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 8
1 oct

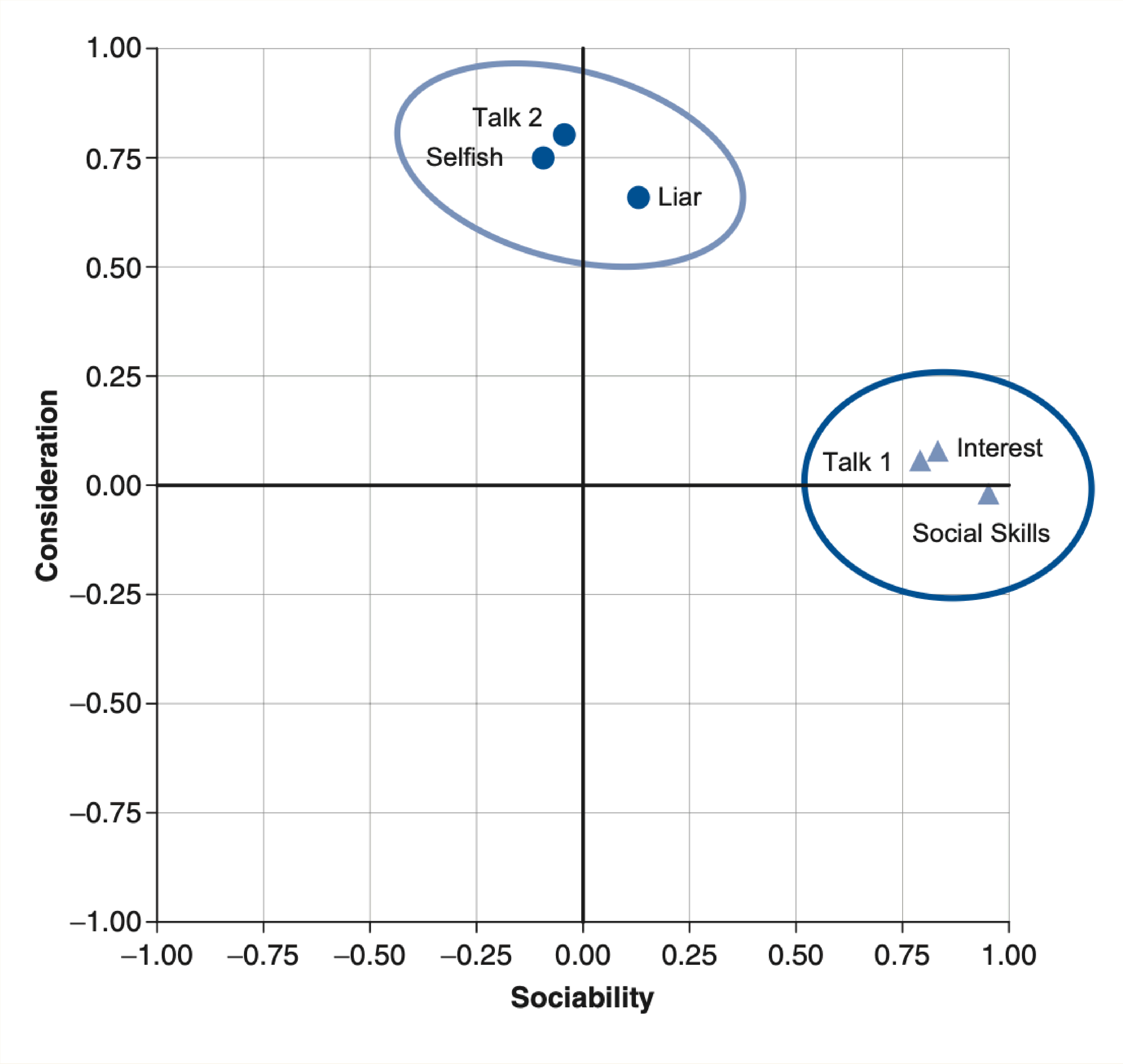
- Análisis Factorial Exploratorio (EFA)
    - Repaso general del tipo de análisis
    - Ejemplo con los elementos centrales
- Ejercicio

CONTENIDOS DEL CURSO

Método	Objetivo	Tipo de variables	Tipo de resultado	Medida base / criterio	Naturaleza del método
ANÁLISIS DE CORRESPONDENCIAS	Representar gráficamente **asociaciones entre categorías** de variables nominales	**Nominales** (categóricas) en tablas de contingencia	**Mapa perceptual** de categorías y casos en espacios 2 dimensiones	Distancia chi-cuadrado	Exploratorio, descriptivo
ANÁLISIS DE COMPONENTES PRINCIPALES	**Reducir la dimensionalidad** manteniendo la mayor varianza posible	**Cuantitativas** continuas (o cuantitativas ordinales tratadas como continuas)	**Nuevas variables** ("componentes principales") no correlacionadas	Varianza y covarianza (o matriz de correlaciones)	Exploratorio, descriptivo
ANÁLISIS FACTORIAL EXPLORATORIO	**Identificar factores latentes** que explican las correlaciones entre variables observadas	**Cuantitativas continuas (o cuantitativas ordinales tratadas como continuas)**	**Factores latentes** y cargas factoriales	Covarianza/correlació n, comunalidades	Exploratorio, modelo estadístico implícito
ANÁLISIS CONGLOMERADOS	**Agrupar casos** (o variales) en función de su similitud	**Nominales o continuas o mixtas** (según el tipo de distancia/similitud elegido)	**Grupos o clústeres** de casos similares entre si	Distancia euclídea, Manhattan, Gower, correlación, etc.	Exploratorio, descriptivo o confirmatorio

Ejemplo

Imaginemos un caso donde se mide “popularidad” mediante varias variables. Buscamos encontrar factores (latentes), en este caso 2, que den cuenta de la esctructura de los datos. En EFA da el siguiente gráfico;



	Talk 1	Social Skills	Interest	Talk 2	Selfish	Liar
Talk 1	1.000					
Social Skills	.772	1.000				
Interest	.646	.879	1.000			
Talk 2	.074	−.120	.054	1.000		
Selfish	−.131	.031	−.101	.441	1.000	
Liar	.068	.012	.110	.361	.277	1.000

