Curso Estadística IV

Sociología
Universidad
Alberto Hurtado

Profesora

Carolina Aguilera
caguilera@uahurtado.cl

Ayudantes

Vicente Díaz - vidiazam@alumnos.uahurtado.cl
Miguel Tognarelli - mtognare@alumnos.uahurtado.cl

UNIVERSIDAD
ALBERTO HURTADO
FACULTAD DE CIENCIAS SOCIALES

Clase 9
8 oct

- Chi cuadrado – prueba
- Repaso Análisis Componentes Principales
- Análisis Factorial exploratorio
    - Diferencia con PCA
- Ejercicio

Chi cuadrado - prueba

Sexo / preferencia musical	Hombre	Mujer
Música tipo 1	83	20
Música tipo 2	15	90

X^2 = 91,72
GL = 1

$X^2 > X^2_{crítico}$ = 3,841, con 95,5 % de certeza

Se rechaza $H_{0,}$ es decir con un 95,5% de certeza se puede afirmar que si existe relación o asociación entre las variables

¿Qué asociación?

Miremos la tabla: $H > M_1$; $M > M_2$

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fit$eig

Variable	eigenvalue	percentage of variance	cumulative percentage of variance	Variable	eigenvalue	percentage of variance	cumulative percentage of variance	Variable
comp 1	4,67768373	58,4710466	58,4710466	comp 1	4,67768373	58,4710466	58,4710466	comp 1
comp 2	1,16535572	14,5669464	73,037993	comp 2	1,16535572	14,5669464	73,037993	comp 2
comp 3	0,54703191	6,83789883	79,8758918	comp 3	0,54703191	6,83789883	79,8758918	comp 3
comp 4	0,43407115	5,42588937	85,3017812	comp 4	0,43407115	5,42588937	85,3017812	comp 4
comp 5	0,36872714	4,60908929	89,9108705	comp 5	0,36872714	4,60908929	89,9108705	comp 5
comp 6	0,29284372	3,66054648	93,571417	comp 6	0,29284372	3,66054648	93,571417	comp 6
comp 7	0,27986199	3,49827492	97,0696919	comp 7	0,27986199	3,49827492	97,0696919	comp 7
comp 8	0,23442465	2,93030809	100	comp 8	0,23442465	2,93030809	100	comp 8

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fit$eig

Variable	eigenvalue	percentage of variance	cumulative percentage of variance	Variable	eigenvalue	percentage of variance	cumulative percentage of variance	Variable
comp 1	4,67768373	58,4710466	58,4710466	comp 1	4,67768373	58,4710466	58,4710466	comp 1
comp 2	1,16535572	14,5669464	73,037993	comp 2	1,16535572	14,5669464	73,037993	comp 2
comp 3	0,54703191	6,83789883	79,8758918	comp 3	0,54703191	6,83789883	79,8758918	comp 3
comp 4	0,43407115	5,42588937	85,3017812	comp 4	0,43407115	5,42588937	85,3017812	comp 4
comp 5	0,36872714	4,60908929	89,9108705	comp 5	0,36872714	4,60908929	89,9108705	comp 5
comp 6	0,29284372	3,66054648	93,571417	comp 6	0,29284372	3,66054648	93,571417	comp 6
comp 7	0,27986199	3,49827492	97,0696919	comp 7	0,27986199	3,49827492	97,0696919	comp 7
comp 8	0,23442465	2,93030809	100	comp 8	0,23442465	2,93030809	100	comp 8

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fitvarcoord

Variable	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
barrio_ide	0,33591998	-0,0528410	0,89086844	-0,2128212	-0,2024523	0,06662419	-0,0006151	-0,0001582
integrado	0,07310189	0,86334077	0,03325243	-0,0569038	0,01208607	-0,2823935	-0,3867384	0,12448379
identifico	0,13414147	0,80345394	-0,0993648	-0,4207129	0,0173528	-0,0293980	0,3625231	-0,1304099
parte_de_	0,04080132	0,59930185	0,15790733	0,70845129	-0,0348317	0,31701348	0,09961183	-0,0264816
amigos	0,68645829	0,04603811	-0,2454731	-0,2681734	-0,0119024	0,59859408	-0,1464554	0,12073004
sociable	0,8095792	-0,1033340	-0,1130784	0,17224467	-0,2231163	-0,2964151	0,20277871	0,33568816
cordialida	0,83567331	-0,1021287	-0,1435026	0,1327969	-0,2257147	-0,1631593	-0,1197634	-0,4013482
colaborac	0,54622824	-0,0735775	0,1559223	0,06746141	0,81219232	-0,0829659	0,02240377	-0,0177934

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fitvarcoord

Variable	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
barrio_ide	0,33591998	-0,0528410	0,89086844	-0,2128212	-0,2024523	0,06662419	-0,0006151	-0,0001582
integrado	0,07310189	0,86334077	0,03325243	-0,0569038	0,01208607	-0,2823935	-0,3867384	0,12448379
identifico	0,13414147	0,80345394	-0,0993648	-0,4207129	0,0173528	-0,0293980	0,3625231	-0,1304099
parte_de_	0,04080132	0,59930185	0,15790733	0,70845129	-0,0348317	0,31701348	0,09961183	-0,0264816
amigos	0,68645829	0,04603811	-0,2454731	-0,2681734	-0,0119024	0,59859408	-0,1464554	0,12073004
sociable	0,8095792	-0,1033340	-0,1130784	0,17224467	-0,2231163	-0,2964151	0,20277871	0,33568816
cordialida	0,83567331	-0,1021287	-0,1435026	0,1327969	-0,2257147	-0,1631593	-0,1197634	-0,4013482
colaborac	0,54622824	-0,0735775	0,1559223	0,06746141	0,81219232	-0,0829659	0,02240377	-0,0177934

