* 1. **Análisis de Componentes Principales**

En esta sección se realizará el análisis de las componentes principales, pero en primera instancia las definiremos. Es importante indicar que las definiciones descritas en esta sección están referidas a la muestra y por tanto se utilizarán los estimadores de los parámetros poblacionales.

Sea Rp, vector aleatorio p variado no necesariamente normal,  es observable; T = (X1 X2 X3 ... Xp), Σ es la matriz de varianzas y covarianzas.

Sean  y **a1, a2, ...,ap** , los valores y vectores propios de Σ; consideremos las siguientes combinaciones lineales:









Se puede probar que:

 donde i=1,2,...,p

 donde i, j=1,2,...,p

Las componentes principales son las combinaciones lineales Y1, Y2,......,Yp que no están correlacionadas entre sí y cuyas varianzas cumplen que Var(Y1)****Var(Y2)**** . . . ****Var(Yp)****0. Se define entonces a las componentes principales como sigue:

 es la primera componente principal y es la combinación de máxima varianza, sujeto a la condición que  = 1, esto restringe los coeficientes de los vectores a la longitud unitaria.

es la segunda componente principal, maximiza Var() =  , sujeto a  = 1 ,  = 0 y 

En general:

es la i-ésima componente principal, maximiza Var() =  sujeto a  = 1,  = 0, para i<j. y 

En si, componentes principales es una técnica multivariada que no hace supuestos de normalidad, sus objetivos son (1) la reducción de datos, trata de disminuir el número de variables observadas en un número menor de componentes principales  que son combinaciones lineales de las variables observadas Xi y (2) la interpretación.

Siendo Σ la matriz de varianzas y covarianzas de  cuya dimensión es p x p, , los pares de valores y vectores propios ortonormales asociados a Σ , siendo , la i-ésima componente principal está dada por:



siendo además:

 donde i=1,2,...,p

 para ij

La mayor proporción del total de varianza de la población explicada por las componentes principales está dado por:

 para k=1,2,...,p

Para el presente trabajo las variables de estudio establecidas son 55, las variables de matemáticas y lenguaje con sus respectivas notas, la edad, sexo, jornada y actividad extra escolar. Con los datos recogidos se calcula la matriz de varianzas y covarianzas de la muestra Sn, matriz con la que se obtienen las componentes principales.

**Resultados con la matriz de datos originales**

Ahora se presentará una generalización de las  variables aleatorias no observables construidas a partir de las n variables aleatorias observables:



Para obtener los  primero tenemos que encontrar los valores propios por medio de la ecuación característica:



Con esto tenemos una ecuación en grado n = 55, debido a que la dimensión de la matriz de varianzas y covarianzas Sn es 55 x 55.

Los 55 valores propios estimados asociados a la matriz de varianzas y covarianzas Sn, ordenados de mayor a menor se presenta en la Tabla LXXXIX.

## TABLA LXXXIX

## Valores propios estimados asociados

**a la matriz de varianzas y covarianzas Sn**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | = | 461,9402464 |  | = | 0,450963475 |  | = | 0,128654491 |
|  | = | 122,3343764 |  | = | 0,420462647 |  | = | 0,099561066 |
|  | = | 3,578751519 |  | = | 0,381996855 |  | = | 0,095950107 |
|  | = | 2,150687721 |  | = | 0,351770791 |  | = | 0,083164875 |
|  | = | 1,830150006 |  | = | 0,318721141 |  | = | 0,079919226 |
|  | = | 1,592514382 |  | = | 0,299191371 |  | = | 0,07539143 |
|  | = | 1,410807437 |  | = | 0,293965405 |  | = | 0,074651837 |
|  | = | 1,376098254 |  | = | 0,274572589 |  | = | 0,066494343 |
|  | = | 1,152593855 |  | = | 0,24706889 |  | = | 0,06380648 |
|  | = | 0,99463429 |  | = | 0,241887051 |  | = | 0,056942776 |
|  | = | 0,843238025 |  | = | 0,227731591 |  | = | 0,044855598 |
|  | = | 0,82542528 |  | = | 0,217013845 |  | = | 0,035645931 |
|  | = | 0,777132026 |  | = | 0,202233931 |  | = | 0,031585629 |
|  | = | 0,734469553 |  | = | 0,187406527 |  | = | 0,026643672 |
|  | = | 0,702453886 |  | = | 0,166488474 |  | = | 0,011592971 |
|  | = | 0,664157298 |  | = | 0,165810914 |  | = | 0,010023144 |
|  | = | 0,635414377 |  | = | 0,157087015 |  | = | 0,00187126 |
|  | = | 0,545258286 |  | = | 0,147758944 |  |  |  |
|  | = | 0,51378578 |  | = | 0,135008207 |  |  |  |

Con cada uno de estos valores calculamos los vectores propios estimados ortonormales , aquellos que son ortogonales todos con todos, y cuya norma es 1. Estos son los coeficientes de las componentes principales.

**Gráfico 4.1**

**Número de la componente Vs. Valor Propio (matriz sn)**

Una de las formas para determinar el número de componentes principales que se deben escoger, es graficando los valores propios obtenidos frente al número de la componente. Si observamos el gráfico 4.1, notaremos que a partir de la segunda componente los valores propios son relativamente pequeños por lo tanto no representarán mayor cantidad de la varianza total.