Factor 1

Factor 2

Field, 2012

Ejemplo

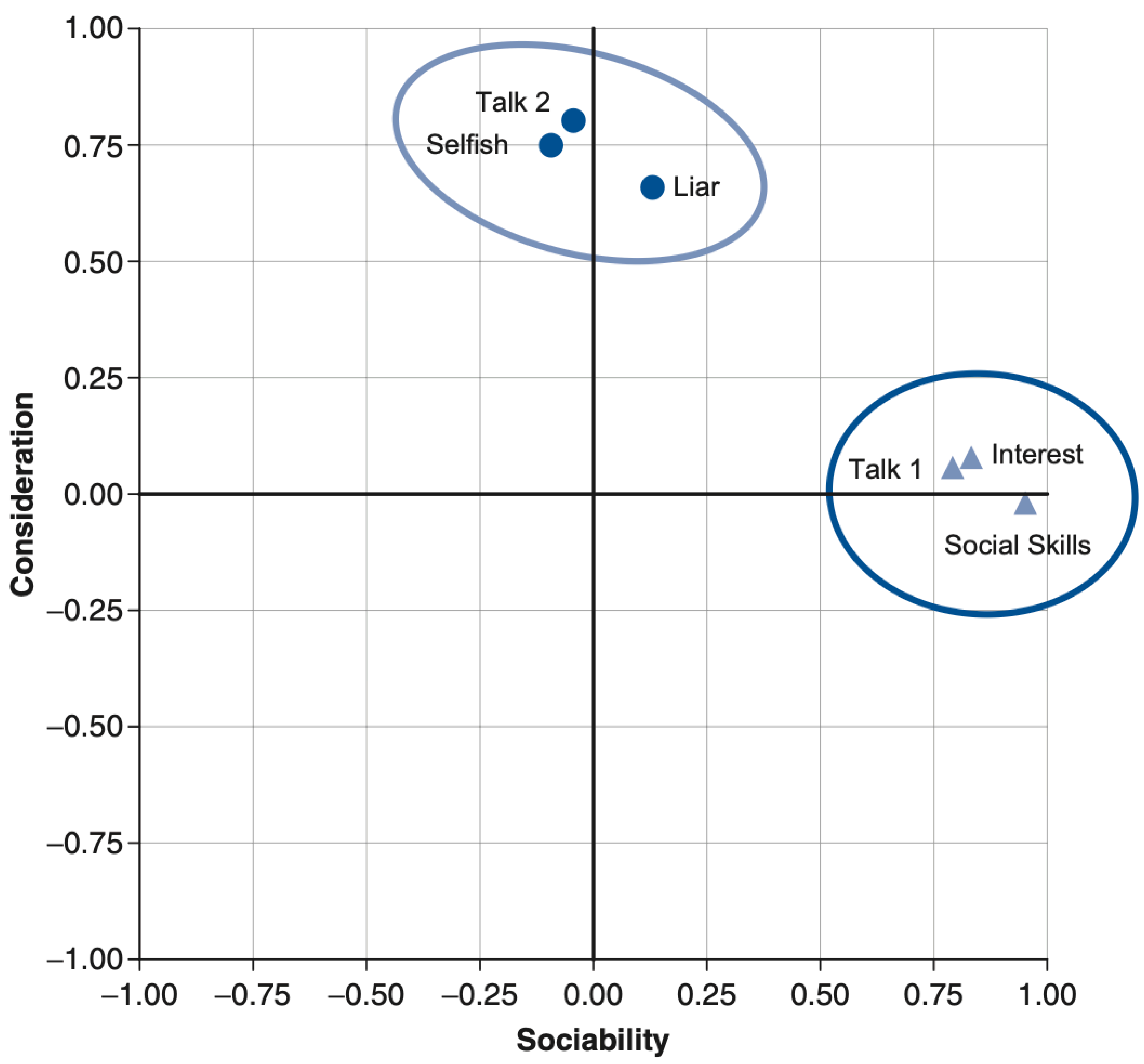


FIGURE 17.3
Example of a factor plot

Para ambos factores, la línea del eje oscila entre −1 y 1, que son los límites externos de un coeficiente de correlación. Por lo tanto, la posición de una variable dada depende de su correlación con los dos factores.

Field, 2012

Ejemplo

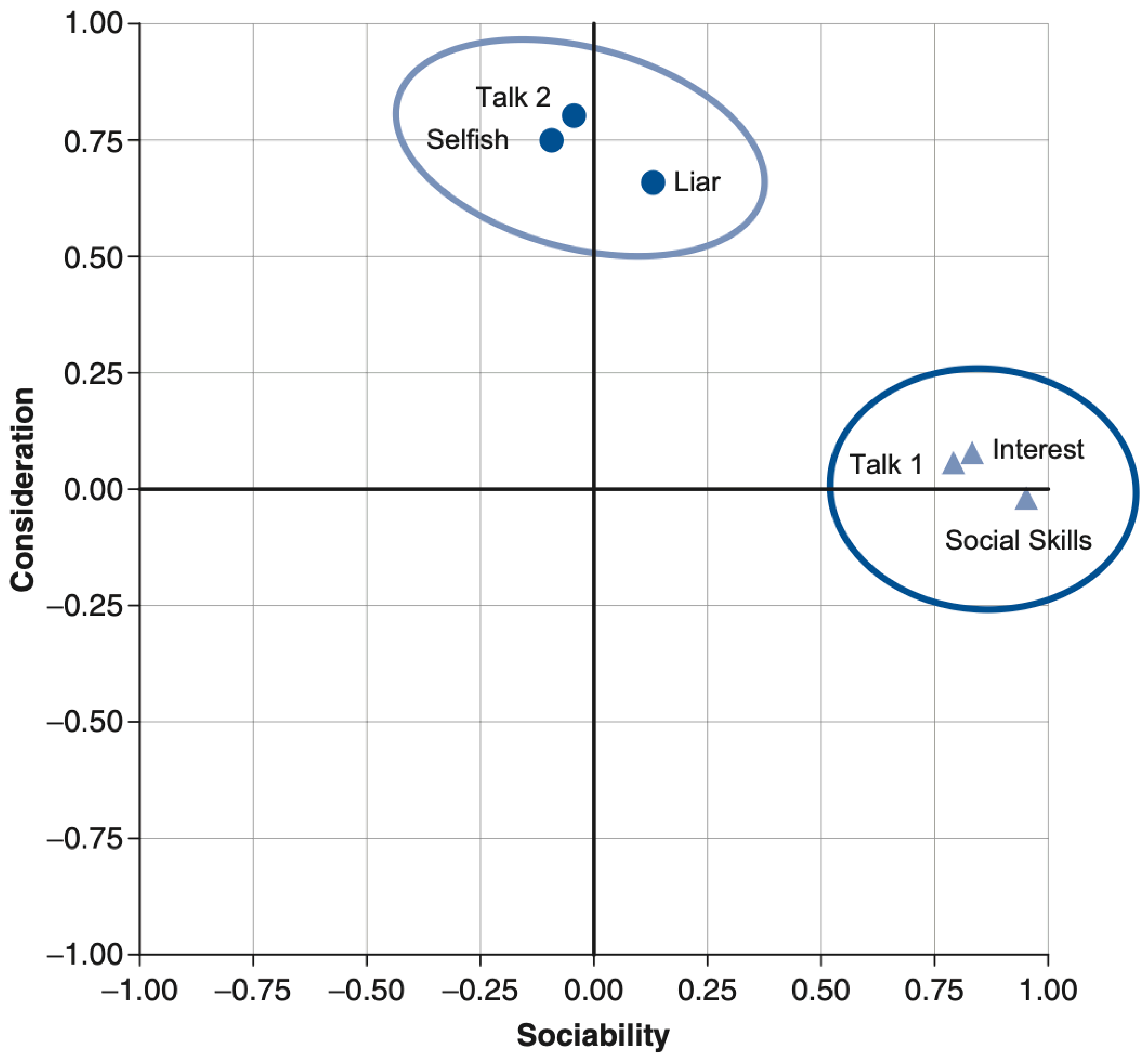


FIGURE 17.3
Example of a factor plot

Los círculos representan las tres variables que tienen una alta correlación con el factor 1 (Sociabilidad: eje horizontal), pero una baja correlación con el factor 2 (Consideración, trato hacia los demás: eje vertical).