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fitvarcoord

Variable	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
barrio_ide	0,33591998	-0,0528410	0,89086844	-0,2128212	-0,2024523	0,06662419	-0,0006151	-0,0001582
integrado	0,07310189	0,86334077	0,03325243	-0,0569038	0,01208607	-0,2823935	-0,3867384	0,12448379
identifico	0,13414147	0,80345394	-0,0993648	-0,4207129	0,0173528	-0,0293980	0,3625231	-0,1304099
parte_de	0,04080132	0,59930185	0,15790733	0,70845129	-0,0348317	0,31701348	0,09961183	-0,0264816
amigos	0,68645829	0,04603811	-0,2454731	-0,2681734	-0,0119024	0,59859408	-0,1464554	0,12073004
sociable	0,8095792	-0,1033340	-0,1130784	0,17224467	-0,2231163	-0,2964151	0,20277871	0,33568816
cordialida	0,83567331	-0,1021287	-0,1435026	0,1327969	-0,2257147	-0,1631593	-0,1197634	-0,4013482
colaborac	0,54622824	-0,0735775	0,1559223	0,06746141	0,81219232	-0,0829659	0,02240377	-0,0177934

Repaso Análisis Componentes Principales

Para comprender bien el EFA, se requiere entender bien el PCA
Ejemplo de la prueba

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

fitvarcoord

Variable	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
barrio_ide	0,33591998	-0,0528410	0,89086844	-0,2128212	-0,2024523	0,06662419	-0,0006151	-0,0001582
integrado	0,07310189	0,86334077	0,03325243	-0,0569038	0,01208607	-0,2823935	-0,3867384	0,12448379
identifico	0,13414147	0,80345394	-0,0993648	-0,4207129	0,0173528	-0,0293980	0,3625231	-0,1304099
parte_de_	0,04080132	0,59930185	0,15790733	0,70845129	-0,0348317	0,31701348	0,09961183	-0,0264816
amigos	0,68645829	0,04603811	-0,2454731	-0,2681734	-0,0119024	0,59859408	-0,1464554	0,12073004
sociable	0,8095792	-0,1033340	-0,1130784	0,17224467	-0,2231163	-0,2964151	0,20277871	0,33568816
cordialida	0,83567331	-0,1021287	-0,1435026	0,1327969	-0,2257147	-0,1631593	-0,1197634	-0,4013482
colaborac	0,54622824	-0,0735775	0,1559223	0,06746141	0,81219232	-0,0829659	0,02240377	-0,0177934

Repaso Análisis Componentes Principales

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

Interpretación

Cada componente (en este caso considero dos, pero también podría considerar una es una combinación lineal de todas las variables, donde algunas pesan más y otras menos (pesos dados por las correlaciones que acabamos de ver:

$$PC_1 = w_{11}X_1 + w_{12}X_2 + ... + w_{1p}X_p$$

$$PC_2 = w_{21}X_1 + w_{22}X_2 + ... + w_{2p}X_p$$

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

En la PC1 las Xi que más pesan son amigos, sociales, cordialidad y colaboración (dim “sociabilidad”)

En la PC2 son integrado, identifico y parte_de_mi (dim “pertenencia”)

Las variables se tienden a agrupar en dos grupos. Por tanto quienes tienen valores en alguna de las variables del grupo 1, también tienden a tener valores altos en las otras variables del grupo. El modelo no dice mucho sobre la relación entre estas variables a nivel de los casos (componentes son ortogonales entre si)