Un dato que ofrece mucha información y que también nos permite decidir el número de componente principales que se deben escoger, es el porcentaje de explicación de cada componente principal, el cual es mostrado en la Tabla XC:



Por lo general, aspiramos obtener en tres o menos componentes principales del 80 al 90% del total de la varianza de la población, permitiéndonos reemplazar para un valor p grande las variables originales por las componentes principales perdiendo un mínimo de información.

## TABLA XC

## Porcentajes de explicación de las componentes principales obtenidas a partir de la matriz de covarianzas muestral Sn

|  |  |  |
| --- | --- | --- |
| **Componente principal** | **Porcentaje de Explicación** | **Porcentaje acumulado de explicación** |
| Y1 | 75,66885166 | 75,6688517 |
| Y2 | 20,03917574 | 95,7080274 |
| Y3 | 0,586223045 | 96,2942504 |
| Y4 | 0,352296799 | 96,6465472 |
| Y5 | 0,299790612 | 96,9463379 |
| Y6 | 0,260864334 | 97,2072022 |
| Y7 | 0,23109954 | 97,4383017 |
| Y8 | 0,225413948 | 97,6637157 |
| Y9 | 0,188802457 | 97,8525181 |
| Y10 | 0,162927641 | 98,0154458 |
| Y11 | 0,138127937 | 98,1535737 |
| Y12 | 0,135210092 | 98,2887838 |
| Y13 | 0,127299339 | 98,4160831 |
| Y14 | 0,120310945 | 98,5363941 |
| Y15 | 0,115066568 | 98,6514607 |
| Y16 | 0,108793335 | 98,760254 |
| Y17 | 0,104085056 | 98,864339 |
| Y18 | 0,089316895 | 98,9536559 |
| Y19 | 0,084161491 | 99,0378174 |
| Y20 | 0,073870784 | 99,1116882 |
| Y21 | 0,068874548 | 99,1805628 |
| Y22 | 0,062573598 | 99,2431364 |
| Y23 | 0,05762237 | 99,3007587 |
| Y24 | 0,05220862 | 99,3529674 |
| Y25 | 0,049009515 | 99,4019769 |
| Y26 | 0,048153467 | 99,4501303 |
| Y27 | 0,044976797 | 99,4951071 |

## TABLA XC

## Porcentajes de explicación de las componentes principales obtenidas a partir de la matriz de covarianzas muestral Sn (continuación)

|  |  |  |
| --- | --- | --- |
| **Componente principal** | **Porcentaje de Explicación** | **Porcentaje acumulado de explicación** |
| Y28 | 0,04047151 | 99,5355786 |
| Y29 | 0,03962269 | 99,5752013 |
| Y30 | 0,037303933 | 99,6125053 |
| Y31 | 0,035548296 | 99,6480536 |
| Y32 | 0,033127248 | 99,6811808 |
| Y33 | 0,030698422 | 99,7118792 |
| Y34 | 0,027271907 | 99,7391511 |
| Y35 | 0,027160919 | 99,7663121 |
| Y36 | 0,025731887 | 99,7920439 |
| Y37 | 0,024203887 | 99,8162478 |
| Y38 | 0,022115233 | 99,8383631 |
| Y39 | 0,021074452 | 99,8594375 |
| Y40 | 0,016308758 | 99,8757463 |
| Y41 | 0,015717259 | 99,8914635 |
| Y42 | 0,013622954 | 99,9050865 |
| Y43 | 0,013091295 | 99,9181778 |
| Y44 | 0,012349612 | 99,9305274 |
| Y45 | 0,012228462 | 99,9427559 |
| Y46 | 0,010892211 | 99,9536481 |
| Y47 | 0,010451921 | 99,9641 |
| Y48 | 0,009327601 | 99,9734276 |
| Y49 | 0,007347642 | 99,9807752 |
| Y50 | 0,005839038 | 99,9866143 |
| Y51 | 0,005173934 | 99,9917882 |
| Y52 | 0,004364409 | 99,9961526 |
| Y53 | 0,001899005 | 99,9980516 |
| Y54 | 0,001641857 | 99,9996935 |
| Y55 | 0,000306525 | 100 |

De acuerdo con los resultados obtenidos, las componentes principales que mayor información aportan son las dos primeras, debido a que su porcentaje de explicación acumulado es:



**TABLA XCI**

**Las dos primeras componentes principales**

**a partir de los datos originales**

|  |  |  |
| --- | --- | --- |
|  | Y1 | Y2 |
| X1 | -0,010 | -0,001 |
| X2 | 0,002 | -0,003 |
| X3 | 0,004 | 0,005 |
| X4 | 0,009 | 0,013 |
| X5 | 0,008 | 0,012 |
| X6 | 0,013 | 0,017 |
| X7 | 0,004 | 0,009 |
| X8 | 0,006 | 0,009 |
| X9 | 0,006 | 0,008 |
| X10 | 0,007 | 0,009 |
| X11 | 0,005 | 0,009 |
| X12 | 0,008 | 0,014 |
| X13 | 0,006 | 0,008 |
| X14 | 0,022 | 0,035 |
| X15 | 0,007 | 0,014 |
| X16 | 0,005 | 0,009 |
| X17 | 0,003 | 0,007 |
| X18 | 0,002 | 0,003 |
| X19 | 0,004 | 0,007 |
| X20 | 0,013 | 0,013 |
| X21 | 0,017 | 0,021 |
| X22 | 0,010 | 0,018 |
| X23 | 0,005 | 0,007 |
| X24 | 0,005 | 0,009 |
| X25 | 0,004 | 0,006 |
| X26 | 0,001 | 0,002 |
| X27 | 0,001 | 0,001 |
| X28 | 0,003 | 0,007 |
| X29 | 0,006 | 0,012 |
| X30 | 0,028 | -0,028 |

**TABLA XCI**

**Las dos primeras componentes principales**

**a partir de los datos originales**

**(Continuación)**

|  |  |  |
| --- | --- | --- |
|  | Y1 | Y2 |
| X31 | 0,005 | -0,002 |
| X32 | 0,009 | -0,003 |
| X33 | 0,008 | -0,007 |
| X34 | 0,028 | -0,016 |
| X35 | 0,041 | -0,001 |
| X36 | 0,015 | 0,004 |
| X37 | 0,031 | -0,019 |
| X38 | 0,019 | -0,011 |
| X39 | 0,022 | -0,014 |
| X40 | 0,024 | -0,016 |
| X41 | 0,025 | -0,021 |
| X42 | 0,025 | -0,017 |
| X43 | 0,018 | -0,010 |
| X44 | 0,013 | -0,007 |
| X45 | 0,013 | -0,007 |
| X46 | 0,012 | -0,006 |
| X47 | 0,012 | -0,006 |
| X48 | 0,020 | -0,007 |
| X49 | 0,038 | -0,023 |
| X50 | 0,021 | -0,015 |
| X51 | 0,019 | -0,010 |
| X52 | 0,512 | 0,855 |
| X53 | 0,851 | -0,511 |
| X54 | -0,004 | -0,010 |
| X55 | 0,001 | -0,001 |

A continuación se mostraran las dos primeras componentes principales originadas a partir de la matriz de varianzas y covarianzas de la muestra.

*Y*1 = - 0,010X1 + 0,002X2 + 0,004X3 + 0,009X4 + 0,008X5 + 0,013X6 + 0,004X7 + 0,006X8 + 0,006X9 + 0,007X10 + 0,005X11 + 0,008X12 + 0,006X13 + 0,022X14 + 0,007X15 + 0,005X16 + 0,003X17 + 0,002X18 + 0,004X19 + 0,013X20 + 0,017X21 + 0,010X22 + 0,005X23 + 0,005X24 + 0,004X25 + 0,001X26 + 0,001X27 + 0,003X28 + 0,006X29 + 0,028X30 + 0,005X31 + 0,009X32 + 0,008X33 + 0,028X34 + 0,041X35 + 0,015X36 + 0,031X37 + 0,019X38 + 0,022X39 + 0,024X40 + 0,025X41 + 0,025X42 + 0,018X43 + 0,013X44 + 0,013X45 + 0,012X46 + 0,012X47 + 0,020X48 + 0,038X49 + 0,021X50 + 0,019X51 + 0,512X52 + 0,851X53 - 0,004X54 + 0,001X55

*Y*2 = - 0,001X1 - 0,003X2 + 0,005X3 + 0,013X4 + 0,012X5 + 0,017X6 + 0,009X7 + 0,009X8 + 0,008X9 + 0,009X10 + 0,009X11 + 0,014X12 + 0,008X13 + 0,035X14 + 0,014X15 + 0,009X16 + 0,007X17 + 0,003X18 + 0,007X19 + 0,013X20 + 0,021X21 + 0,018X22 + 0,007X23 + 0,009X24 + 0,006X25 + 0,002X26 + 0,001X27 + 0,007X28 + 0,012X29 - 0,028X30 - 0,002X31 - 0,003X32 - 0,007X33 - 0,016X34  - 0,001X35 + 0,004X36 - 0,019X37 - 0,011X38 - 0,014X39 - 0,016X40 - 0,021X41 - 0,017X42 - 0,010X43 - 0,007X44 - 0,007X45 - 0,006X46 - 0,006X47 - 0,007X48 - 0,023X49 - 0,015X50 - 0,010X51 + 0,855X52 - 0,511X53 - 0,010X54 - 0,001X55

**Datos estandarizados.**

Las variables estudiadas serán estandarizadas, debido a que sus unidades de medidas no son conmensurables (diferentes escalas).

Así tenemos la variable edad, las notas de lenguaje y matemáticas, el sexo de los alumnos y los diferentes códigos para cada variable de las pruebas de lenguaje y matemáticas.

Puede ocurrir que variables medidas en escalas mayores (este es el caso de las variables: nota de lenguaje y nota de matemáticas, las cuales pueden ser valores entre 0 y 100) resten importancia a otras variables medidas en escalas inferiores.

Al estandarizar las variables, sus magnitudes serán del mismo orden, y Xi  (Zi) tendrá un papel importante en la construcción de las componentes.