Por el contrario, los triángulos representan variables que tienen una alta correlación con la consideración hacia los demás, pero una baja correlación con la sociabilidad.

Field, 2012

Ejemplo

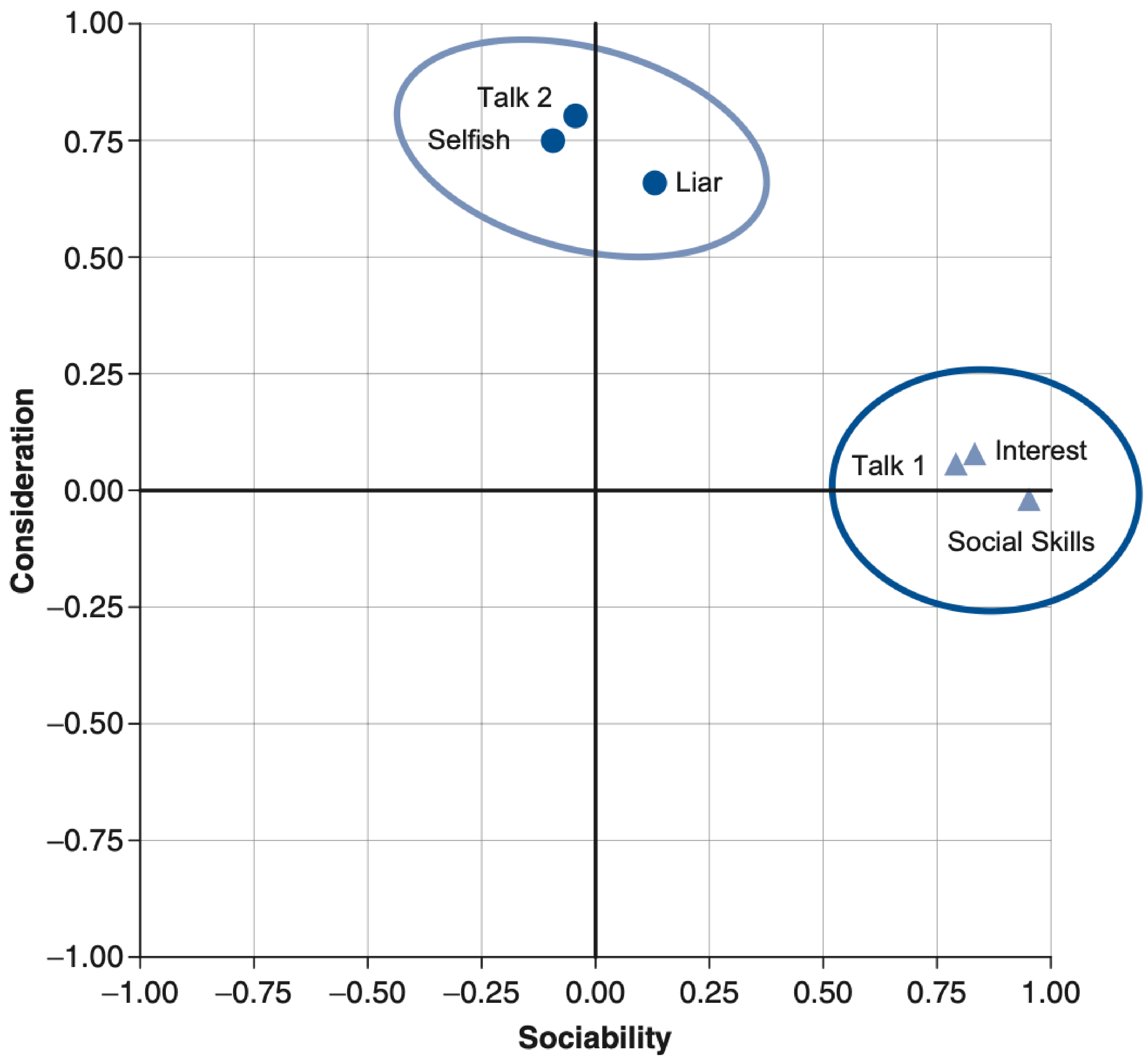


FIGURE 17.3
Example of a factor plot

Egoísmo (selfish): la cantidad de veces que una persona habla de sí misma y su propensión a mentir contribuyen a un factor que podría denominarse consideración hacia los demás.

Por el contrario, el interés de una persona por otras, su interés por otras y su nivel de habilidades sociales contribuyen a un segundo factor: la sociabilidad.