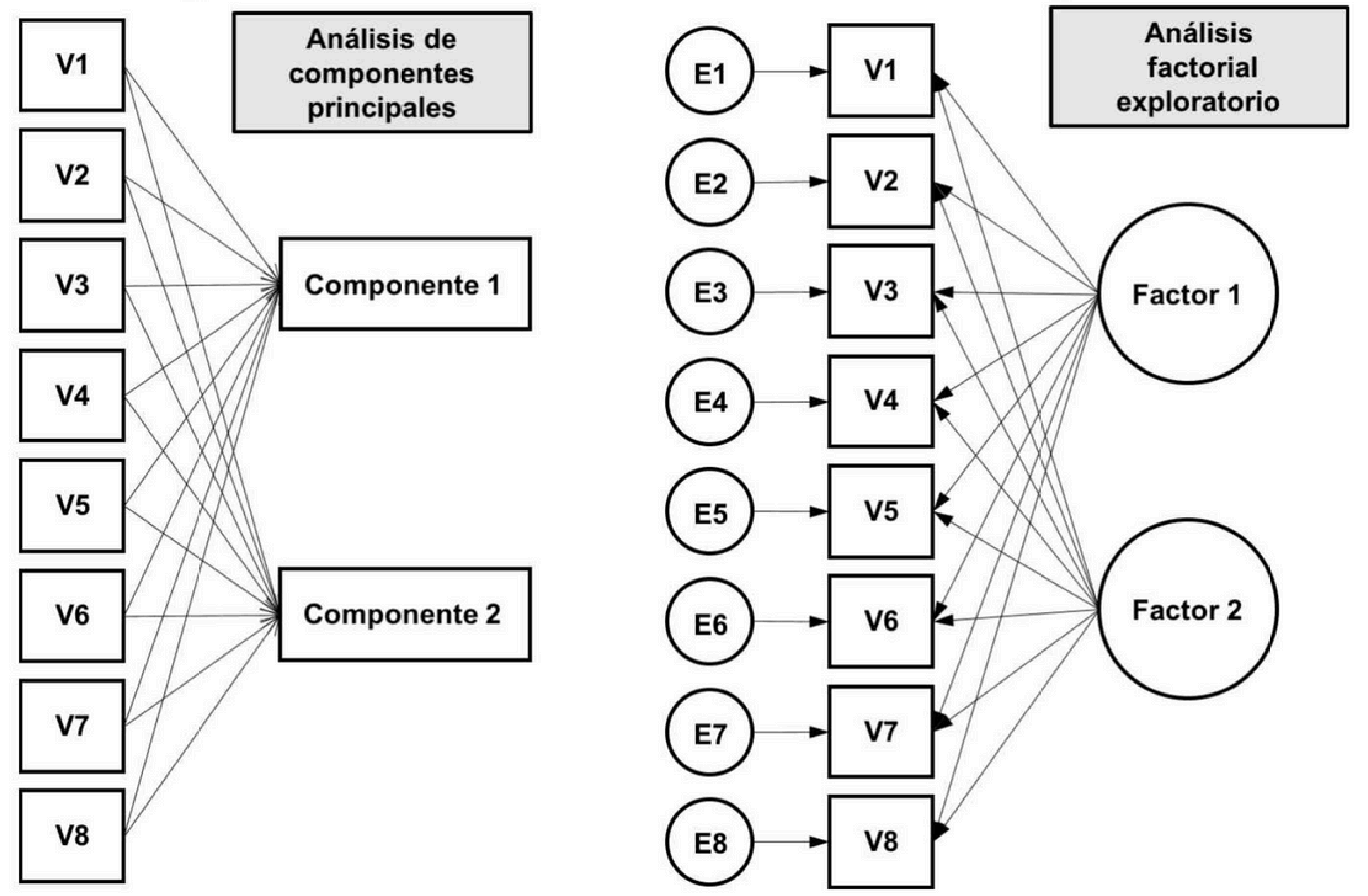
Eso no da variables latentes, solo agrupa variables

¿Qué pasa ahora con el EFA?

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

Figura 12.1.: Visión conceptual diferencial del PCA y del EFA



Fuente: Osborne y Banjanovic (2016).

$$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \cdots \lambda_{1m}\xi_m + \varepsilon_1$$

$$x_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \cdots \lambda_{2m}\xi_m + \varepsilon_2$$

$$\cdots$$

$$x_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \cdots \lambda_{pm}\xi_m + \varepsilon_p$$

χi = variables manifiestas (observadas)

λij = es el peso del factor j en la variable i (ca factoriales

ξi = son los factores comunes

εi = son factores únicos o errores

Ya no buscamos agrupar variables (PCA), sino descubrir la existencia de variables latentes que EXPLICAN la estructura latente de las variables. Se asume teóricamente que esto existe.

Aldás y Uriel, 2017

$$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \cdots \lambda_{1m}\xi_m + \varepsilon_1$$

$$x_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \cdots \lambda_{2m}\xi_m + \varepsilon_2$$

$$\cdots$$

$$x_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \cdots \lambda_{pm}\xi_m + \varepsilon_p$$

Todas las variables originales vienen influidas por todos los factores comunes

Existe un factor único que es específico para cada variable.

Tanto los factores comunes como los factores únicos no
son observables.

χi = variables manifiestas (observadas)

λij = es el peso del factor j en la variable i (cargas factoriales

ξi = son los factores comunes

εi = son factores únicos o errores

Aldás y Uriel, 2017

Diferencia entre PCA y EFA

El **PCA es una versión computacionalmente más sencilla** del EFA.

El EFA fue desarrollado previamente al PCA (Hotelling, 1933) gracias a los trabajos de Spearman (1904), pero en esa época previa a los computadores, el EFA era demasiado exigente para cálculos manuales que se realizaban y el PCA se generó como una alternativa menos costosa en términos de esfuerzo de cálculo (Gorsuch, 1990).

Aldás y Uriel, 2017

Diferencia entre PCA y EFA

En el PCA no importa cuál es la **estructura latente de las variables**, es decir, si hay **factores que están provocando que esas variables estén correlacionadas entre sí**.

Para el PCA las variables son en sí mismas el objeto del interés, no su estructura subyacente. En buena medida esto convierte al PCA en una herramienta similar a la regresión, al generar combinaciones lineales ponderadas de las variables.

El objetivo del EFA, sin embargo, es otro. Lo que **busca es intentar detectar si hay variables latentes (no observadas) que explican por qué las variables manifiestas (indicadores) están correlacionadas entre sí y pueden agruparse en un proceso de reducción de datos.**

Aldás y Uriel, 2017

Fundamentos del EFA

Se estiman la o las variables latentes a un conjunto de indicadores, sin una especificación previa de la estructura factorial.

Preguntas a responder:

- ¿Cuántos factores subyacen a un conjunto de indicadores?
- ¿Cómo se relacionan los indicadores con los factores?
- ¿Cómo es la calidad del modelo estimado?

