

EJERCICIO 1
 EJERCICIO 2
 EJERCICIO 3
 EJERCICIO 4
 EJERCICIO 5
 EJERCICIO 6

Examen II Estadística Multivariada

Code ▾

Hairo Ulises Miranda Belmonte

30 de Mayo del 2019

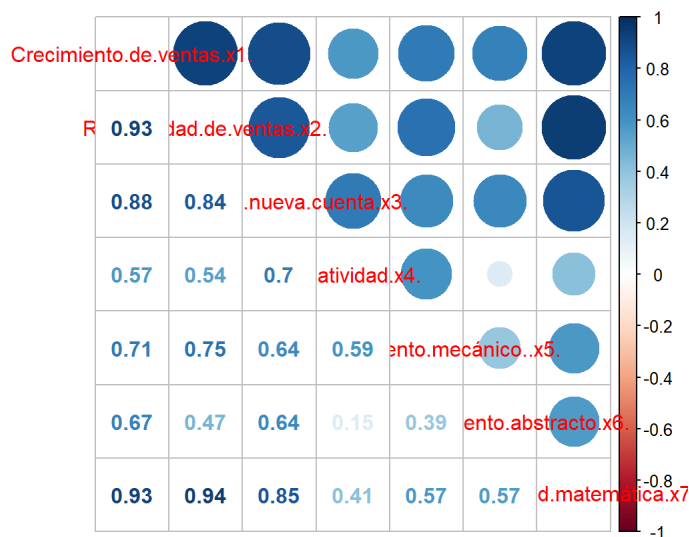
EJERCICIO 1

(a) Asumiendo un modelo ortogonal de factores para las variables estandarizadas. Obten la solución por máxima verosimilitud de L y ϕ para $m = 2$ y $m = 3$ factores, considerando una rotación varimax, e interpreta las soluciones con $m = 2$ y $m = 3$ factores.

Code

Antes de realizar factor checar correlación

Code



sí, hay correlación, procedemos al análisis de factores.

Se evalúa si se puede utilizar 2 y 3 factores.

Con dos factores; sí, se puede

Code

```
## [1] TRUE
```

Con tres factores; sí, se puede

Code

```
## [1] TRUE
```

Con m=2

Se obtiene una solución con MV y considerando una rotación varimax.

Code

Se observan las cargas de los factores (coeficientes)

Code

```
##
## Loadings:
##               Factor1 Factor2
## Crecimiento.de.ventas.x1.  0.852  0.452
## Rentabilidad.de.ventas.x2.  0.868  0.419
## Ventas.de.nueva.cuenta.x3.  0.717  0.602
## Creatividad.x4.           0.148  0.987
## Razonamiento.mecánico..x5.  0.501  0.525
## Razonamiento.abstracto.x6.  0.619
## Capacidad.matemática.x7.   0.946  0.277
##
##               Factor1 Factor2
## SS loadings      3.545  2.071
## Proportion Var   0.506  0.296
## Cumulative Var   0.506  0.802
```

El factor uno se encuentra representado por el crecimiento de ventas, rentabilidad en las ventas, y las capacidades matemáticas. Este factor puede indicar la efectividad que el vendedor tiene con el uso de su capacidad matematica, o aquellos vendedores que utilizan herramientas matemáticas. para mejorar su desempeño.

El segundo factor se representa por la creatividad, que tan bueno es el vendedor para generar ventas, que es algo más cualitativo en contraste a los del primer factor.

Con m=3

Se obtiene una solución con MV y considerando una rotación varimax.

Code

Se observan las cargas de los factores (coeficientes)

Code

```
##
## Loadings:
##               Factor1 Factor2 Factor3
## Crecimiento.de.ventas.x1.  0.793  0.374  0.438
## Rentabilidad.de.ventas.x2.  0.911  0.317  0.185
## Ventas.de.nueva.cuenta.x3.  0.651  0.544  0.438
## Creatividad.x4.           0.255  0.964
## Razonamiento.mecánico..x5.  0.542  0.465  0.207
## Razonamiento.abstracto.x6.  0.299          0.950
## Capacidad.matemática.x7.   0.917  0.180  0.298
##
##               Factor1 Factor2 Factor3
## SS loadings      3.175  1.718  1.453
## Proportion Var   0.454  0.245  0.208
## Cumulative Var   0.454  0.699  0.906
```

El primer factor se compone de la rentabilidad de ventas, crecimiento de ventas y habilidades matemáticas., que se interpreta como se hizo en el caso de dos factores.

El segundo factor es el mismo que en el caso anterior, la creatividad del vendedor.

El tercer factor lo componen los vendedores que tienen mucha capacidad de abstracción, lo cual no tiene una relación con aquellos que tienen buen razonamiento matemático, o se liga a su rendimiento en las ventas.

(b) A partir de las estimaciones de los parametros las comunialidades, las varianzas espec para las soluciones en m = 2 y m = 3 factores. Compara los resultados. Qué elección de m prefieres en este punto?

con m=2 Ahora mostramos los siguiente:

Code

Varianza Especifica:

Code

```
##                                [,1]
## Crecimiento.de.ventas.x1.  0.06919160
## Rentabilidad.de.ventas.x2. 0.07038038
## Ventas.de.nueva.cuenta.x3. 0.12330883
## Creatividad.x4.           0.00500000
## Razonamiento.mecánico..x5. 0.47358490
## Razonamiento.abstracto.x6. 0.61363862
## Capacidad.matemática.x7.   0.02881701
```

Comunalidades:

Code

```
##                                [,1]
## Crecimiento.de.ventas.x1.  0.9308083
## Rentabilidad.de.ventas.x2. 0.9296182
## Ventas.de.nueva.cuenta.x3. 0.8766896
## Creatividad.x4.           0.9950121
## Razonamiento.mecánico..x5. 0.5264121
## Razonamiento.abstracto.x6. 0.3863622
## Capacidad.matemática.x7.   0.9711829
```

Se calcula el residual de la aproximación a la matriz de correlación.