Field, 2012

Ejemplo

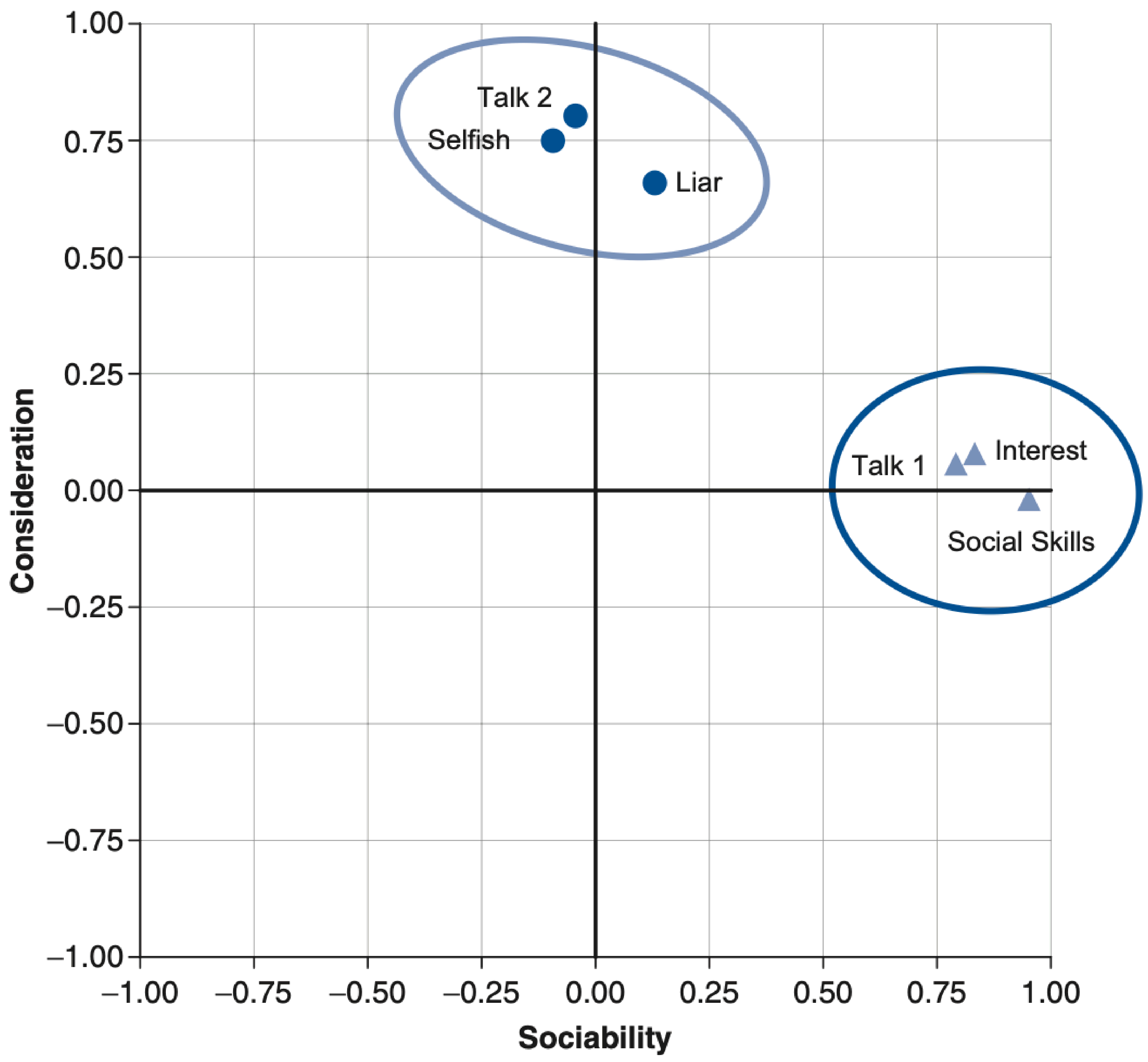


FIGURE 17.3
Example of a factor plot

Si existiera un tercer factor en estos datos, este podría representarse mediante un tercer eje (creando un gráfico tridimensional).

Si existen (o se consideran) más de tres factores en un conjunto de datos, un dibujo tridimensional no puede representarlos todos.

Field, 2012

Fundamentos del EFA

Variables en el mundo social suelen no ser directamente observables.

Constructos hipotéticos latentes

Medir variables latentes:

- a partir de indicadores observables
- ítems de un módulo o batería de un cuestionario

Carrasco, Olivares y Nuñez, 2024

Descripción general EFA

- **Objetivo teórico:** relacionar datos con dimensiones latentes basadas en conceptos (validez de constructo)

- **Objetivo pragmático**: hacer sentido de un conjunto de datos, reducción de dimensiones y obtención de puntajes

- **Objetivo metodológico**: aislar el error (varianza única) de la varianza común

Carrasco, Olivares y Nuñez, 2024

Descripción general

Técnica multivariadas de “interdependencia” (analiza relación mutua entre un conjunto de variables)
Finalidad: agrupación de variables, en función de la variabilidad que cada variable comparte con otras variables (varianza o covarianza).

Serie menor de variables latentes (factores o componentes) : aglutinan variables empíricas que están bastante correlacionadas entre sí y escasamente correlacionadas con aquellas variables empíricas que conforman otra estructura latente (o dimensión del concepto que se analice).

 Reduce la información de una matriz de correlaciones a partir de la construcción de funciones lineales.
 Descifra patrones de dependencia a partir del análisis de correlaciones múltiples.
 Identifica dimensiones que representen esquemas conceptuales de análisis.
 Valida la construcción de instrumentos de medida, particularmente escalas.

(Zamora y Esnaloa, s/f)

Descripción general

El análisis de componentes principales (PCA) se caracteriza por analizar la varianza total del conjunto de variables observadas. De ellas trata de determinar las dimensiones básicas (o "componentes") que las definen.

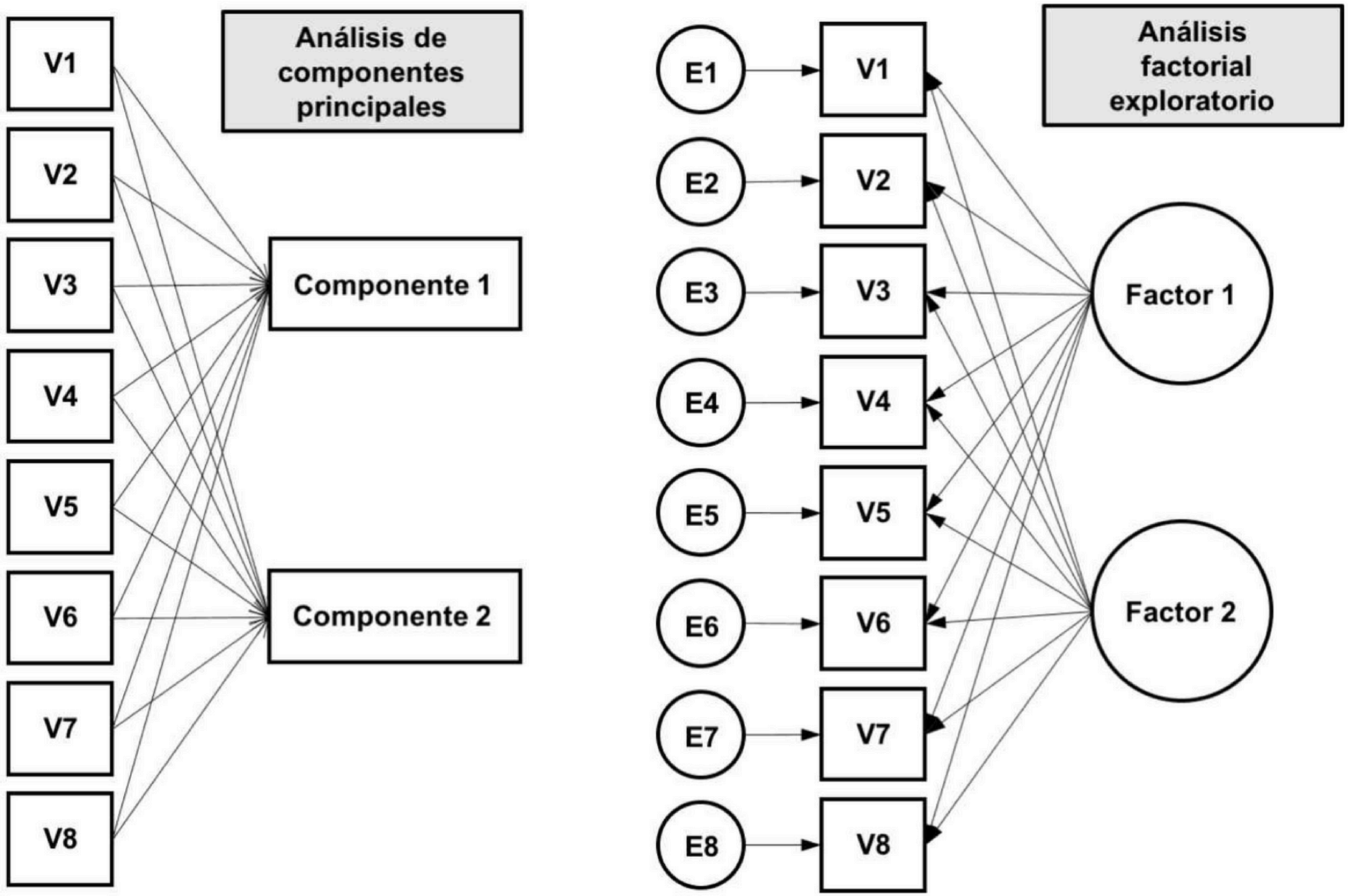
En el análisis de factor común (EFA) el estudio de las interrelaciones entre las variables **se restringe, en cambio, a la varianza común (o covarianza),** es decir, a la búsqueda de un número reducido de "factores" q**ue expresen lo que es "común"** al conjunto de variables observadas.