Carrasco, Olivares y Nuñez, 2024

Fundamentos del EFA

Lo **latente** puede ser entendido como **la varianza compartida** por diferentes indicadores observados

La medición de variables latentes se encuentra asociada al modelo de factor común (Thurstone) y al análisis factorial

Carrasco, Olivares y Nuñez, 2024

Pasos del EFA

- Estimación de matriz de correlaciones
- Pruebas medición de supuestos
- Decisión sobre número de factores
- Extracción de factores (varios métodos)
- Rotación
- Obtención de puntajes factoriales
- Interpretación y reporte

Carrasco, Olivares y Nuñez, 2024

Ejemplo. Confianza en las Instituciones (ELSOC)

Considere un estudio simulado en que se analiza la cohesión barrial con 8 variables de escala de Likert que miden diferentes aspectos de relación con el barrio, con los siguientes resultados

barrio_ideal	Este es el barrio ideal para mí
Integrado	Me siento integrado/a en este barrio
Identifico	Me identifico con la gente de este barrio
parte_de_mi	Este barrio es parte de mí
amigos	En este barrio es fácil hacer amigos
sociable	La gente en este barrio es sociable
cordialidad	La gente en este barrio es cordial
colaboracion	La gente en este barrio es colaboradora

Matriz de Correlaciones

barrio_ideal	*integrado*	*identifico*	*parte_de_mi*	*amigos*	*sociable*	*cordialidad*	*colaboracion*
1.00	0.70	0.63	0.67	0.37	0.36	0.45	0.35
0.70	1.00	0.74	0.72	0.47	0.45	0.49	0.42
0.63	0.74	1.00	0.73	0.49	0.48	0.52	0.44
0.67	0.72	0.73	1.00	0.45	0.43	0.48	0.41
0.37	0.47	0.49	0.45	1.00	0.65	0.54	0.48
0.36	0.45	0.48	0.43	0.65	1.00	0.67	0.56
0.45	0.49	0.52	0.48	0.54	0.67	1.00	0.57
0.35	0.42	0.44	0.41	0.48	0.56	0.57	1.00

Interpretación de R (matriz de correlaciones).

Valores de correlación cercanos a 1 = existe correlación, valores cercanos a 0 = no existe correlación

- identifico_m con integrado_m = 0.58: Personas que se identifican con el barrio consideran sentirse integradas.
- sociable_m con cordialidad = 0.67 Quienes perciben más facilidad para hacer amigos tienden a considerar a la gente del barrio cordial.

Entonces, tiene sentido hacer un EFA, para buscar variables latentes (factores).

Paso 2. Supuesto de adecuación de matriz para AFE KMO (Kaiser, Meyer, Olkin Measure of Sampling Adequacy):

1. Test de adecuación de matriz para AFE

Varía entre 0 y 1. Contrasta si las correlaciones parciales entre las variables son pequeñas

Valores pequeños (menores a 0.5) indican que los datos no serían adecuados para AFE, ya que las correlaciones entre pares de variables no pueden ser explicadas por otras variables.

```
corMat  <- proc_data %>% select(barrio_ideal:colaboracion) %>%
  cor(use = "complete.obs")  # estimar matriz pearson

KMO(corMat)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = corMat)
## Overall MSA =  0.9
## MSA for each item =
## barrio_ideal    integrado   identifico  parte_de_mi       amigos     sociable
##         0.90         0.89         0.90         0.90         0.90         0.85
##  cordialidad colaboracion
##         0.90         0.93
```

Interpretación: el valor Overall MSA = 0,9, muy cercano a 1, así que las variables si pueden explicarse por variables latentes. Ninguna variable individual arroja un valor de 0,5 o menos, por lo que se pueden considerar todas las variables

Varía entre 0 y 1. Contrasta si las correlaciones parciales entre las variables son pequeñas

Valores pequeños (menores a 0.5) indican que los datos no serían adecuados para AFE, ya que las correlaciones entre pares de variables no pueden ser explicadas por otras variables.

Paso 2. Supuesto nivel de correlaciones de la matriz: test de esfericidad de Barlett

```
cortest.bartlett(corMat, n = 3417)
```

```
## $chisq
## [1] 16081.04
##
## $p.value
## [1] 0
##
## $df
## [1] 28
```

Interpretación: el valor p = 0, por lo que se rechaza la Ho, y se concluye que sí hay significación estadística de las correlaciones