Code

```

## Crecimiento.de.ventas.x1.
## Crecimiento.de.ventas.x1. 0.000
## Rentabilidad.de.ventas.x2. -0.003
## Ventas.de.nueva.cuenta.x3. 0.001
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. 0.044
## Razonamiento.abstracto.x6. 0.120
## Capacidad.matemática.x7. -0.004
## Rentabilidad.de.ventas.x2.
## Crecimiento.de.ventas.x1. -0.003
## Rentabilidad.de.ventas.x2. 0.000
## Ventas.de.nueva.cuenta.x3. -0.032
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. 0.091
## Razonamiento.abstracto.x6. -0.097
## Capacidad.matemática.x7. 0.007
## Ventas.de.nueva.cuenta.x3. Creatividad.x4.
## Crecimiento.de.ventas.x1. 0.001 0.000
## Rentabilidad.de.ventas.x2. -0.032 0.000
## Ventas.de.nueva.cuenta.x3. 0.000 0.001
## Creatividad.x4. 0.001 0.000
## Razonamiento.mecánico..x5. -0.038 -0.001
## Razonamiento.abstracto.x6. 0.161 -0.004
## Capacidad.matemática.x7. 0.008 0.000
## Razonamiento.mecánico..x5.
## Crecimiento.de.ventas.x1. 0.044
## Rentabilidad.de.ventas.x2. 0.091
## Ventas.de.nueva.cuenta.x3. -0.038
## Creatividad.x4. -0.001
## Razonamiento.mecánico..x5. 0.000
## Razonamiento.abstracto.x6. 0.045
## Capacidad.matemática.x7. -0.044
## Razonamiento.abstracto.x6.
## Crecimiento.de.ventas.x1. 0.120
## Rentabilidad.de.ventas.x2. -0.097
## Ventas.de.nueva.cuenta.x3. 0.161
## Creatividad.x4. -0.004
## Razonamiento.mecánico..x5. 0.045
## Razonamiento.abstracto.x6. 0.000
## Capacidad.matemática.x7. -0.035
## Capacidad.matemática.x7.
## Crecimiento.de.ventas.x1. -0.004
## Rentabilidad.de.ventas.x2. 0.007
## Ventas.de.nueva.cuenta.x3. 0.008
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. -0.044
## Razonamiento.abstracto.x6. -0.035
## Capacidad.matemática.x7. 0.000

```

Se observan valores pequeños, y alguno que otro con valor cero, dando a entender una buena aproximación a la matriz de correlación.

Se calcula la norma uno para la matriz de residuales y contrastar contra 3 factores.

Code

```
## [1] 0.19
```

con m=3

A continuación seguimos con el análisis de factores

Code

Varianza Especifica:

Code

```
##                                [,1]
## Crecimiento.de.ventas.x1.  0.03857165
## Rentabilidad.de.ventas.x2. 0.03448071
## Ventas.de.nueva.cuenta.x3. 0.08812176
## Creatividad.x4.            0.00500000
## Razonamiento.mecánico..x5. 0.44662048
## Razonamiento.abstracto.x6. 0.00500000
## Capacidad.matemática.x7.   0.03750980
```

Comunalidades:

Code

```
##                                [,1]
## Crecimiento.de.ventas.x1.  0.9614284
## Rentabilidad.de.ventas.x2. 0.9655192
## Ventas.de.nueva.cuenta.x3. 0.9118756
## Creatividad.x4.            0.9950434
## Razonamiento.mecánico..x5. 0.5533820
## Razonamiento.abstracto.x6. 0.9950317
## Capacidad.matemática.x7.   0.9624901
```

Se calcula el residual de la aproximación a la matriz de correlación.

Code

```

## Crecimiento.de.ventas.x1.
## Crecimiento.de.ventas.x1. 0.000
## Rentabilidad.de.ventas.x2. 0.003
## Ventas.de.nueva.cuenta.x3. -0.028
## Creatividad.x4. 0.001
## Razonamiento.mecánico..x5. 0.013
## Razonamiento.abstracto.x6. 0.001
## Capacidad.matemática.x7. 0.002
## Rentabilidad.de.ventas.x2.
## Crecimiento.de.ventas.x1. 0.003
## Rentabilidad.de.ventas.x2. 0.000
## Ventas.de.nueva.cuenta.x3. -0.005
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. 0.066
## Razonamiento.abstracto.x6. 0.000
## Capacidad.matemática.x7. -0.004
## Ventas.de.nueva.cuenta.x3. Creatividad.x4.
## Crecimiento.de.ventas.x1. -0.028 0.001
## Rentabilidad.de.ventas.x2. -0.005 0.000
## Ventas.de.nueva.cuenta.x3. 0.000 0.001
## Creatividad.x4. 0.001 0.000
## Razonamiento.mecánico..x5. -0.059 0.000
## Razonamiento.abstracto.x6. 0.001 0.000
## Capacidad.matemática.x7. 0.027 0.000
## Razonamiento.mecánico..x5.
## Crecimiento.de.ventas.x1. 0.013
## Rentabilidad.de.ventas.x2. 0.066
## Ventas.de.nueva.cuenta.x3. -0.059
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. 0.000
## Razonamiento.abstracto.x6. 0.002
## Capacidad.matemática.x7. -0.068
## Razonamiento.abstracto.x6.
## Crecimiento.de.ventas.x1. 0.001
## Rentabilidad.de.ventas.x2. 0.000
## Ventas.de.nueva.cuenta.x3. 0.001
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. 0.002
## Razonamiento.abstracto.x6. 0.000
## Capacidad.matemática.x7. -0.001
## Capacidad.matemática.x7.
## Crecimiento.de.ventas.x1. 0.002
## Rentabilidad.de.ventas.x2. -0.004
## Ventas.de.nueva.cuenta.x3. 0.027
## Creatividad.x4. 0.000
## Razonamiento.mecánico..x5. -0.068
## Razonamiento.abstracto.x6. -0.001
## Capacidad.matemática.x7. 0.000

```

Se calcula la norma uno para la matriz de residuales y contrastar contra 3 factores.

Code

```
## [1] 0.06
```

En conclusión si tomamos la norma mínima la matriz de correlación se aproxima mejor si se utilizan utilizan 3 factores

(c) realiza una prueba para $m = 2$ y $m = 3$. A partir de estos resultados y de la parte b), que elección de m parece ser la adecuada?

Code

Code

```
## $estadistico
## [1] 117.3065
##
## $gl
## [1] 8
##
## $critico
## [1] 20.09024
##
## $Rehaza
## [1] TRUE
```

Code

```
## $estadistico
## [1] 62.83714
##
## $gl
## [1] 3
##
## $critico
## [1] 11.34487
##
## $Rehaza
## [1] TRUE
```

De acuerdo con la prueba se necesita un m mayor a tres a lo menos, y utilizando la norma para la matriz de los residuales, se toma $m=3$.

(d) De acuerdo al número de factores elegido en c), calcula las puntuaciones de los factores (factor scores) para los vendedores mediante: i) mínimos cuadrados ponderados y ii) mediante el enfoque de regresión. ¿Existe algún patrón de agrupamiento de los vendedores de acuerdo a sus puntuaciones factoriales?, si es así ¿cómo se caracterizan los vendedores de cada grupo, de acuerdo a la interpretación de los factores?

