix}$$

- Proyección de la matriz de datos estandarizados asociada a los vectores propios de la matriz de correlaciones:

Esta proyección la podemos encontrar haciendo el producto $A \cdot u$ donde A es la matriz de datos estandarizados y u es la matriz compuesta por los vectores propios correspondientes a la matriz de correlaciones. En nuestro caso:

$$\begin{pmatrix} 1.89 & 53 & 4.5 & 120 \\ 1.92 & 62 & 5.2 & 136 \\ 1.95 & 71 & 5.4 & 145 \\ 2.11 & 75 & 6.1 & 156 \\ 1.78 & 58 & 5.0 & 127 \\ 2.12 & 77 & 6.5 & 160 \\ 2.06 & 82 & 7.6 & 169 \\ 2.19 & 78 & 7.3 & 164 \\ 1.93 & 60 & 5.5 & 142 \\ 1.84 & 72 & 5.7 & 147 \end{pmatrix} \cdot \begin{pmatrix} -0.005759331 & -0.01300728 & 0.001249545 & 0.9998980345 \\ -0.517026145 & 0.85544272 & -0.028891125 & 0.0081861926 \\ -0.050711126 & -0.06431056 & -0.996640617 & 0.0001167914 \\ -0.854446708 & -0.51372435 & 0.076623899 & -0.0117001348 \end{pmatrix}$$

$$= \begin{pmatrix} -130.1751 & -16.62244 & 3.181117 & 0.9201849 \\ -148.5351 & -17.18845 & 3.449468 & 0.8367372 \\ -160.8887 & -14.12624 & 3.679773 & 0.8351320 \\ -172.3921 & -16.40253 & 3.709623 & 0.8992407 \\ -138.7661 & -15.97202 & 3.074571 & 0.7692845 \\ -176.8643 & -16.77240 & 3.559692 & 0.8788583 \\ -187.1949 & -17.18867 & 3.008472 & 0.7546226 \\ -180.8401 & -18.02421 & 3.040072 & 0.9103302 \\ -152.6430 & -22.00111 & 3.668014 & 0.7601980 \\ -163.1292 & -14.31611 & 3.505000 & 0.7099642 \end{pmatrix}$$

10. Punto 9

11. Punto 10

- Gráfico en 2 dimensiones de la proyección encontrada de la matriz de datos original y los vectores propios de la matriz de varianzas y covarianzas:



FIGURA 1: Gráfico de la proyección de la matriz de datos original

- Gráfico en 2 dimensiones de la proyección encontrada de la matriz de datos centrada y los vectores propios de la matriz de varianzas y covarianzas:

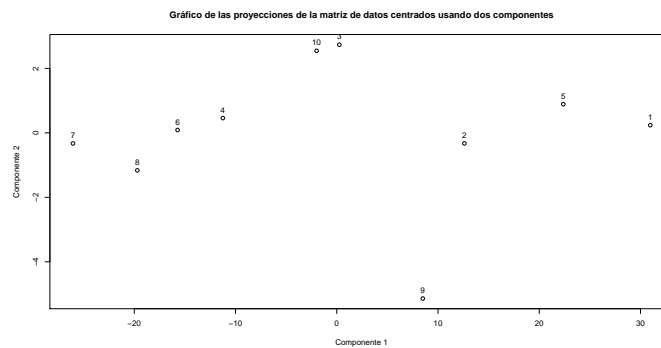


FIGURA 2: Gráfico de la proyección de la matriz de datos centrada

- Gráfico en 2 dimensiones de la proyección encontrada de la matriz de datos estandarizada y los vectores propios de la matriz de correlaciones:

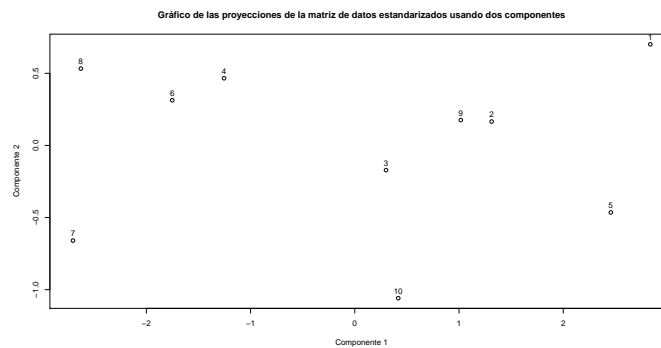


FIGURA 3: Gráfico de la proyección de la matriz de datos estandarizada

?????