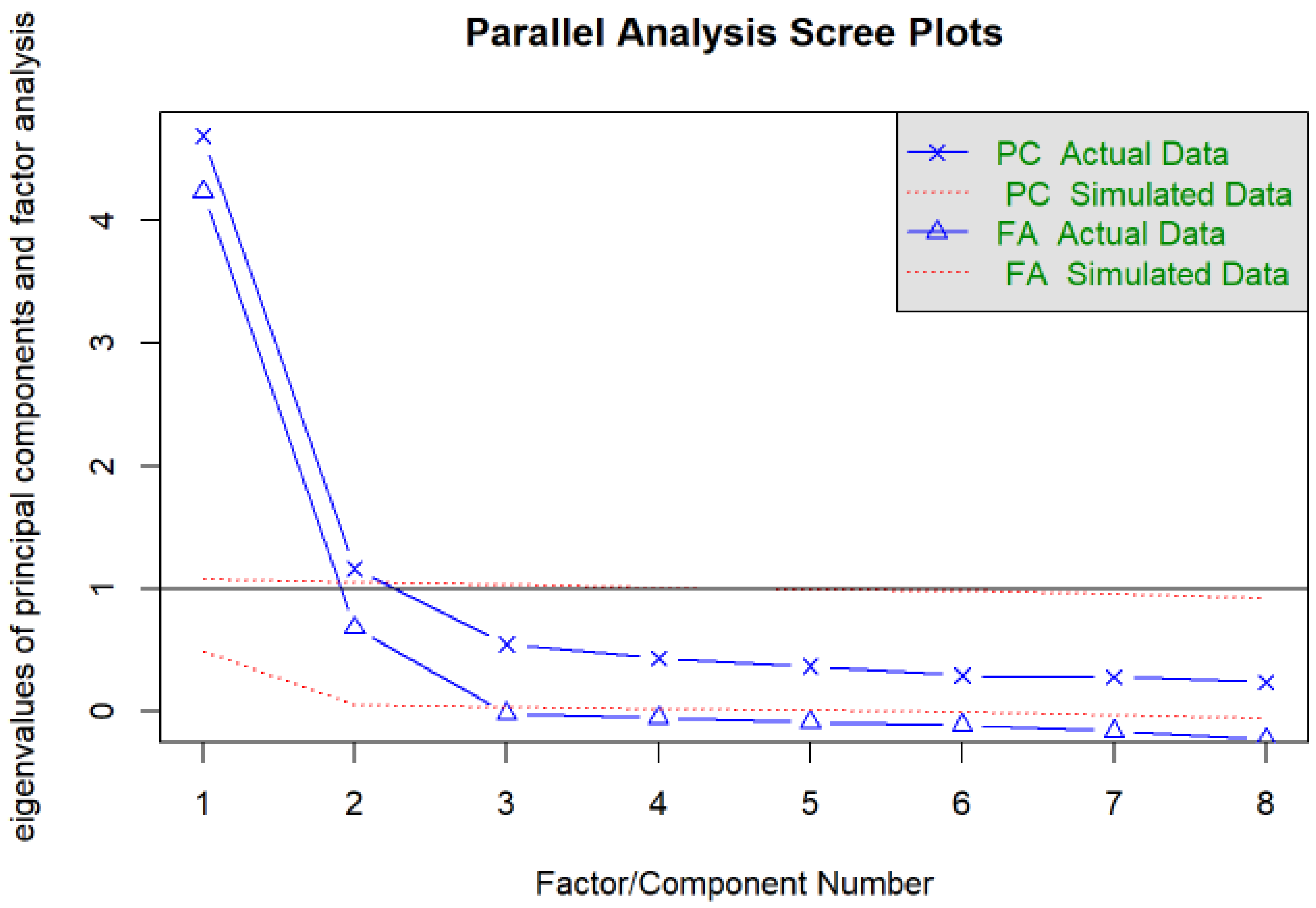
Se utiliza para evaluar la hipótesis que la matriz de correlaciones es una matriz identidad (diagonal=1 y bajo la diagonal=0): las variables no se correlacionan entre sí.

Se busca significación ($p < 0.05$) que indica que las variables estén correlacionadas

Paso 3. Definición de número de factores. Método de Análisis Paralelo

Análisis paralelo para decidir número de factores:

```
fa.parallel(corMat, n.obs=3417)
```




```
## Parallel analysis suggests that the number of factors =  2  and the number of components =  2
```

Interpretamos: el número de factores a seleccionar es 2. Este análisis es más robusto y nos quedamos con este resultado.

¿Qué nos dice este método?

¿Con cuántos factores nos quedamos?

Paso 4. Extracción de los factores con el método de Maximum Likelihood. Considerando que se considerarán 2 factores

Método Maximum likelihood

Maximiza la posibilidad de que los parámetros reproduzcan los datos observados

```
fac_ml <- proc_data %>% select(barrio_ideal:colaboracion) %>% fa(nfactors = 2, fm= "ml")
fac_ml
```

```
## Factor Analysis using method =  ml
## Call: fa(r = ., nfactors = 2, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                ML1   ML2   h2   u2 com
## barrio_ideal   0.82 -0.06 0.61 0.39 1.0
## integrado      0.86  0.01 0.76 0.24 1.0
## identifico     0.79  0.09 0.72 0.28 1.0
## parte_de_mi    0.85  [illegible] 0.71 0.29 1.0
## amigos         0.10  0.67 0.54 0.46 1.0
## sociable      -0.09  0.94 0.78 0.22 1.0
## cordialidad    0.12  0.69 0.61 0.39 1.1
## colaboracion   0.10  0.60 0.44 0.56 1.1
##
##                        ML1  ML2
## SS loadings           2.90 2.28
## Proportion Var        0.36 0.28
## Cumulative Var        0.36 0.65
## Proportion Explained  0.56 0.44
## Cumulative Proportion 0.56 1.00
##
```

- SS loadings: Suma de cuadrados de las cargas factoriales → mide la “fuerza” de cada factor.
- Proportion Var: Porcentaje de la varianza total explicada por cada factor.
- Cumulative Var: Porcentaje acumulado de varianza explicada hasta ese factor.
- Proportion Explained: Proporción de la varianza explicada solo considerando los factores retenidos.
- Cumulative Proportion: Porcentaje acumulado (siempre llega a 1 = 100%).

Tenemos dos factores que explican la estructura de los datos. En este caso, a diferencia del PCA, la varibale barrio_ideal está correlacionada fuertemente con el ML1.

Ambos factores tienen una fuerza similar, aunque el ML1 es mayor. El % de la varianza tota l explicada por MLA es 36%, y si se toman ambas como el total el ML1 explica el 56% de la varianza total.