Con tres Factores

NOTA: las siguientes gráficas están realizadas en plotly, el usuario puede interactuar dando rotaciones al plot, lo que se describe es de acuerdo a alguna rotación que se encuentra y se realiza una captura de pantalla.

(Dar click si aparecen en blanco)

Code

mediante el enfoque de regresión.

Code

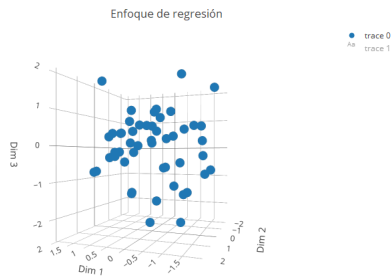
Enfoque de regresión

● trace 0
Aa trace 1

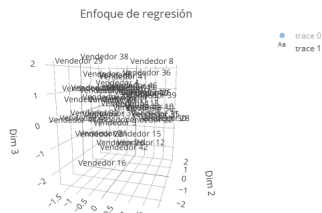
Code

Utilizando el enfoque regresión se observa la separación del espacio en dos, un grupo en la parte superior y otro en la parte inferior, inclusive se detectan, que aquellos que parecen outliers caen o tienen más peso sobre el tercer factor, siendo posiblemente aquellos que son mejores en abstracción, el grupo de abajo tienen pesos más fuerte en el primer factor, siendo aquellos con buen desempeño en el crecimiento y rentabilidad, cuya capacidad matemáticas sobre sale de los demás; el grupo de enmedio cae sobre el factor donde las habilidades de creatividad sobre salen.

Code



Code



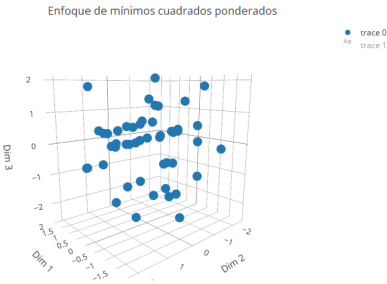
mediante el enfoque de mínimos cuadrados ponderados

Code

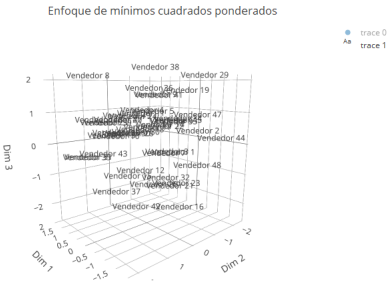
Enfoque de mínimos cuadrados ponderados

trace 0
Aa
trace 1

Code



Code



EJERCICIO 2

En el archivo SitiosArqueologicos se muestran las “distancias” entre sitios arqueol de diferentes periodos, basados en las frecuencias de ciertos tipos de alfar encontrados en esos sitios. En la tabla de distancias, el n despu de ”, indica la fecha probable en la que se reali la cer. por ejemplo P198_0918 se refiere al sitio P198 y 0918 al año posible de elaboración (d.c).

Al utilizar el enfoque de mínimos cuadrados ponderados, la representación en general es la misma que el enfoque regresión; sin embargo, la separación de los posibles grupos que caen en los factores es clara que el del enfoque por regresión.

- a. Considerando estas distancias como una medida de disimilaridad entre los sitios, aplica escalamiento multidimensional (MDS) bajo el enfoque de mínimos cuadrados y registra el STRESS1 de las soluciones obtenidas en un rango de 2 a 5 dimensiones. Utiliza como configuración inicial del algoritmo la solución Torgerson o solución clasica, considerando 200 iteraciones y un criterio de convergencia de 1e-06.

Code

Se realiza una función que nos de el valor del STRESS1 y se presenta la tabla con las dimensiones y el valor stress correspondiente.

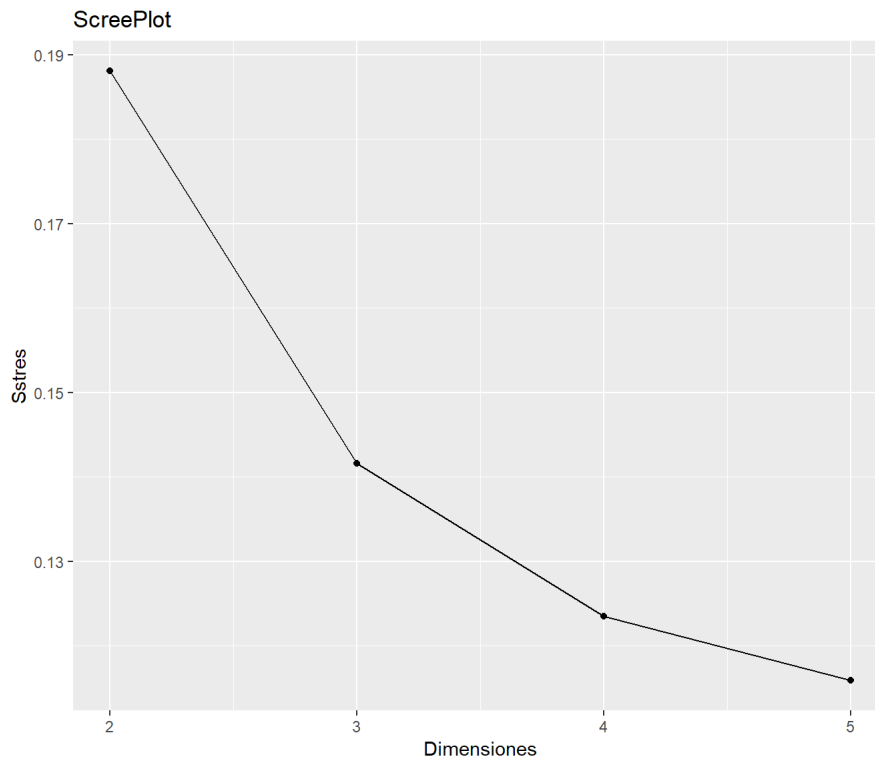
Code

| Dimensiones | Stress1 |
|-------------|-----------|
| 2 | 0.1880604 |
| 3 | 0.1415703 |
| 4 | 0.1234701 |
| 5 | 0.1159110 |

(b) Identifica la dimensionalidad adecuada de la representación MDS de los sitios graficando el STRESS versus cada número de dimensiones (gráfico de sedimentación o de codo).

Se gráfica el screplot, donde el eje de las x, representan las dimensiones que se seleccionan para el MDS, y el de las ordenadas nos indica el nivel de sstres.