Las componentes principales se obtienen de las variables estandarizadas:









Donde la matriz de varianzas y covarianzas muestral de las variables estandarizadas Zp es igual a la matriz de correlación muestral (vea Anexo 5) generada de la matriz de datos.

La Tabla XCII presenta los valores propios estimados asociados a la matriz de correlación:

## TABLA XCII

## Valores propios estimados asociados

**a la matriz de correlación R**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | = | 9,14467432 |  | = | 0,92637609 |  | = | 0,56773939 |
|  | = | 2,93801805 |  | = | 0,88009672 |  | = | 0,54833961 |
|  | = | 2,22390255 |  | = | 0,86737494 |  | = | 0,53743365 |
|  | = | 2,07918046 |  | = | 0,84887179 |  | = | 0,52461721 |
|  | = | 1,65176147 |  | = | 0,83675879 |  | = | 0,48559846 |
|  | = | 1,58787602 |  | = | 0,81744366 |  | = | 0,44694296 |
|  | = | 1,48690328 |  | = | 0,81083619 |  | = | 0,44252516 |
|  | = | 1,43166451 |  | = | 0,78819653 |  | = | 0,38735278 |
|  | = | 1,36158126 |  | = | 0,77807618 |  | = | 0,38089606 |
|  | = | 1,28621594 |  | = | 0,75436643 |  | = | 0,36467939 |
|  | = | 1,18391727 |  | = | 0,73605086 |  | = | 0,34056278 |
|  | = | 1,1599005 |  | = | 0,71066657 |  | = | 0,29444496 |
|  | = | 1,1328471 |  | = | 0,69439355 |  | = | 0,27064952 |
|  | = | 1,10468174 |  | = | 0,67181339 |  | = | 0,20453418 |
|  | = | 1,05305089 |  | = | 0,66194036 |  | = | 0,14973126 |
|  | = | 1,01774099 |  | = | 0,64071781 |  | = | 0,00416117 |
|  | = | 1,01244177 |  | = | 0,62332201 |  | = | 0,00243157 |
|  | = | 0,98495023 |  | = | 0,60679573 |  |  |  |
|  |  | 0,96279693 |  |  | 0,58915704 |  |  |  |

**Gráfico 4.2**

**Número de la componente Vs. Valor Propio (matriz R)**

En el gráfico 4.2, se muestran los valores propios frente a al número de su componente principal. Notaremos que a partir de la cuarta componente los valores propios son relativamente pequeños, sin embargo, si consultamos la Tabla XCII, el porcentaje de explicación acumulado para las cuatro primeras componentes principales es 29,79%, lo que no aporta mayor explicación.

El porcentaje de explicación de cada componente será igual al valor de su varianza dividido para 55 (número de variables) y por 100. Utilizando de la estandarización de las variables se tiene que las cincuenta y cinco variables se han reducido a diecisiete componentes las cuales explican el 59,74% de la varianza total.

## TABLA XCIII

## Porcentajes de explicación de las componentes principales obtenidas a partir de la matriz de correlación muestral R

|  |  |  |
| --- | --- | --- |
| **Componente Principal** | **Porcentaje de explicación** | **Porcentaje acumulado de explicación** |
| Y1 | 16,6266806 | 16,6266806 |
| Y2 | 5,34185099 | 21,9685316 |
| Y3 | 4,04345918 | 26,0119908 |
| Y4 | 3,78032811 | 29,7923189 |
| Y5 | 3,00320267 | 32,7955215 |
| Y6 | 2,8870473 | 35,6825688 |
| Y7 | 2,7034605 | 38,3860293 |
| Y8 | 2,60302637 | 40,9890557 |
| Y9 | 2,4756023 | 43,464658 |
| Y10 | 2,33857443 | 45,8032324 |
| Y11 | 2,15257685 | 47,9558093 |
| Y12 | 2,10891 | 50,0647193 |
| Y13 | 2,05972199 | 52,1244413 |
| Y14 | 2,00851226 | 54,1329535 |

## TABLA XCIII

## Porcentajes de explicación de las componentes principales obtenidas a partir de la matriz de correlación muestral R

## (continuación)

|  |  |  |
| --- | --- | --- |
| **Componente Principal** | **Porcentaje de explicación** | **Porcentaje acumulado de explicación** |
| Y15 | 1,91463798 | 56,0475915 |
| Y16 | 1,85043816 | 57,8980297 |
| Y17 | 1,84080321 | 59,7388329 |
| Y18 | 1,79081859 | 61,5296515 |
| Y19 | 1,75053988 | 63,2801914 |
| Y20 | 1,68432016 | 64,9645115 |
| Y21 | 1,60017586 | 66,5646874 |
| Y22 | 1,57704535 | 68,1417327 |
| Y23 | 1,54340326 | 69,685136 |
| Y24 | 1,52137962 | 71,2065156 |
| Y25 | 1,48626121 | 72,6927768 |
| Y26 | 1,47424762 | 74,1670244 |
| Y27 | 1,43308459 | 75,600109 |
| Y28 | 1,41468395 | 77,014793 |
| Y29 | 1,37157532 | 78,3863683 |
| Y30 | 1,33827429 | 79,7246426 |
| Y31 | 1,29212105 | 81,0167636 |
| Y32 | 1,26253372 | 82,2792974 |
| Y33 | 1,22147889 | 83,5007763 |
| Y34 | 1,20352794 | 84,7043042 |
| Y35 | 1,16494148 | 85,8692457 |
| Y36 | 1,13331274 | 87,0025584 |
| Y37 | 1,10326496 | 88,1058234 |
| Y38 | 1,07119461 | 89,177018 |
| Y39 | 1,03225343 | 90,2092714 |
| Y40 | 0,9969811 | 91,2062525 |
| Y41 | 0,97715209 | 92,1834046 |