Paso 4. Extracción de los factores con el método de Maximum Likelihood. Considerando que se considerarán 2 factores

Método Maximum likelihood

Maximiza la posibilidad de que los parámetros reproduzcan los datos observados

```
fac_ml <- proc_data %>% select(barrio_ideal:colaboracion) %>% fa(nfactors = 2, fm= "ml")
fac_ml
```

```
## Factor Analysis using method =  ml
## Call: fa(r = ., nfactors = 2, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                ML1   ML2   h2   u2 com
## barrio_ideal  0.82 -0.06 0.61 0.39 1.0
## integrado     0.86  0.01 0.76 0.24 1.0
## identifico    0.79  0.09 0.72 0.28 1.0
## parte_de_mi   0.85  0.00 0.71 0.29 1.0
## amigos        0.10  0.67 0.54 0.46 1.0
## sociable     -0.09  0.94 0.78 0.22 1.0
## cordialidad   0.12  0.69 0.61 0.39 1.1
## colaboracion  0.10  0.60 0.44 0.56 1.1
##
##                        ML1  ML2
## SS loadings           2.90 2.28
## Proportion Var        0.36 0.28
## Cumulative Var        0.36 0.65
## Proportion Explained  0.56 0.44
## Cumulative Proportion 0.56 1.00
##
```

h2 (comunalidad): proporción de varianza explicada por todos los factores juntos (ideal > 0.4, cercano a 1).

u2 (unicidad): parte no explicada (1 - h2).

com (complejidad): cuántos factores influyen en una variable (1 = carga simple; > 2 = más compleja).

Paso 4. Extracción de los factores con el método de Maximum Likelihood. Considerando que se considerarán 2 factores

```
##  With factor correlations of
##      ML1  ML2
## ML1 1.00 0.64
## ML2 0.64 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model =  28  with the objective function =  4.72 with Chi Square =  16095.21
## df of  the model are 13  and the objective function was  0.06
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.02
##
## The harmonic n.obs is  3395 with the empirical chi square  53.05  with prob <  0.00000089
## The total n.obs was  3417  with Likelihood Chi Square =  189.6  with prob <  0.000000000000000000000000000000000019
##
## Tucker Lewis Index of factoring reliability =  0.976
## RMSEA index =  0.063  and the 90 % confidence intervals are  0.055 0.071
## BIC =  83.83
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                   ML1  ML2
## Correlation of (regression) scores with factors   0.95 0.94
## Multiple R square of scores with factors          0.91 0.88
## Minimum correlation of possible factor scores     0.82 0.76
```

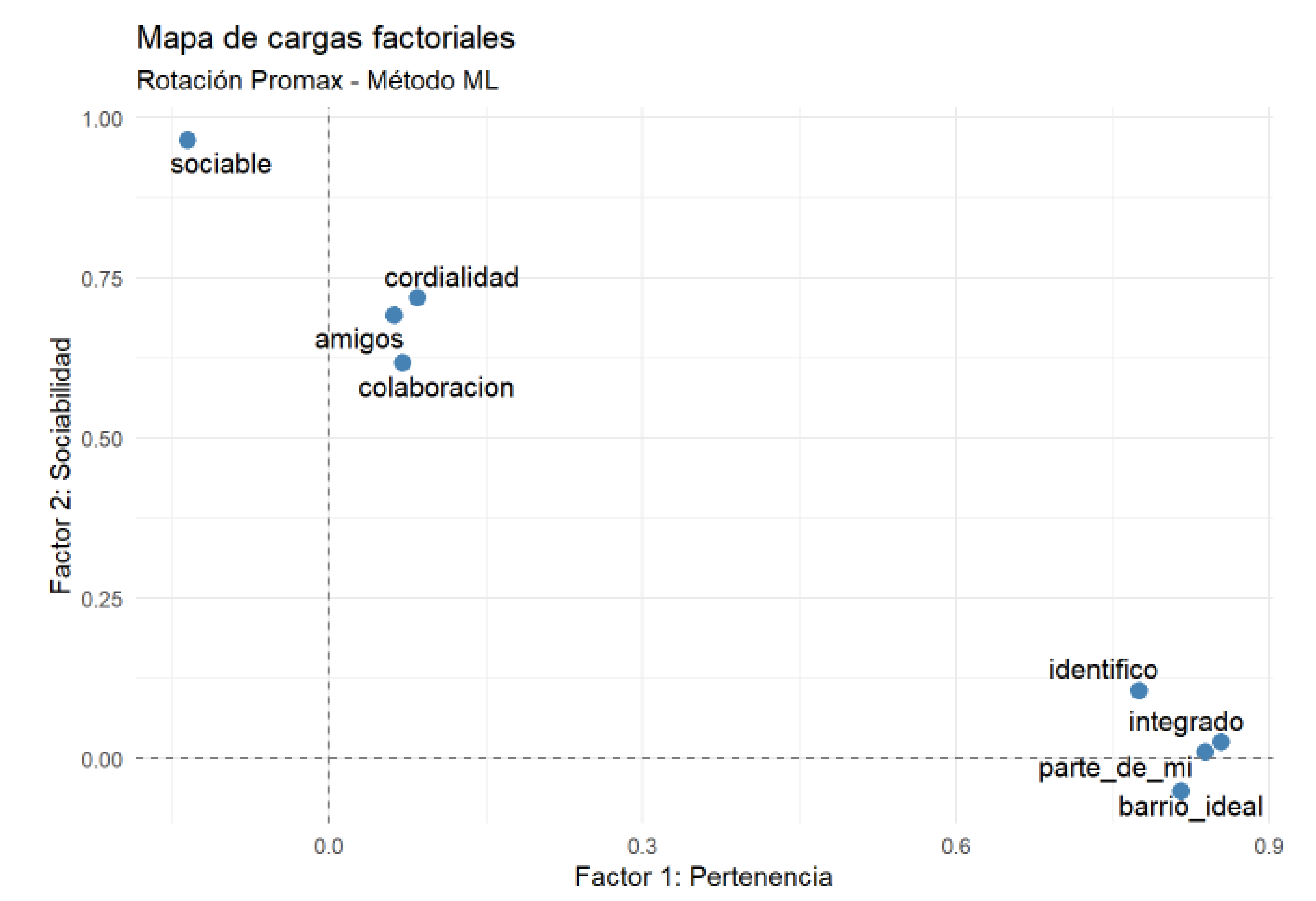
Matriz de correlación entre factores. Justifica rotación oblicua

Correlación entre los puntajes estimados y los factores verdaderos (latentes). Mientras más cerca de 1 mejor

El cuadrado de esa correlación (proporción de varianza compartida). Mide la fiabilidad de los puntajes factoriales.