Code



Se puede observar que en la dimensión tres, se observa la figura del codo; por lo tanto, utilizar tres dimensiones para la representación son buenas.

(c) Considerando la representación MDS en dos dimensiones, y tomando en cuenta los periodos asociados con los sitios, interpreta esta configuración de dos dimensiones

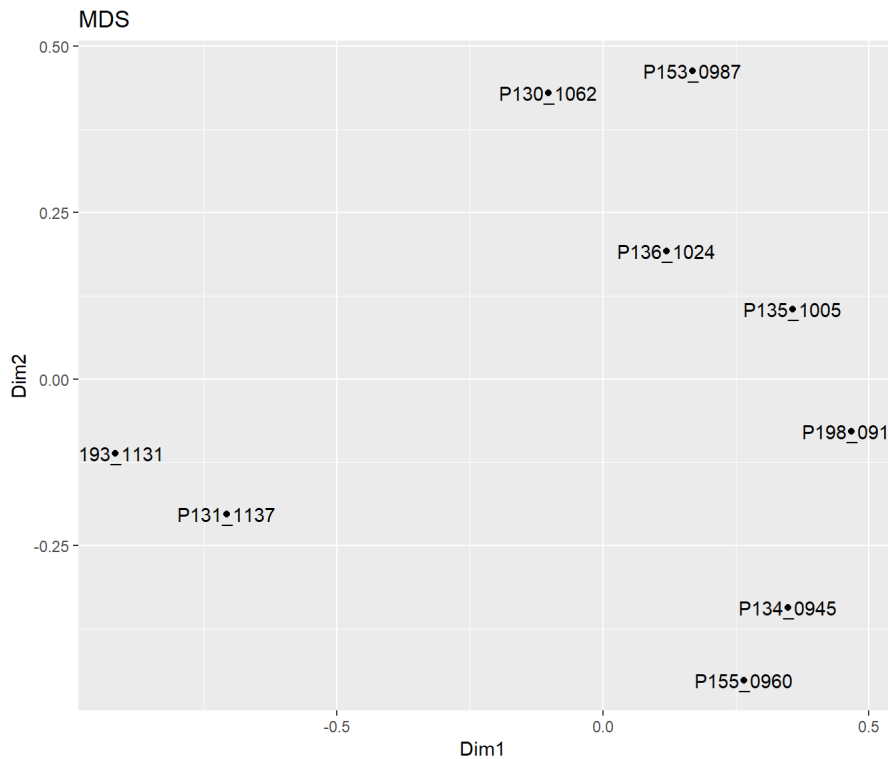
Se estima MDS con 3 dimensiones de acuerdo al resultado anterior y se visualiza en dos dimensiones, y se toma en cuenta los periodos asociados con los sitios.

[Hide](#)

```
resul_mds_mc <- smacofSym(disimilaridades, ndim=3, init = "torgerson", itmax = 200, eps = 1e-06)
config_nations_mc<-resul_mds_mc$conf

nombres <- obs[,1] %>% as.character()

config_nations_mc %>% as.data.frame %>% ggplot() + aes(x=config_nations_mc[,1], y=config_nations_mc[,2],label=nombres)
+ geom_point() + labs(title = "MDS",x="Dim1",y="Dim2") +
  geom_text(aes(x=config_nations_mc[,1], y=config_nations_mc[,2]))
```



Observando la representación se tiene que la alfarería del sitio P136 y P135 de los años 1024 y 1005 d.c, se encuentra relacionadas: también el tipo de alfarería en los sitios P155, P134 y P198, de los años respectivos, 960, 945 y 918 d.c.; asimismo la relación con los sitios P130 y P153, de los años 1062 y 987 d.c.

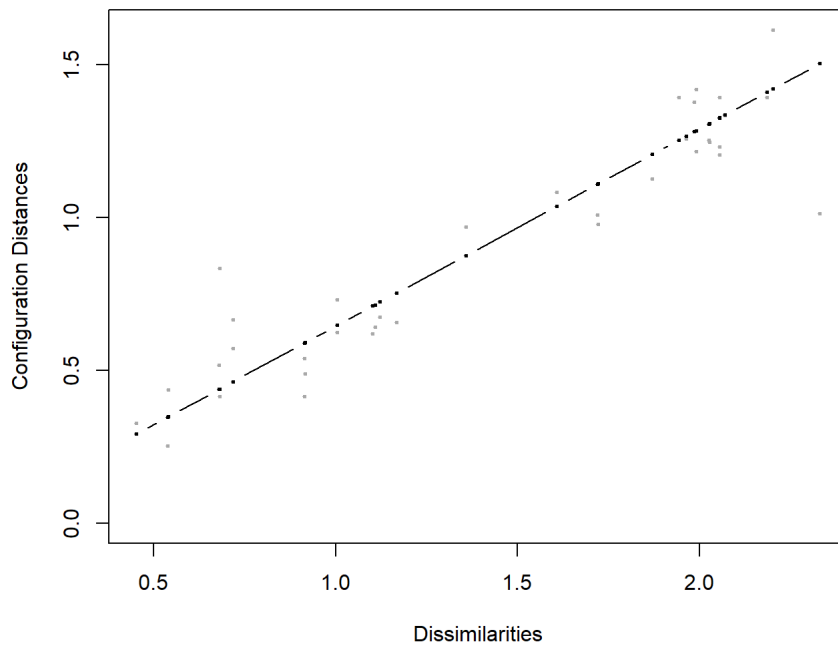
Por otro lado se tiene, la diferencia notable entre las alfarerías de los sitios P193 y P131, de los años 1131 y 1137 d.c, respecto a los otros.

De esta manera, el gráfico muestra los tipos de alfarerías encontrados en los distintos sitios, por periodo en el tiempo, donde se observa que la alfarería del siglo X (1 de enero de 901 al 31 de diciembre de 1000) conocido como el siglo del hierro, se asemejaba un poco a la del siglo XI (comienza el 1 de enero de 1001 y termina el 31 de diciembre de 1100), conocido como siglo de las cruzadas; sin embargo difiere notablemente de aquella del siglo XII (comienza el 1 de enero de 1101 y termina el 31 de diciembre de 1200) fechas donde se presentaron la segunda y tercera cruzada y consigo cambios en aspectos culturales en Europa.

O de manera general, aquella alfarería después del siglo XI, se diferencia notablemente de aquella previa a este siglo.