En el PCA el propósito es maximizar la proporción total de la varianza explicada. En cambio, **EFA está orientado al análisis de la covarianza (varianza en común o comunalidad), no de la varianza total.**

En EFA la varianza se descompone en varianza común (o comunalidad) y varianza específica. La comunalidad de cada variable expresa la porción de la varianza total de la variable “x” que es compartida con las p-1 variables observadas restantes. La varianza específica es, por el contrario, la porción de la varianza total de la variable que no es explicada por los "factores comunes" (que está compuesta por la varianza específica y el error). **En palabras simples, en el EFA los factores explican las variables, y en el PCA las variables explican los factores.**

Diferencia entre PCA y EFA

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



Fuente: Osborne y Banjanovic (2016).

Aldás y Uriel, 2017

Descripción general

- **Describe las covarianzas entre variables observadas en función de otras no observables que subyacen bajo ellas, denominadas factores**

- Si **un grupo de variables** manifiestas guarda una fuerte correlación entre ellas pero a su vez la correlación con **otro grupo de variables** es relativamente baja, es sensato pensar que **cada grupo pueda ser reflejo de un factor subyacente** que cause ese comportamiento diferenciado.

- Se usa cuando tenemos una batería o módulo de variables que están midiendo diferentes aspectos de un mismo fenómeno. Permite construir índices.

- **Se usa para variables de intervalo** (aunque se puede aplicar con precaucion para variables ordinales como escalas Likert con al menos 5 categorías, idealmente 7 ó 10)

- **Al igual que con el PCA se trabaja con las variables tipificadas (varianza = 1 y media = 0)**

Carrasco, Olivares y Nuñez, 2024

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

Fundamentos del EFA

Permite identificar la varianza común a una serie de indicadores

Establecer la contribución de cada indicador a la varianza común

Estimar posteriormente un índice (puntaje factorial) para cada factor, con mayor precisión que un promedio bruto

Carrasco, Olivares y Nuñez, 2024

Fundamentos del EFA

- Basado en la matriz de correlaciones

- Modelo estandarizado (varianza factores = 1; promedio = 0)

- Diferentes métodos de extracción de factores

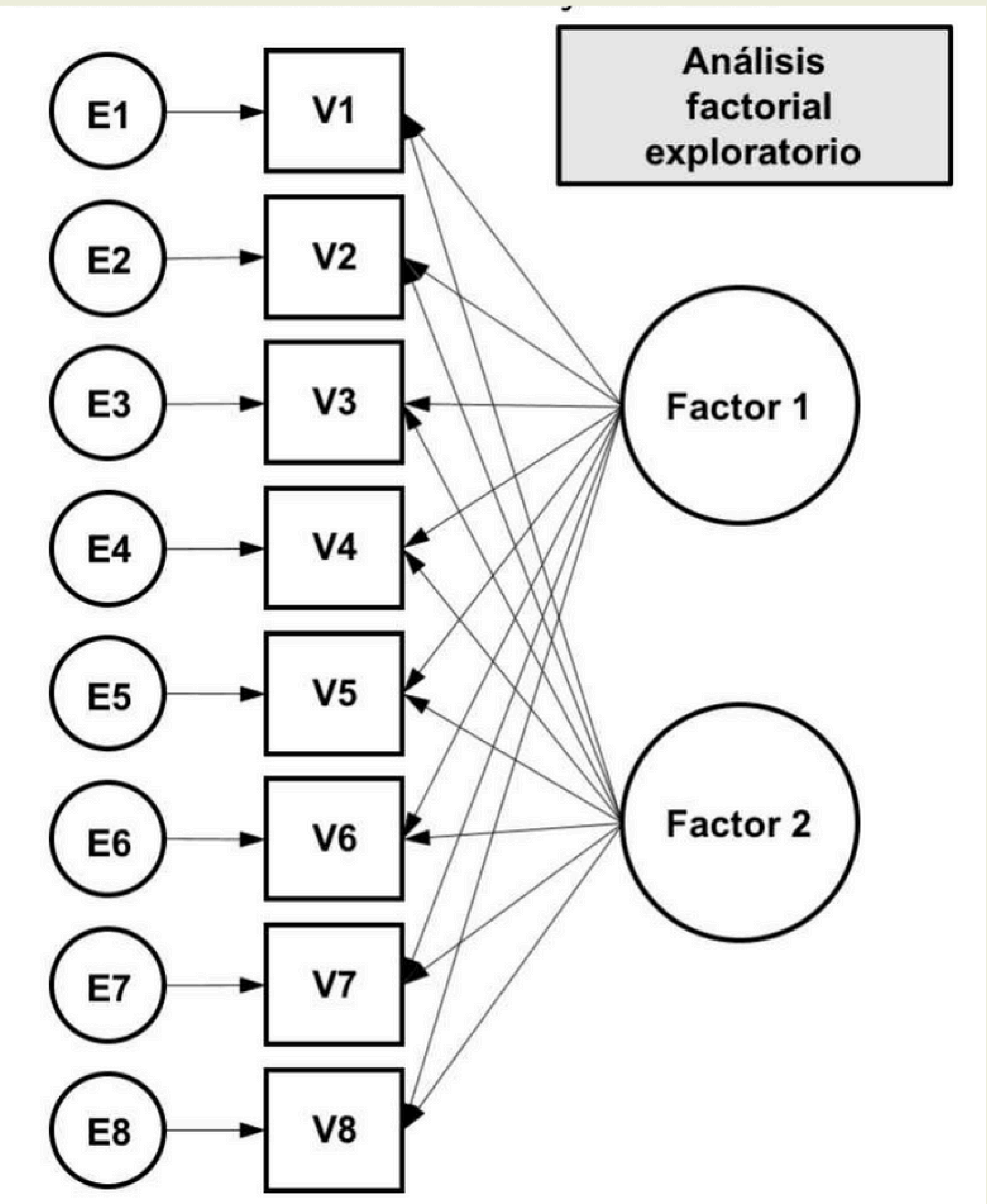
- Determinación del número y "calidad" de las dimensiones (continuas) subyacentes a una escala