## TABLA XCIII

## Porcentajes de explicación de las componentes principales obtenidas a partir de la matriz de correlación muestral R

**(continuación)**

|  |  |  |
| --- | --- | --- |
| **Componente Principal** | **Porcentaje de explicación** | **Porcentaje acumulado de explicación** |
| Y42 | 0,95384948 | 93,1372541 |
| Y43 | 0,88290629 | 94,0201604 |
| Y44 | 0,81262356 | 94,8327839 |
| Y45 | 0,8045912 | 95,6373751 |
| Y46 | 0,70427779 | 96,3416529 |
| Y47 | 0,69253829 | 97,0341912 |
| Y48 | 0,66305344 | 97,6972446 |
| Y49 | 0,61920506 | 98,3164497 |
| Y50 | 0,53535448 | 98,8518042 |
| Y51 | 0,49209004 | 99,3438942 |
| Y52 | 0,37188032 | 99,7157745 |
| Y53 | 0,27223866 | 99,9880132 |
| Y54 | 0,00756577 | 99,995579 |
| Y55 | 0,00442104 | 100 |

*Porcentaje de explicación de*

*la i-ésima componente principal* =  (100) i =1,2,...,p

Por lo tanto el porcentaje de explicación acumulado de las primeras 17 componentes principales es 59,74%.

## TABLA XCIV

## Las primeras diecisiete componentes principales a partir de los datos estandarizados



## TABLA XCIV

## Las primeras diecisiete componentes principales a partir de los datos estandarizados

## (Continuación)



Para formar las combinaciones lineales de las componentes principales, observemos los coeficientes en la Tabla XCIV:

Así la primera componente principal será:

*Y*1 = - 0,067Z1 + 0,031Z2 + 0,064Z3 + 0,116Z4 + 0,100Z5 + 0,142Z6 + 0,101Z7 + 0,129Z8 + 0,105Z9 + 0,141Z10 + 0,096Z11 + 0,125Z12 + 0,096Z13 + 0,167Z14 + 0,062Z15 + 0,117Z16 + 0,107Z17 + 0,066Z18 + 0,112Z19 + 0,106Z20 + 0,113Z21 + 0,150Z22 + 0,102Z23 + 0,118Z24 + 0,105Z25 + 0,056Z26 + 0,043Z27 + 0,084Z28 + 0,099Z29 + 0,130Z30 + 0,101Z31 + 0,119Z32 + 0,109Z33 + 0,130Z34 + 0,171Z35 + 0,120Z36 + 0,135Z37 + 0,165Z38 + 0,153Z39 + 0,163Z40 + 0,150Z41 + 0,152Z42 + 0,143Z43 + 0,108Z44 + 0,184Z45 + 0,189Z46 + 0,189Z47 + 0,166Z48 + 0,185Z49 + 0,168Z50 + 0,168Z51 + 0,271Z52 + 0,297Z53 - 0,068Z54 + 0,012Z55

La segunda componente principal será:

*Y*2 = - 0,027Z1 - 0,081Z2 + 0,048Z3 + 0,106Z4 + 0,107Z5 + 0,136Z6 + 0,233Z7 + 0,266Z8 + 0,151Z9 + 0,226Z10 + 0,114Z11 + 0,164Z12 + 0,073Z13 + 0,179Z14 + 0,018Z15 + 0,186Z16 + 0,154Z17 + 0,088Z18 + 0,154Z19 + 0,035Z20 + 0,046Z21 + 0,113Z22 + 0,112Z23 + 0,125Z24 + 0,099Z25 + 0,028Z26 + 0,095Z27 + 0,080Z28 + 0,073Z29 - 0,130Z30 - 0,033Z31 - 0,053Z32 - 0,104Z33 - 0,044Z34 - 0,008Z35 + 0,066Z36 - 0,059Z37 - 0,146Z38 - 0,167Z39 - 0,185Z40 - 0,150Z41 - 0,116Z42 - 0,098Z43 - 0,070Z44 - 0,164Z45 - 0,156Z46 - 0,147Z47 - 0,077Z48 - 0,152Z49 - 0,207Z50 - 0,178Z51 + 0,272Z52 - 0,224Z53 - 0,157Z54 - 0,088Z55

Para obtener las demás componentes principales seguimos el mismo patrón, observando la Tabla XCIV

Con los resultados obtenidos podemos darnos cuenta que en la primera componente las variables de nota de matemáticas y nota de lenguaje son las que más aportan. Además su interpretación podría ser difícil, pues se puede observar que las cargas de cada componente no se encuentran bien distribuidas, por lo tanto es conveniente aplicar rotación VARIMAX, a fin de conseguir una mejor agrupación de variables para su posterior rotulación.