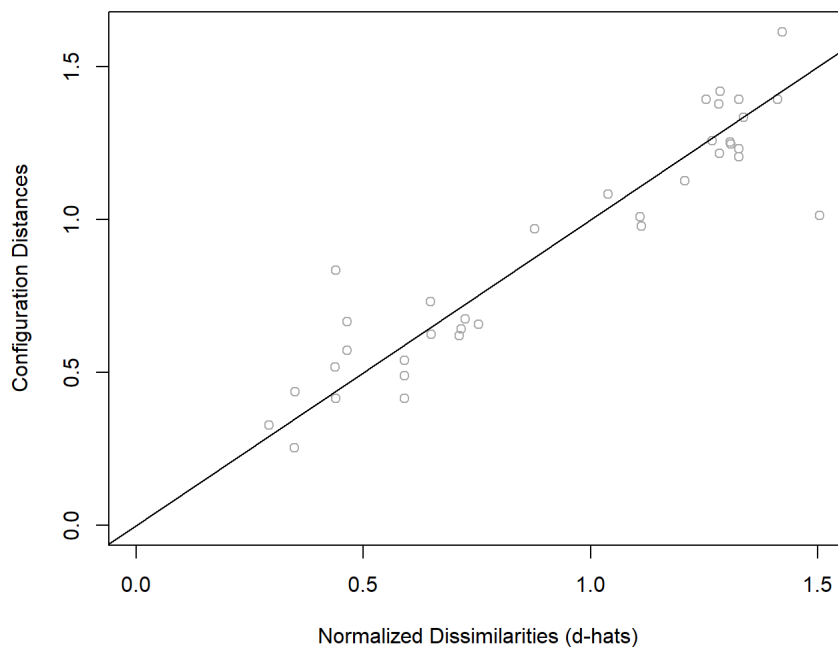
Dejando de lado la interpretación, ahora observamos el ajuste por las tres dimensiones por medio del diagrama de Shepard.

Code

Diagrama de Shepard

Se observa que la configuraciones van sobre la linea.

Gráficandola diparidad vs la distancia ajustada se puede notar un buen ajuste a la linea de 45 grado. Por lo tanto con 3 dimensiones se observa una buena representación de los datos por medio de las disimilaridades.

[Code](#)
Grafica de las disparidades vs las distancias ajustadas

EJERCICIO 3

3. Otra forma de derivar los resultados del análisis de correspondencia simple es encontrando una matriz \hat{P} de dimensión con rango reducido $t < \min(r, s)$ que aproxime P minimizando el criterio de mínimos cuadrados ponderados:

$$tr(D_r^{-\frac{1}{2}}(P - \hat{P})D_c^{-1}(P - \hat{P})D_r^{-\frac{1}{2}})$$

Usando el teorema de Eckart-Young, encuentre la matrix \hat{P} que arroje la mejor aproximación de rango reducido de P en este sentido. Muestre que la mejor aproximación de rango 1 de P es la solución trivial $\hat{P} = rc'$.

Se considera la versión escalada $B = D_r^{-1} P D_c^{-1}$ de la matrix P .

Sabemos que la mejor aproximación de \hat{B} a $D_r^{-1} P D_c^{-1}$ se encuentra dada por los primeros s valores de la descomposición SVD.

$$D_r^{-1} P D_c^{-1} = \sum_{k=1}^J \lambda_k u_k v_k$$

donde:

$$D_r^{-1} P D_c^{-1} v_k = \lambda_k u_k$$

$$u_k' D_r^{-1} P D_c^{-1} = \lambda_k v_k$$

y

$$[(D_r^{-1} P D_c^{-1})(D_r^{-1} P D_c^{-1}) - \lambda_k I] = 0$$

entonces la aproximación de P se encuentra dada por

$$\hat{P} = D_r^{-1} \hat{B} D_c^{-1} = \sum_{k=1}^J \lambda_k (D_r^{-1} u_k)(D_c^{-1} v_k)$$

Ahora, cualquier sea el valor de P , el termino rc' , provee la mejor aproximación de rango. Para esto, sea $u_k = D_r^{-1} 1$ y $v_k = D_c^{-1} 1$, y con basea lo anterior se tiene:

$$u_k' (D_r^{-1} P D_c^{-1}) = (D_r^{-1} 1)' (D_r^{-1} P D_c^{-1}) = 1 P D_c^{-1} = c' D_c^{-1} = (D_c^{-1} 1)' = v_k'$$

por otro lado

$$(D_r^{-1} P D_c^{-1}) v_k = (D_r^{-1} P D_c^{-1}) (D_c^{-1} 1) = D_c^{-1} P 1 = D_r^{-1} r = (D_r^{-1} 1)' = u_k$$

esto es

$$(u_k, v_k) = (D_r^{-1} 1, D_c^{-1} 1)$$

los cuales son los vectores singulares asociados con el valor singular $\lambda_k = 1$, y por lo tanto para toda matrix de correspondencia, el termino común en su expansión es

$$D_r^{-1} u_k v_k' D_c^{-1} 1 = rc'$$

lo que da como resultado que la aproximación

$$D_r^{-1} P D_c^{-1} = \sum_{k=1}^J \lambda_k u_k v_k$$

se da por

$$P = rc' + \sum_{k=2}^J \lambda_k (D_r^{-1} u_k)(D_c^{-1} v_k)'$$

y pasando el termino rc' al otro lado de la ecuación se tiene que la mejor aproximación de $D_r^{-1} (P - rc) D_c^{-1}$, se da por $\sum_{k=1}^J \lambda_k u_k v_k$. La mejor aproximación de $(P - rc)$ se da por:

$$P - rc' = \sum_{k=2}^J \lambda_k (D_r^{-1} u_k)(D_c^{-1} v_k)'$$

EJERCICIO 4

- Supongamos que queremos evaluar el efecto de las especies de robles en los vinos tinto rojos envejecidos en barriles. Primero, se añejo el vino de una misma cosecha de Pinot Noir en seis barriles diferentes hechos con dos tipos de roble. Los vinos 1, 5 y 6 se envejecieron con el primer tipo de roble, mientras que los vinos 2, 3 y 4 se envejecieron con el segundo. Después, se le a tres expertos en vinos que eligieran de dos a cinco variables para describir los vinos. Para cada vino y para cada variable, se pidió al experto que calificara la intensidad. La respuesta dada por el experto se codificó como una respuesta binaria (es decir, frutal frente a no afrutado) o como una respuesta ternaria (es decir, no vainilla, un poco de vainilla, claro olor a vainilla). Cada respuesta binaria está representada por 2 columnas binarias (por ejemplo, la respuesta afrutado está representada por el 1 o 0 y no afrutado es 0 1). Una respuesta ternaria representada por 3 columnas binarias (es decir, la respuesta algo de vainilla representada por el 0 1 0). Los resultados se presentan en la table inferior.