Carrasco, Olivares y Nuñez, 2024

Fundamentos del EFA

- Cada indicador (variable observada) en un set de medidas observadas es una función lineal de uno o más factores comunes y un factor único

Carrasco, Olivares y Nuñez, 2024



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

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

$$\cdots$$

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

χi = variables manifiestas (observadas)

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

ξi = son los factores comunes

εi = son factores únicos o errores

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

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

Supuestos

Supuestos a evaluar (1)

- Nivel de medición de variables, normalidad (eventualmente test de normalidad multivariado, ej: Shapiro Wilk multivariado)
- Test de adecuación muestal (KMO)
    - 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 EFA, ya que las correlaciones entre pares de variables no pueden ser explicadas por otras variables

Supuestos

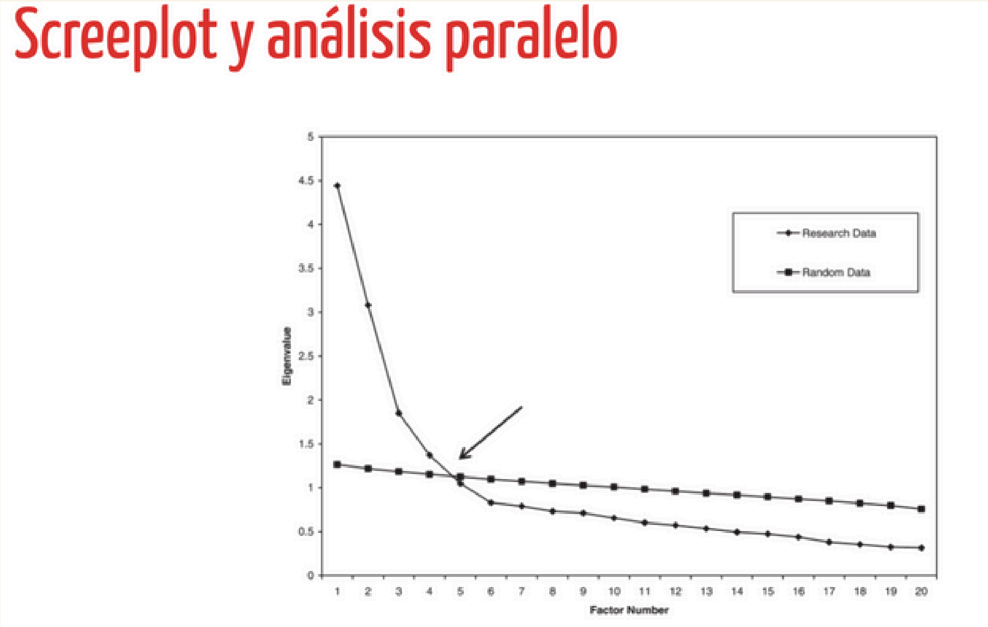
Supuestos a evaluar (2)

- Nivel de correlaciones de la matriz: test de esfericidad de Bartlett
  - se utiliza para evaluar la hipótesis que la matriz de correlaciones es una matriz identidad (en la diagonal=1 y bajo la diagonal=0)
  - se busca significación ($p < 0.05$), ya que se espera que las variables estén correlacionadas

Determinación de número de factores a considerar

Similar al PCA, existen diferentes métodos

- considerar los factores que tienen autovalores mayores a 1 (criterio de Kaiser)
- cómo del gráfico de sedimentación (graficar el % de varianza explicada por cada factor y retener tantos factores según cuando se quiebra la curva)
- “análisis paralelo”: comparación de los autovalores de la muestra con eigenvalues de datos aleatorios. Nº apropiado de factores: numero de eigenvalues de los datos reales que son mayores que sus correspondientes eigenvalues de datos aleatorios



Carrasco, Olivares y Nuñez, 2024

Métodos de extracción

Técnicas para determinar los factores / variables latentes a las variables observadas.

Los tres métodos principales son:

- Factores principales
- Factores principales iterados
- Maximum likelihood

Métodos de extracción

- Factores principales

Más común

Se basa en la descomposición de la matriz de correlaciones para identificar los factores que explican la mayor cantidad de varianza compartida por las variables.

Es útil cuando el objetivo es reducir la dimensionalidad manteniendo el máximo de información posible.

Métodos de extracción

- Factores principales iterados:

Variante del anterior.

Estima las comunalidades (la cantidad de varianza de cada variable explicada por los factores) iterativamente.

Reemplaza los valores iniciales de las comunalidades en la matriz de correlaciones con las comunalidades estimadas a partir de los factor loadings (cargas factoriales) y repite el proceso hasta que se alcance una solución estable.

Mejora la precisión de la estimación de los factores

Carrasco, Olivares y Nuñez, 2024

Métodos de extracción

- Maximum likelihood:

Busca encontrar los parámetros del modelo que maximicen la probabilidad de que los datos observados sean replicados por el modelo factorial.

Es útil cuando se quiere hacer inferencia estadística sobre los factores, ya que permite realizar pruebas de hipótesis y obtener intervalos de confianza para los factores y sus cargas.

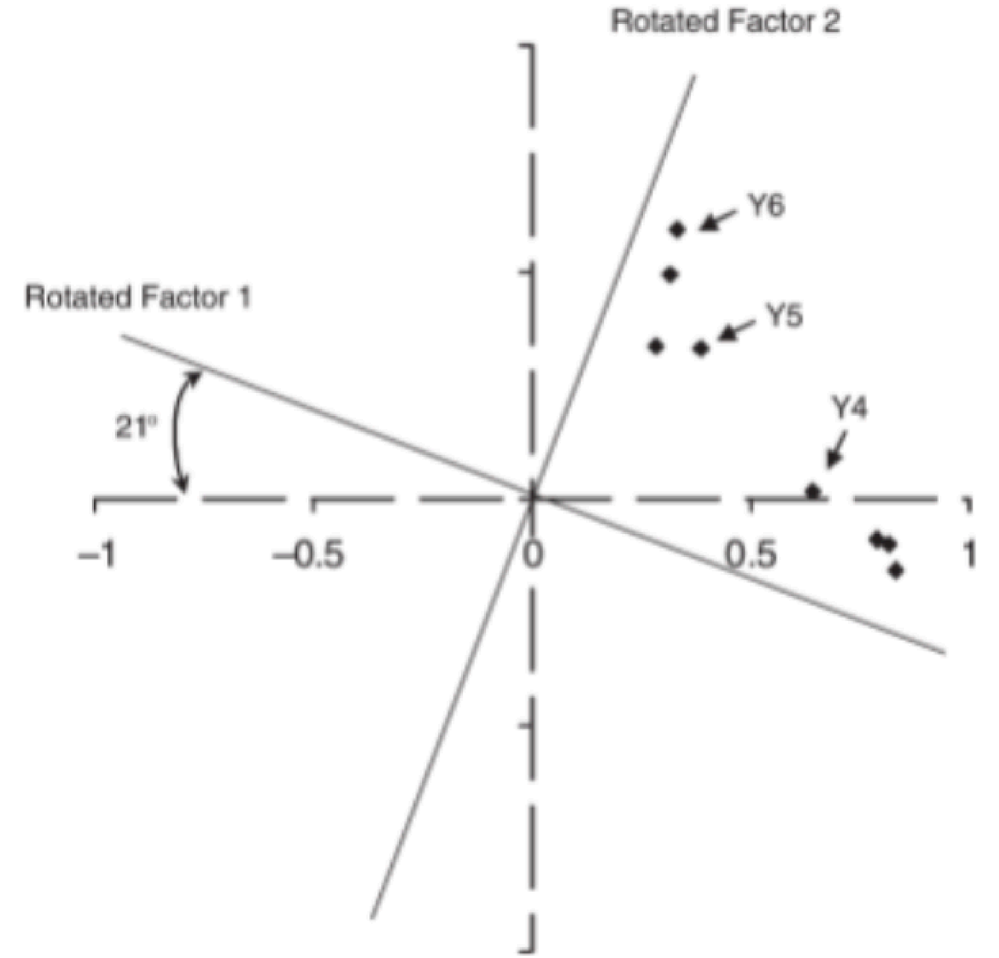
Es más robusto, pero requiere que los datos cumplan ciertos supuestos como normalidad multivariada.