**Rotación Varimax.**

Si aplicamos rotación Varimax a las variables originales, se construirán diecisiete componentes principales con un porcentaje de explicación de 59,74%.

## TABLA XCV

## Porcentaje de explicación de las componentes principales aplicando rotación varimax

|  |  |  |
| --- | --- | --- |
| **Componente Principal** | **Porcentaje de explicación** | **Porcentaje acumulado de explicación** |
| Y1 | 16,6266806 | 16,6266806 |
| Y2 | 5,34185099 | 21,9685316 |
| Y3 | 4,04345918 | 26,0119908 |
| Y4 | 3,78032811 | 29,7923189 |
| Y5 | 3,00320267 | 32,7955215 |
| Y6 | 2,8870473 | 35,6825688 |
| Y7 | 2,7034605 | 38,3860293 |
| Y8 | 2,60302637 | 40,9890557 |
| Y9 | 2,4756023 | 43,464658 |
| Y10 | 2,33857443 | 45,8032324 |
| Y11 | 2,15257685 | 47,9558093 |
| Y12 | 2,10891 | 50,0647193 |
| Y13 | 2,05972199 | 52,1244413 |
| Y14 | 2,00851226 | 54,1329535 |
| Y15 | 1,91463798 | 56,0475915 |
| Y16 | 1,85043816 | 57,8980297 |
| Y17 | 1,84080321 | 59,7388329 |

## TABLA XCVI

## Componentes principales obtenidas utilizando rotación VARIMAX



## TABLA XCVI

## Componentes principales obtenidas utilizando rotación VARIMAX

## (Continuación)



Para formar las combinaciones lineales de las componentes principales, observemos los coeficientes en la Tabla XCVI:

Así la primera componente principal será:

*Y*1 = - 0,052Z1 - 0,013Z2 + 0,028Z3 + 0,079Z4 + 0,028Z5 + 0,085Z6 - 0,003Z7 - 0,008Z8 + 0,106Z9 + 0,061Z10 - 0,006Z11 + 0,025Z12 + 0,066Z13 + 0,066Z14 + 0,031Z15 + 0,065Z16 + 0,003Z17 + 0,002Z18 - 0,021Z19 + 0,064Z20 + 0,105Z21 + 0,014Z22 + 0,104Z23 + 0,020Z24 - 0,024Z25 + 0,058Z26 - 0,010Z27 + 0,003Z28 + 0,020Z29 - 0,047Z30 - 0,030Z31 + 0,065Z32 + 0,024Z33 + 0,014Z34 + 0,082Z35 + 0,073Z36 + 0,074Z37 + 0,075Z38 + 0,035Z39 + 0,041Z40 + 0,009Z41 + 0,082Z42 + 0,108Z43 + 0,044Z44 + 0,480Z45 + 0,511Z46 + 0,501Z47 + 0,174Z48 + 0,091Z49 + 0,094Z50 + 0,089Z51 + 0,089Z52 + 0,235Z53 + 0,067Z54 + 0,035Z55

La segunda componente principal será:

*Y*2 = - 0,063Z1 + 0,001Z2 + 0,028Z3 + 0,024Z4 + 0,016Z5 + 0,023Z6 + 0,000Z7 + 0,017Z8 + 0,038Z9 + 0,037Z10 + 0,035Z11 + 0,020Z12 + 0,044Z13 + 0,040Z14 + 0,060Z15 + 0,016Z16 + 0,002Z17 + 0,058Z18 + 0,043Z19 + 0,027Z20 + 0,015Z21 + 0,089Z22 + 0,017Z23 - 0,039Z24 + 0,027Z25 + 0,024Z26 - 0,033Z27 + 0,118Z28 + 0,073Z29 + 0,063Z30 + 0,048Z31 + 0,247Z32 + 0,192Z33 + 0,002Z34 + 0,076Z35 + 0,007Z36 + 0,075Z37 + 0,366Z38 + 0,489Z39 + 0,482Z40 + 0,185Z41 + 0,119Z42 + 0,112Z43 + 0,062Z44 + 0,056Z45 + 0,063Z46 + 0,056Z47 + 0,170Z48 + 0,089Z49 + 0,082Z50 + 0,060Z51 + 0,105Z52 + 0,289Z53 - 0,051Z54 + 0,050Z55

Para obtener las demás componentes principales seguimos el mismo patrón, observando la Tabla XCVI.

Se procederá a rotular las primeras diez componentes principales obtenidas a través de la rotación VARIMAX. Nos fijaremos en los coeficientes de mayor peso en las componentes (Tabla XCVI).

***Primera componente principal***

Las variables que más pesan en esta componente principal son:

* X45 = *Identificación de palabras agudas.*
* X46 = *Identificación de palabras graves.*
* X47 = *Identificación de palabras esdrújulas*.

Por lo tanto esta variable puede ser rotulada como *ACENTO DE LAS PALABRAS*.