El objetivo es doble. Primero queremos obtener una tipo de los vinos y segundo queremos saber si existe un acuerdo entre las escalas utilizadas por los expertos. Usaremos el tipo de roble como una variable complementaria (o ilustrativa) que se proyecta en el análisis de correspondencia múltiple. Además, después de realizar la prueba de los seis vinos, los catadores encontraron y probaron una botella desconocida de Pinot Noir. Este vino será utilizado como una observación complementaria. Para este vino, cuando un experto no estaba seguro de usar un descriptor, se utilizó un pat de respuesta .5 .5 para representar su valor.

En este caso MCA codifica los datos creando varias columnas binarias para cada variable con la restricción de que una y solo una de las columnas obtiene el valor 1. Este esquema de codificación crea dimensiones adicionales artificiales ya que una variable categórica se codifica con varias columnas.

Aplique MCA y responda las siguientes preguntas:

Code

| | E1_fruity_y | E1_fruity_n | E1_woody_a | E1_woody_b | E1_woody_c | E1_coffe_y | E1_coffe_n | E2_fruit_y | E2_fruit_n |
|----|-------------|-------------|------------|------------|------------|------------|------------|------------|------------|
| w1 | 1 | 0 | 0 | 0 | 1 | 0.0 | 1.0 | 1 | 0 |
| w2 | 0 | 1 | 0 | 1 | 0 | 1.0 | 0.0 | 0 | 1 |
| w3 | 0 | 1 | 1 | 0 | 0 | 1.0 | 0.0 | 0 | 1 |
| w4 | 0 | 1 | 1 | 0 | 0 | 1.0 | 0.0 | 0 | 1 |
| w5 | 1 | 0 | 0 | 0 | 1 | 0.0 | 1.0 | 1 | 0 |
| w6 | 1 | 0 | 0 | 1 | 0 | 0.0 | 1.0 | 1 | 0 |
| w? | 0 | 1 | 0 | 1 | 0 | 0.5 | 0.5 | 1 | 0 |

Code

| | E1_fruity_y | E1_fruity_n | E1_woody_a | E1_woody_b | E1_woody_c | E1_coffe_y | E1_coffe_n | E2_fruit_y | E2_fruit_n |
|------|-------------|-------------|------------|------------|------------|------------|------------|------------|------------|
| w1 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0131579 | 0.0131579 | 0.0000000 |
| w2 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 |
| w3 | 0.0000000 | 0.0131579 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 |
| w4 | 0.0000000 | 0.0131579 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 |
| w5 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0131579 | 0.0131579 | 0.0000000 |
| w6 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0000000 | 0.0131579 | 0.0131579 | 0.0000000 |
| w? | 0.0000000 | 0.0131579 | 0.0000000 | 0.0131579 | 0.0000000 | 0.0065789 | 0.0065789 | 0.0131579 | 0.0000000 |
| CTOT | 0.0394737 | 0.0526316 | 0.0263158 | 0.0394737 | 0.0263158 | 0.0460526 | 0.0460526 | 0.0526316 | 0.0394737 |

Valores singulares:

Code

```
## [1] 0.8838 0.4612 0.3182 0.2123 0.1550 0.0000
```

Primeros dos factores, scored Fila

Code

```
##           [,1]      [,2]
## [1,] 0.97535822 -0.4888391
## [2,] -0.80865349 0.3227360
## [3,] -1.04771515 -0.4128870
## [4,] -0.98781732 -0.2609001
## [5,] 1.03525605 -0.3368523
## [6,] 0.80342873 0.4637015
## [7,] 0.03315725 0.7843452
```

Primeros dos factores, scored Columna

Code

```
##          [,1]          [,2]
## [1,]  1.06131491 -0.26164465
## [2,] -0.79513355  0.23488720
## [3,] -1.15155007 -0.73051540
## [4,]  0.01053472  1.13535452
## [5,]  1.13745324 -0.89520897
## [6,] -0.91408344  0.02547633
## [7,]  0.91505788  0.01869934
## [8,]  0.80536511  0.22895733
## [9,] -1.07268330 -0.25373815
## [10,] -0.79513355  0.23488720
## [11,]  1.06131491 -0.26164465
## [12,] -1.15155007 -0.73051540
## [13,]  0.01053472  1.13535452
## [14,]  1.13745324 -0.89520897
## [15,] -0.91408344  0.02547633
## [16,]  0.91505788  0.01869934
## [17,]  0.25005711  0.35252227
## [18,] -0.33227264 -0.41849142
## [19,] -0.91408344  0.02547633
## [20,]  0.91505788  0.01869934
## [21,] -1.07268330 -0.25373815
## [22,]  0.80536511  0.22895733
## [23,]  1.06131491 -0.26164465
## [24,] -1.07268330 -0.25373815
```

Inercia:

Code

```
## [1] 0.7811 0.2127 0.1013 0.0451 0.0240 0.0000
```

Inercia o varianza explicada:

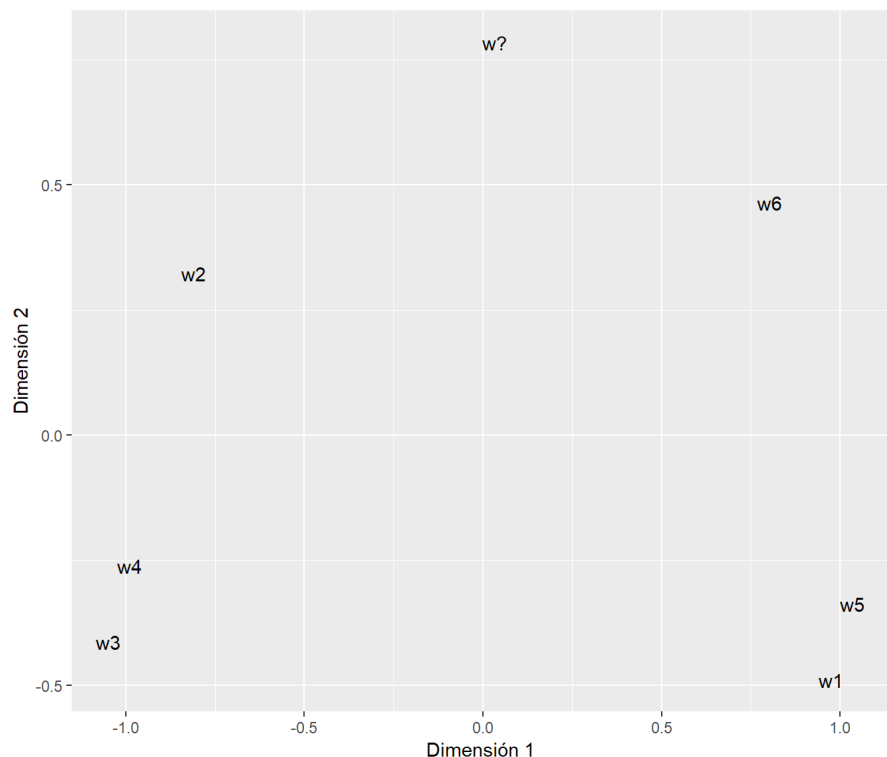
Code

Ahora se visualiza y se contestan las preguntas.