Carrasco, Olivares y Nuñez, 2024

Rotación

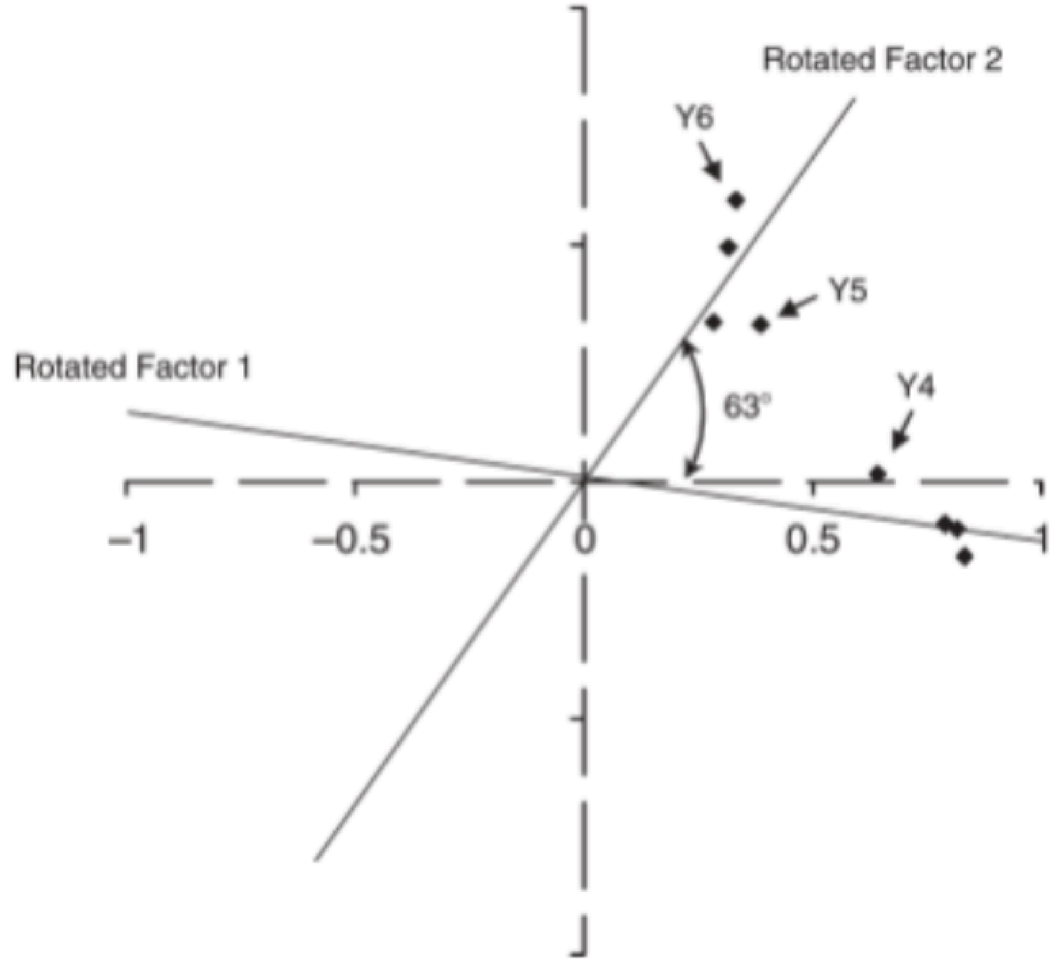
La rotación es una operación que se hace para mejorar la representación de los ejes para la interpretación.

B. Orthogonally Rotated Factor Matrix (Varimax)



	Factor 1	Factor 2
Y1	.836	.150
Y2	.794	.199
Y3	.767	.201
Y4	.594	.244
Y5	.242	.445
Y6	.098	.673
Y7	.114	.576
Y8	.145	.416

C. Obliquely Rotated Factor Matrix (Promax)



	Factor 1	Factor 2
Y1	.875	–.062
Y2	.817	.003
Y3	.788	.012
Y4	.588	.106
Y5	.154	.418
Y6	–.059	.704
Y7	–.018	.595
Y8	.055	.413

Carrasco, Olivares y Nuñez, 2024

Puntajes factoriales

Los puntajes factoriales son “estimaciones” (predicciones) de puntajes en los factores para cada observación en los datos.

Estos puntajes pueden utilizarse en análisis posteriores. Se pueden calcular puntajes para cada observación en cada factor utilizando un método de regresión.

Estas nuevas variables se estandarizan con media 0 y desviación estándar 1

Ejemplo. Confianza en las Instituciones (ELSOC)
Carrasco, Olivares y Nuñez, 2024
https://estadisticaiv.netlify.app/practicos/07-content

Se analiza si existen factores latentes a las preguntas sobre confianza en las instituciones en la Encuesta ELSOC. Se usa la ola de 2019. Se busca reducir el número de variables para explicar la confianza en las instituciones.

Se usan las siguientes variables de escala Likert:

- Grado de confianza en el El Gobierno
- Grado de confianza en los Partidos políticos
- Grado de confianza en los Carabineros
- Grado de confianza en os sindicatos
- Grado de confianza en las empresas privadas
- Grado de confianza en el congreso nacional
- Grado de confianza en el Presidente de la República

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)

Paso 1.

Ejemplo. Confianza en las Instituciones (ELSOC)

Paso 1.

Calcular la matriz de correlación

¿Para qué se hace esto?