***Segunda componente principal***

Las variables que más pesan en esta componente principal son:

* X38 = *Conjugar verbo en el tiempo presente.*
* X39 = *Conjugar verbo en el tiempo pasado.*
* X40 = *Conjugar verbo en el tiempo futuro*.

A esta segunda componente principal se la puede rotular como *CONJUGACIÓN DE VERBOS*.

***Tercera componente principal***

Las variables que más pesan en esta componente principal son:

* X7 = *Suma de fracciones.*
* X8 = *Resta de fracciones.*
* X9 = *Multiplicación de fracciones*.
* X10 = *División de fracciones*.

A esta tercera componente principal se la puede rotular como *OPERACIONES CON NUMEROS FRACCIONARIOS.*

***Cuarta componente principal***

Las variables que más pesan en esta componente principal son:

* X50 = *Primera pregunta analítica.*
* X51 = *Segunda pregunta analítica.*

A esta cuarta componente principal se la puede rotular como *EXPRESIÓN Y CAPACIDAD DE ANÁLISIS.*

***Quinta componente principal***

Las variables que más pesan en esta componente principal son:

* X34 = *Partes de la oración 1.*
* X35 = *Partes de la oración 2.*
* X36 = *Partes de la oración 3.*
* X37 = *Partes de la oración 4.*

A esta quinta componente principal se la puede rotular como *NOCIONES GENERALES DE ORACION.*

***Sexta componente principal***

Las variables que más pesan en esta componente principal son:

* X25 = *Unión de conjuntos.*
* X26 = *Intersección de conjuntos.*
* X29 = *Pintar intersección de conjunto.*

A esta sexta componente principal se la puede rotular como *OPERACIONES CON CONJUNTOS*.

***Séptima componente principal***

Las variables que más pesan en esta componente principal son:

* X16 = *Medida de longitud.*
* X17 = *Medida de peso.*
* X18 = *Medida de capacidad.*
* X19 = *Medida de tiempo.*

A esta séptima componente principal se la puede rotular como *SISTEMA DE MEDIDAS*.

***Octava componente principal***

Las variables que más pesan en esta componente principal son:

* X3 = *Suma de enteros.*
* X4 = *Resta de enteros.*
* X5 = *Multiplicación de enteros.*
* X6 = *División de enetros.*

A esta octava componente principal se la puede rotular como *OPERACIONES BASICAS CON NUMEROS ENTEROS*.

***Novena componente principal***

Las variables que más pesan en esta componente principal son:

* X30 = *Sustantivo propio y común.*
* X31 = *Palabras sinónimas.*

A esta novena componente principal se la puede rotular como *CLASES DE SUSTANTIVOS*.

***Décima componente principal***

Las variables que más pesan en esta componente principal son:

* X11 = *Suma de decimales.*
* X12 = *Resta de decimales.*
* X13 = *Multiplicación de decimales.*

A esta décima componente principal se la puede rotular como *OPERACIONES BASICAS CON NUMEROS DECIMALES*.

* 1. **Análisis de Varianza**

El objetivo del análisis de varianza es identificar variables importantes en un estudio y determinar como interactúan y afectan a la respuesta.

Se realizó una prueba de bondad de ajuste, Kolmogorov-Smirnov para determinar si la variable *NOTA DE LENGUAJE*  tiene una distribución normal.

H0 : La muestra ha sido tomada de una población normal con media 53,8 y desviación estándar 19.

Vs.

H1 : Niega H0.

El valor p para esta prueba es 0,209, lo que nos permite decir que existe suficiente evidencia estadística para considerar que la muestra ha sido tomada de una población normal con media 53,8 y una desviación estándar 19.

Procederemos al respectivo análisis de varianza utilizando la variable cuantitativa *NOTA DE LENGUAJE* como variable dependiente.

Para lo cual se construye el modelo:

yijkl =  + i + j + k + ()ij + ()ik + ()jk + ()ijk + ijkl

Donde: ijkl N (0,2)

i=1,2

j=1,2,3

k=1,2,3

l=1,2,…,n

i representa el efecto del i-ésimo tratamiento del factor A: Sexo, el cual tiene dos niveles posibles.

1. : Hombres
2. : Mujeres

j representa el efecto del j-ésimo tratamiento del factor B: Jornada, el cual tiene tres posible niveles:

0 : Matutina

1 : Vespertina

2 : Nocturna

k representa el efecto del k-ésimo tratamiento del factor C: Actividad extra educativa, el cual tiene los siguientes niveles:

0 : Estudiar, jugar, hacer deberes

1 : Trabaja

2 : Quehaceres domésticos

Las hipótesis para este modelo factorial son:















El modelo factorial para el análisis de varianza utilizando la variable cuantitativa *NOTA DE LENGUAJE* como variable dependiente es:

i=1,2

yijkl =  + i + j + k + ()ij + ()ik + ()jk + ()ijk + ijkl j=1,2,3

k=1,2,3

l=1,2,…,n

Donde: ijkl N (0,2)

**TABLA XCVII**

**Análisis de varianza para la nota de lenguaje**



Si observamos la Tabla XCVII nos daremos cuenta que el valor p para el factor Jornada(B) es 0,010, por lo tanto se rechaza la hipótesis nula para este factor, es decir, el efecto sobre la calificación de lenguaje del estudiante de al menos de una de las jornadas es diferente de cero. En el caso de los factores de Sexo (A) y Actividades extra educativas (B) así como de sus interacciones se acepta la hipótesis nula, es decir no existirá influencia alguna en el rendimiento del alumno en lo referente al área de lenguaje así realice alguna actividad extra educativa o el hecho de que sea hombre o mujer.