(a) ¿Cual es la tipología de los vinos?

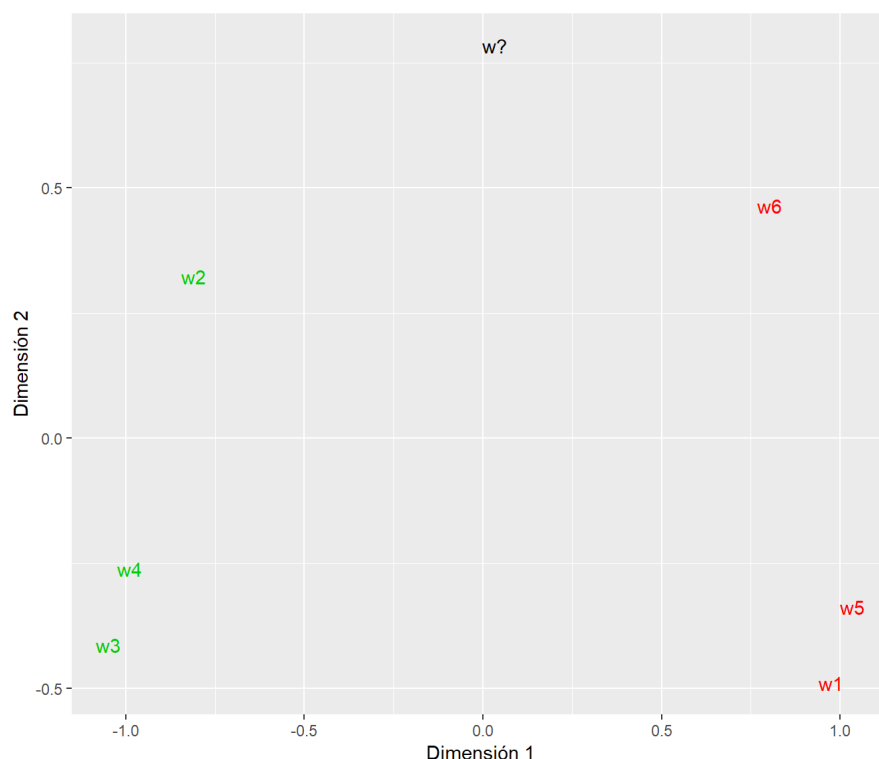
El vino 2, 3, y 4 han sido agrupados en el lado negativo los factores y el vino 1,5, y 6 en el lado positivo. Por lo tanto el vino 1,2 y 3 son del roble son del tipo de roble 2, y el 1,5 y6 del roble envejecido.

Code



(b) ¿En que tipo de roble se añejo la botella desconocida? Se colorea la representación de las filas por tipo de vino y se encuentra que el vino suplementario parece no pertenecer a ningún grupo.

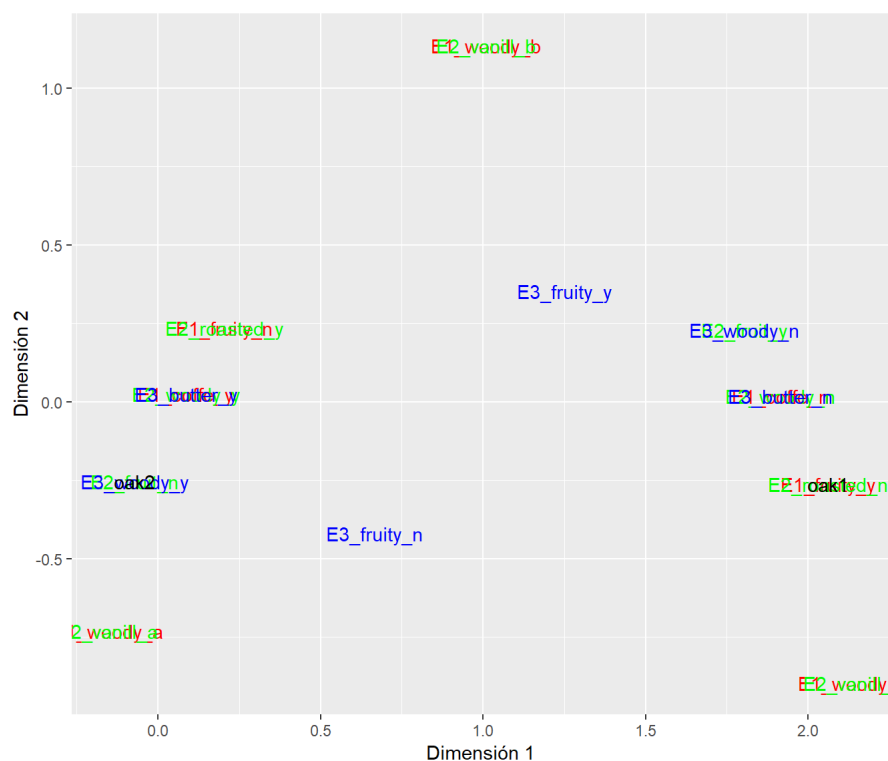
Code



El análisis de las columnas muestran que el lado negativo del factor es caracterizado como no fruity, no-woody, y coffe, esto por el experto 1. Por el experto 2, como roasted, non fruity, baja en vanilla y woody. Por último, el experto 3 indica que es buterry y woody.

NOTA: experto uno en color rojo, experto dos en color verde y experto 3 en color azul.

Code



(c) ¿Las variables están correlacionadas? El lado positivo, presenta patrones reversivos: i.e, los elementos suplementarios, oak1 y oak2, indican que el lado negativo está correlacionado con el segundo tipo de oak, donde el lado positivo está correlacionado con el primer tipo de oak.

NOTA: los puntos que distinguen del tipo de oak, se encuentran en color negro.

(d) ¿Como se ve afectado el cálculo de la inercia al tener variables ternarias?

Al codificar en binario se crea dimensiones artificiales ya que una variable categorica se codifica con varias columnas. Como consecuencia la inercia, que es la varianza, del espacio solución es artificialmente inflada y por lo tanto el porcentaje de inercia explicada por la primera dimensión es severamente subestimada.