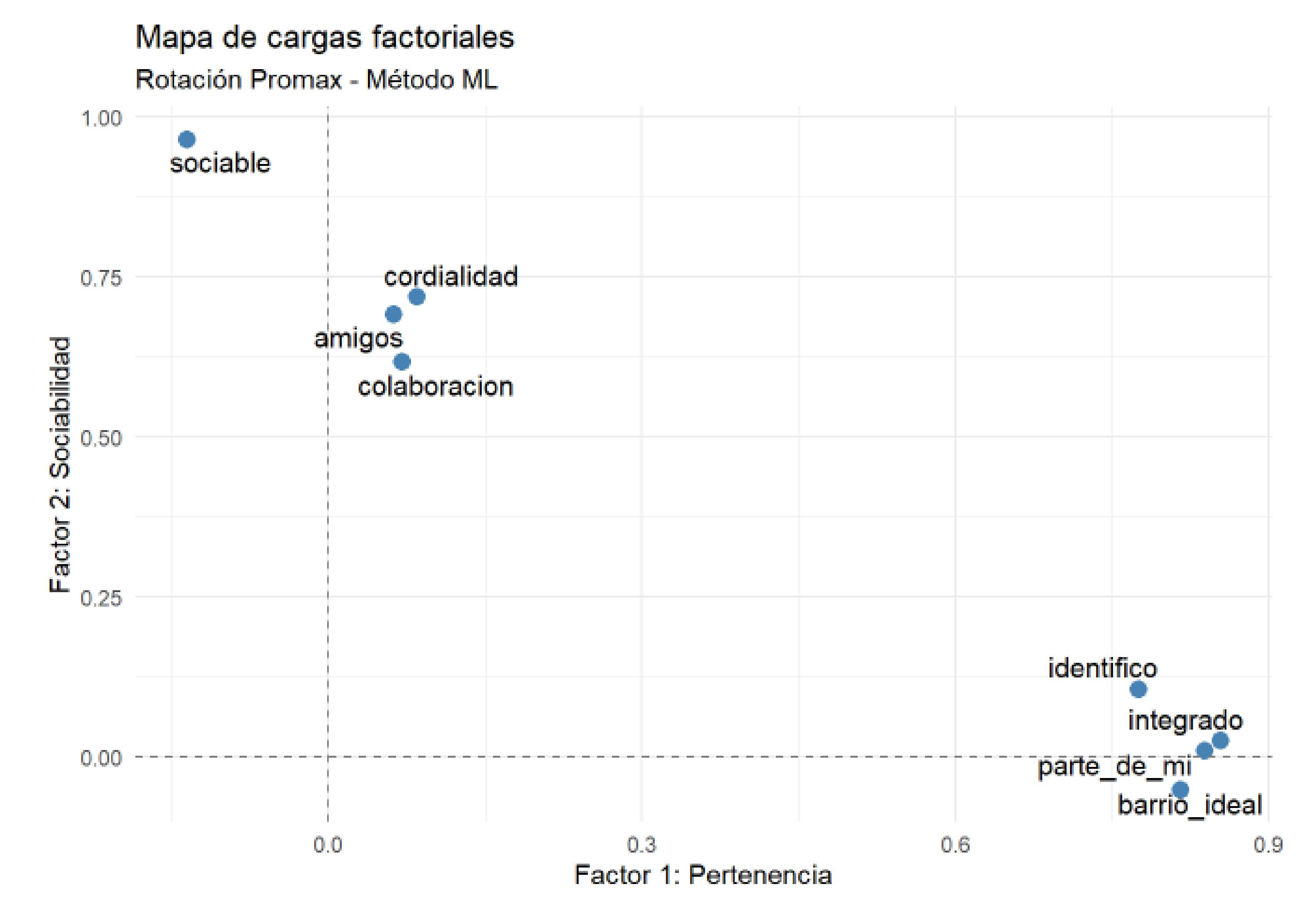
Correlación mínima esperada entre los puntajes estimados y los factores verdaderos. Cuanto más alto, mejor.