* 1. **Correlación Canónica.**

A continuación se presenta el análisis de correlación canónica donde intervienen dos grupos de variables. En el primer grupo constarán las p=22 variables que miden el nivel de conocimiento en lenguaje de los 913 alumnos de séptimo año de educación básica de las escuelas fiscales urbanas del cantón Guayaquil que pertenecen a la muestra, es representado por un vector aleatorio de (p x 1) X(1). En el segundo grupo constarán las q=27 variables que miden el nivel de conocimiento en matemáticas, representado por un vector aleatorio de (q x 1) X(2).

Se crearan combinaciones lineales con cada grupo de variables de tal manera que estén correlacionadas entre sí. Cada variable artificial creada a partir de las variables observables es denominada como variable canónica.

El número de correlaciones canónicas muestrales será el min(p,q) = min(22,27) = 22, las cuales se presentarán en la Tabla XCV.

**TABLA XCVIII**

**Correlaciones Canónicas entre los conjuntos**

**de variables de lenguaje y matemáticas**



El primer par de variable canónicas muestrales tienen una correlación canónica de 0,587. El segundo par de variables canónicas tienen una correlación canónica de 0,346.

La Tabla XCIX muestra los coeficientes de U1 y U2, las que representan las dos primeras variables canónicas para lenguaje.

**TABLA XCIX**

**Coeficientes de las dos primeras**

**variables canónicas de lenguaje**

|  |  |  |
| --- | --- | --- |
| **Variables** | **U1** | **U2** |
| X30 | -0,066 | 0,201 |
| X31 | -0,092 | -0,228 |
| X32 | -0,167 | 0,082 |
| X33 | -0,037 | 0,235 |
| X34 | 0,021 | -0,002 |
| X35 | -0,252 | -0,314 |
| X36 | -0,195 | 0,666 |
| X37 | -0,018 | -0,18 |
| X38 | -0,113 | -0,387 |
| X39 | -0,061 | 0,329 |
| X40 | 0,003 | -0,154 |
| X41 | -0,075 | -0,265 |
| X42 | -0,071 | 0,016 |
| X43 | -0,075 | -0,388 |
| X44 | -0,015 | 0,242 |
| X45 | -0,059 | 0,243 |
| X46 | -0,075 | -0,202 |
| X47 | -0,154 | 0,164 |
| X48 | -0,225 | 0,182 |
| X49 | -0,216 | 0,007 |
| X50 | 0,111 | -0,099 |
| X51 | -0,075 | -0,055 |

La Tabla C muestra los coeficientes de V1 y V2, las dos primeras variables canónicas para matemáticas.

**TABLA C**

**Coeficientes de las dos primeras**

**variables canónicas de matemáticas**

|  |  |  |
| --- | --- | --- |
| **Variables** | **V1** | **V2** |
| X3 | -0,135 | -0,193 |
| X4 | -0,123 | -0,334 |
| X5 | -0,071 | -0,093 |
| X6 | -0,131 | -0,149 |
| X7 | -0,016 | 0,127 |
| X8 | -0,026 | 0,205 |
| X9 | -0,063 | -0,125 |
| X10 | -0,132 | -0,103 |
| X11 | -0,094 | -0,134 |
| X12 | -0,064 | 0,008 |
| X13 | -0,106 | 0,008 |
| X14 | -0,311 | -0,139 |
| X15 | -0,079 | 0,112 |
| X16 | -0,022 | 0,343 |
| X17 | 0,084 | 0,218 |
| X18 | -0,092 | 0,117 |
| X19 | -0,097 | -0,087 |
| X20 | -0,171 | -0,088 |
| X21 | -0,078 | 0,279 |
| X22 | -0,206 | 0,185 |
| X23 | -0,129 | 0,52 |
| X24 | -0,064 | -0,105 |
| X25 | -0,018 | 0,016 |
| X26 | -0,078 | -0,115 |
| X27 | 0,035 | -0,126 |
| X28 | -0,032 | -0,097 |
| X29 | -0,145 | -0,419 |

Así pues:

***Primer par de variables canónicas.***





con correlación canónica 

 está primada por las variables:

X35 *Partes de la segunda oración*

X48 *Signos de puntuación*

X49 *Lectura comprensiva*

 está primada por:

X14 *Perímetro área de un rectángulo*

X22 *Regla de tres*

***Segundo par de variables canónicas.***





 está primada por las variables:

X43 *Primer par de palabras homófonas*

X38 *Conjugar verbo en el tiempo presente*

 está representada por:

X29 *Pintar intersección de conjuntos*

X16 *Equivalencias de medidas de longitud*

Este último par de variables canónicas tienen una correlación canónica .

Si desea conocer los coeficientes de el resto de variables canónicas, consulte el Anexo 6.