(e) ¿Como se puede corregir este efecto para calcular correctamente la inercia?

Benzacri (1979) y Greenacre (1993) proponen dos maneras de realizarlo, que son tomar en cuenta que los valores propios que sean menores a $1/k$ son codificados para las dimensiones extras de MCA es equivalente a analizar la matriz de Burt cuyos valores propios son igual al cuadrado de analizar la matriz de contingencia simple. Con esto los autores indican que se tiene una mejor estimación de la inercia, extrayendo así los valores propios.

EJERCICIO 5

Considere la matriz de covarianza:

$$\text{cov} \begin{pmatrix} X_1^{(1)} \\ X_2^{(2)} \\ X_1^{(3)} \\ X_2^{(4)} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.95 \\ 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}$$

(a) Verifica que el primer par de variables canonicas son $U_1 = X_2^{(1)}$ y $V_1 = X_1^{(2)}$ con correlación canonicas $p_1^* = 0.95$

Se tiene lo siguiente

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \begin{pmatrix} 0 & 0 \\ 0 & (0.95)^2 \end{pmatrix}$$

con valores propios de $p_1^{*2} = (0.95)^2$ y $p_2^{*2} = 0$, con vectores propios normalizados que son:

$$e_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$e_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Entonces, las variables canonicas están dadas por:

$$U_1 = e_1' \Sigma_{11}^{-1} X^{(1)} = (0 \quad 1) \begin{pmatrix} .1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_1^{(1)} \\ X_2^{(1)} \end{pmatrix} = X_2^{(1)}$$

y sabiendo que

$$f_1' \Sigma_{22}^{-1} = (0 \quad 1)$$

con

$$V_1 = f_1' \Sigma_{22}^{-1} X^{(2)} = (1 \quad 0) \begin{pmatrix} 1 & 0 \\ 0 & .1 \end{pmatrix} \begin{pmatrix} X_1^{(2)} \\ X_2^{(2)} \end{pmatrix} = X_1^{(2)}$$

Entonces se tiene que $U_1 = X_2^{(1)}$ y $V_1 = X_1^{(2)}$ y $p_1^* = 0.95$

(b) Al observar la matriz de covarianzas se puede concluir que $U_1 = X_2^{(1)}$ y $V_1 = X_1^{(2)}$ proporcionan un buen resumen de la variabilidad de sus respectivos conjuntos?

$U_1 = X_2^{(1)}$ provee un pobre resumen de la variabilidad en el primer conjunto. Practicamente la mayor parte de la variabilidad recae en el conjunto $X_1^{(1)}$, el cual no se encuentra correlacionado con U_1 . Esto mismo sucede en la situación donde $V_1 = X_1^{(2)}$, en el segundo conjunto de datos.

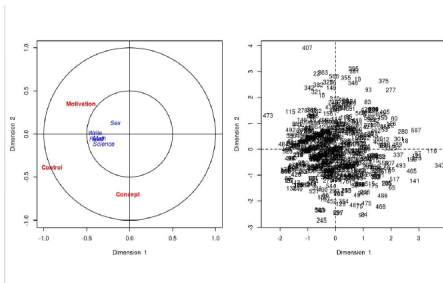
Apendice de código

Code

EJERCICIO 6

Un investigador en ciencias sociales ha recopilado datos de 600 individuos (hombres y mujeres) en 6 variables, de las cuales 3 son psicoclas (Control, Concept y Motivation), 4 variables académicas (Read, Write, Math, Science) que corresponden a scores estandarizados de diversas pruebas realizadas, y finalmente, una variable que corresponde al genero (Female). El investigador es interesado en la relación que existe entre las variables psicológicas y las variables académicas junto con la del genero. Un ejemplo de los datos se muestra a continuación:

Code



En base a los resultados podemos construir una ecuación que nos indique que variables pesan más en los primeros pares de variables canónicas, los cuales son la combinación lineal de las otras variables; i.e, es un resumen de las variables en términos de una sola.

El primer par de variables que resumen la información de los dos grupos de covariables, las psicológicas y académicas están dadas por:

$$U_1 = -1.2z_1^{(1)} + 0.35z_2^{(1)} - 1.26z_3^{(1)}$$

$$V_1 = -0.04z_1^{(2)} - 0.035z_2^{(2)} - 0.023z_3^{(2)} - 0.005z_4^{(2)} - 0.63z_5^{(2)}$$

con una correlación canónica de 0.46.

Eso significa U_1 la representa la variable psicológica concepto y a V_1 las variables académicas ciencia y matemáticas; de esta manera existe relación moderada entre la variable que mide el desempeño conceptual y el desempeño académico en las ciencias y matemáticas; observando el gráfico se puede observar esta relación en la parte inferior izquierda del gráfico de la izquierda, donde matemáticas y ciencias se encuentran cercanas, y la variable concepto de bajo de ellas.

Por otro lado, si vemos el segundo par de variables canónicas, se encuentra una correlación canónica del .16; i.e, existe una ligera relación entre las variables motivación y la variable que mide el desempeño de la habilidad escrita; asimismo, la motivación se encuentra relacionada de acuerdo al género. Esto se puede observar en las siguientes ecuaciones:

$$U_2 = -0.62z_1^{(1)} - 1.185z_2^{(1)} + 2.0226z_3^{(1)}$$

$$V_2 = -0.0004z_1^{(2)} + 0.042z_2^{(2)} + 0.0042z_3^{(2)} - 0.085z_4^{(2)} - 1.08z_5^{(2)}$$

En U_2 , se observa que está representada por la variable psicológica motivación, y en el segundo conjunto de variables, el resumen de ellas V_2 , está representada por la variable sexo y escritura, las cuales se observan en el gráfico izquierdo, en el cuadrante superior izquierdo, donde motivación está cerca de write y sex, haciendo notar la relación entre variables.

Entonces U_1 , son aquellos cuya habilidad conceptual sobresalen, y V_1 , aquellos que son buenos en las habilidades académicas cuantitativas, lo cual está muy relacionado por la exigencia en esta rama académica.

Para U_2 , representa motivación, y V_2 , representa la habilidad de escribir en función de ser mujer u hombre.

En base lo anterior podemos utilizar el gráfico izquierdo e indicar a posibles candidatos que tengan estas habilidades; por ejemplo. el sujeto 407, puede ser un individuo cuyo estado motivacional este en función de su género, al igual que su forma de redactar.

Por otro lado, el individuo 245, puede ser un sujeto que es bueno en la ciencia o matemáticas, y tiene gran habilidad para conceptualizar el contenido de estas ramas.