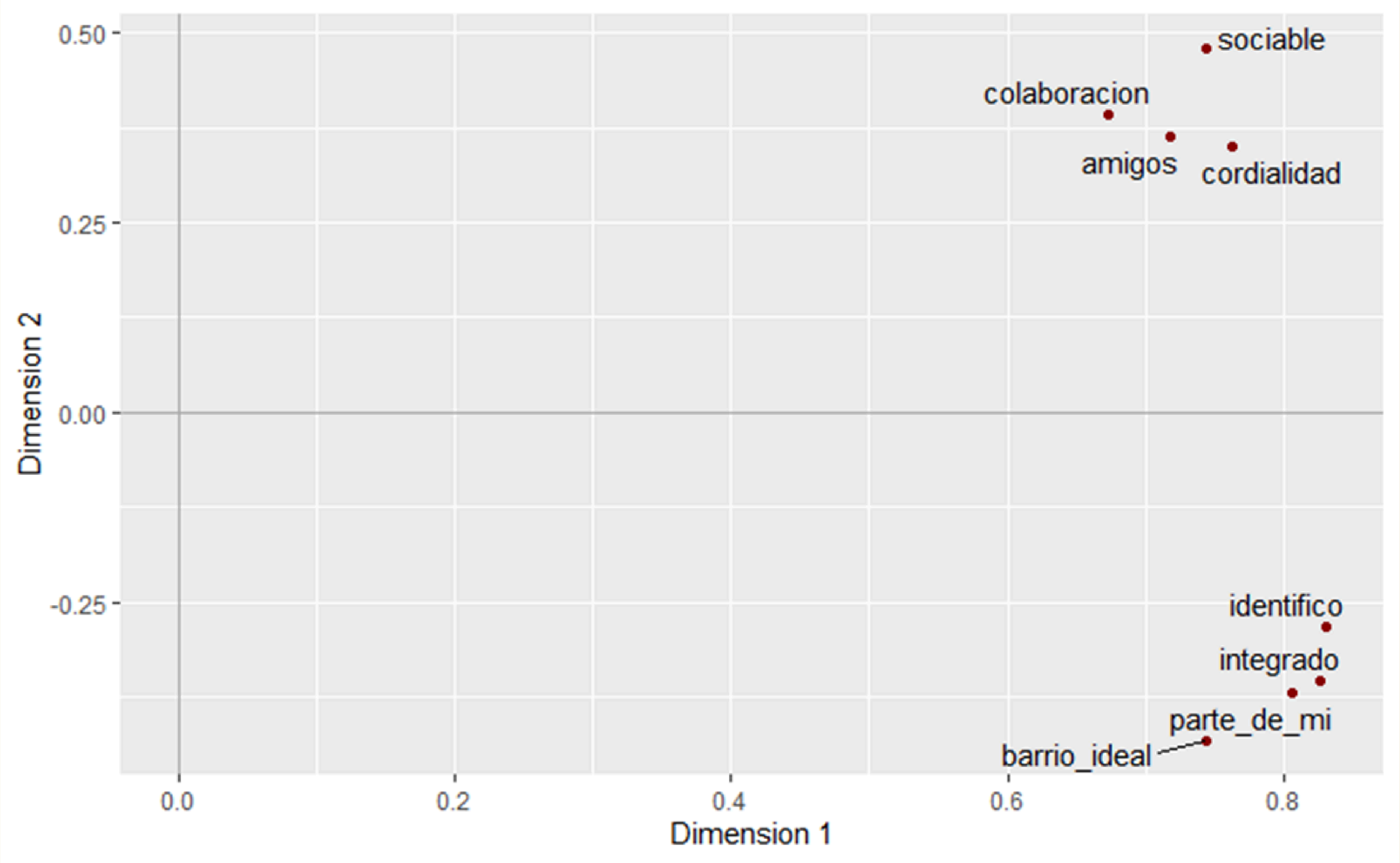
Paso 5. Mapa o gráfico de las cargas factoriales



Una vez que tenemos el modelo factorial, podemos graficas (cuando tenemos 2 factores), como se agrupan entre sí en torno a los dos factores latentes

Comparamos EFA con PCA





variable	ML1	ML2
barrio_ideal	0,815085	-0,05112
integrado	0,8539	0,02611
identifico	0,774921	0,107285
parte_de_mi	0,838834	0,010005
amigos	0,062257	0,693055
sociable	-0,13582	0,965324
cordialidad	0,084577	0,719777
colaboracion	0,069957	0,617681

```
                       ML1  ML2
SS loadings           2.90 2.28
Proportion Var        0.36 0.28
Cumulative Var        0.36 0.65
Proportion Explained  0.56 0.44
Cumulative Proportion 0.56 1.00
```

Variable	Dim.1	Dim.2
barrio_ideal	0,33591998	-0,05284101
integrado	0,07310189	0,86334077
identifico	0,13414147	0,80345394
parte_de_mi	0,04080132	0,59930185
amigos	0,68645829	0,04603811
sociable	0,8095792	-0,10333405
cordialidad	0,83567331	-0,10212876
colaboracion	0,54622824	-0,07357758

Variable	eigenvalue	percentage of	cumulative
comp 1	4,677683	58,47104	58,47104
comp 2	1,165355	14,56694	73,03799
comp 3	0,547031	6,837898	79,87589

Paso 6. Puntajes factoriales

Una vez que tenemos el modelo factorial, con las cargas factoriales, podemos calcular para cada persona (o caso) un puntaje que aproxima su valor en cada factor latente.

Combinar con los datos originales

```
proc_data_scores <- cbind(proc_data, fac_ml$scores)
```

Vemos los valores para los primeros 10 casos:

```
head(proc_data_scores[, c("pertenencia", "sociabilidad")], 10)
```

```
##    pertenencia sociabilidad
## 1  -0.19031088    0.4035489
## 2  -0.03008579    0.4594585
## 3  -0.88757110   -0.6815509
## 4   0.29062166   -0.3395309
## 5  -0.17697314   -0.3497488
## 6   0.33641159    0.4762188
## 7   0.33641159    0.4762188
## 8   0.33641159    0.4762188
## 9   0.30830384    0.3108405
## 10 -0.51783704   -0.1316368
```

Ejercicio Módulo 2

Comparar resultados entre un PCA y un EFA utilizando algún otro módulo de la encuesta ELSOC. Seguir los códigos utilizados el documento adjunto.

Alternativas:

- Usar el EFA utilizado la clase anterior y calcular un PCA y comparar
- Usar un PCA realizado anteriormente y a esas variables calcularle un EFA

La correcta realización del ejercicio dará 0,2 ptos para el Control 2