Carrasco, Olivares y Nuñez, 2024

Paso 1. Calcular la matriz de correlación

▼ Código

```
data %>% select(conf_gob:conf_pres) %>% tab_corr(triangle = "lower")
```

	Grado de confianza: El Gobierno	*Grado de confianza: Los Partidos Politicos*	*Grado de confianza: Carabineros*	*Grado de confianza: Los Sindicatos*	*Grado de confianza: Las Empresas Privadas*	*Grado de confianza: El Congreso Nacional*	*Grado de confianza: El Presidente/a de la Republica*
Grado de confianza: El Gobierno							
Grado de confianza: Los Partidos Politicos	0.451***						
Grado de confianza: Carabineros	0.517***	0.232***					
Grado de confianza: Los Sindicatos	0.175***	0.280***	0.201***				
Grado de confianza: Las Empresas Privadas	0.417***	0.291***	0.409***	0.317***			
Grado de confianza: El Congreso Nacional	0.470***	0.523***	0.321***	0.331***	0.450***		
Grado de confianza: El Presidente/a de la Republica	0.719***	0.355***	0.524***	0.127***	0.378***	0.442***	

Computed correlation used pearson-method with listwise-deletion.

¿Cómo la interpretamos?

Carrasco, Olivares y Nuñez, 2024

Paso 1. Calcular la matriz de correlación

▼ Código

```
data %>% select(conf_gob:conf_pres) %>% tab_corr(triangle = "lower")
```

	Grado de confianza: El Gobierno	Grado de confianza: Los Partidos Politicos	Grado de confianza: Carabineros	Grado de confianza: Los Sindicatos	Grado de confianza: Las Empresas Privadas	Grado de confianza: El Congreso Nacional	Grado de confianza: El Presidente/a de la Republica
Grado de confianza: El Gobierno							
Grado de confianza: Los Partidos Politicos	0.451***						
Grado de confianza: Carabineros	0.517***	0.232***					
Grado de confianza: Los Sindicatos	0.175***	0.280***	0.201***				
Grado de confianza: Las Empresas Privadas	0.417***	0.291***	0.409***	0.317***			
Grado de confianza: El Congreso Nacional	0.470***	0.523***	0.321***	0.331***	0.450***		
Grado de confianza: El Presidente/a de la Republica	0.719***	0.355***	0.524***	0.127***	0.378***	0.442***	
Computed correlation used pearson-method with listwise-deletion.							

Todas las correlaciones son positivas y son significativas

Es dificil decir cómo se relacionan las variables entre si en conjunto

Hacemos un EFA

Carrasco, Olivares y Nuñez, 2024

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

▼ Código

```
corMat  <- data %>% select(conf_gob:conf_pres) %>%
  cor(use = "complete.obs")  # estimar matriz pearson

KMO(corMat)
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = corMat)
Overall MSA =  0.81
MSA for each item =
  conf_gob  conf_part conf_carab  conf_sind conf_empre  conf_cong  conf_pres
      0.78       0.81       0.86       0.80       0.86       0.83       0.77
```

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.

Carrasco, Olivares y Nuñez, 2024

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

▼ Código

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

```
$chisq
[1] 7942.816

$p.value
[1] 0

$df
[1] 21
```

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.

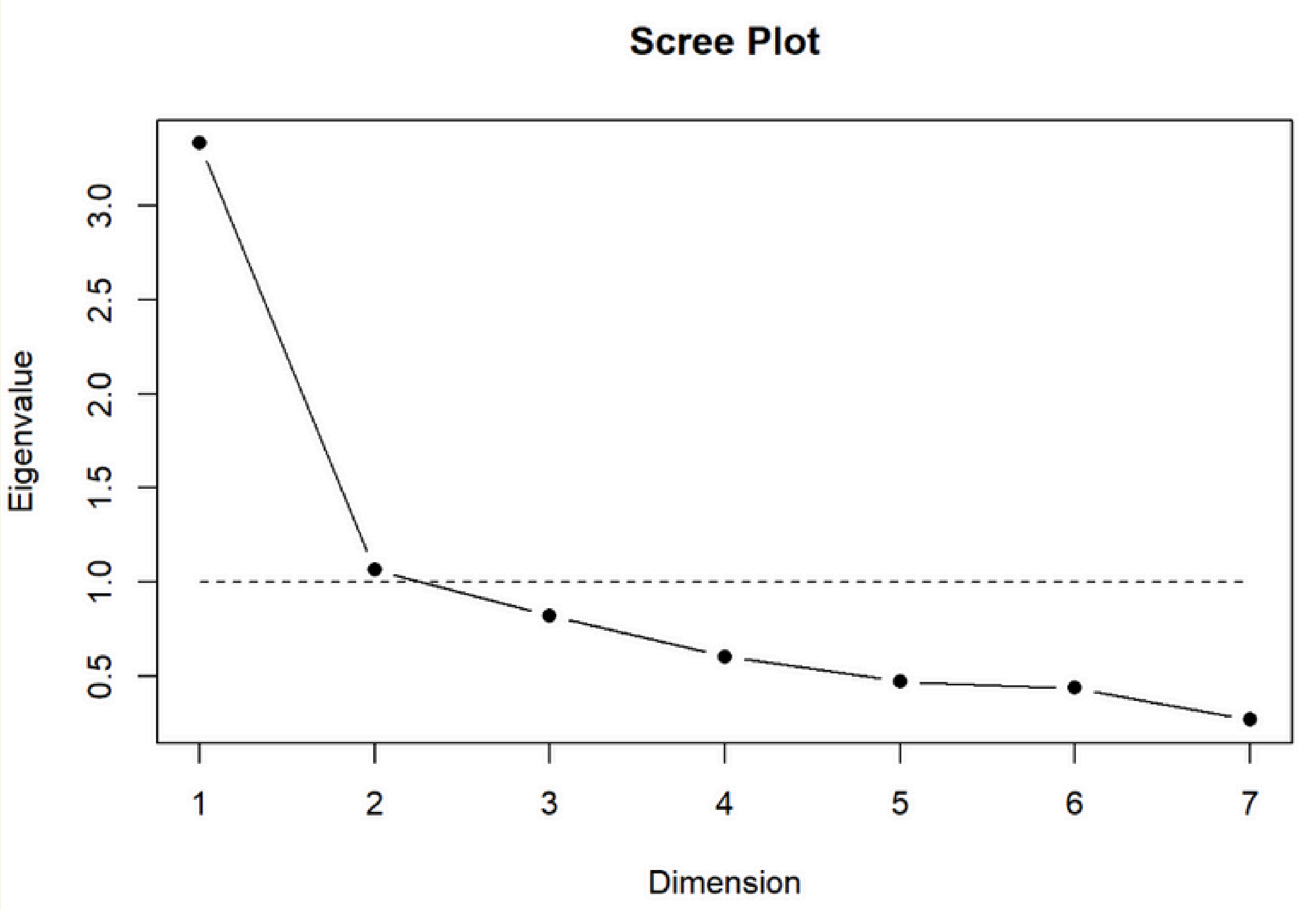
Carrasco, Olivares y Nuñez, 2024

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

Paso 3. Definición de número de factores. Método gráfico de sedimentación

▼ Código

```
data %>% select(conf_gob:conf_pres) %>% scree.plot()
```



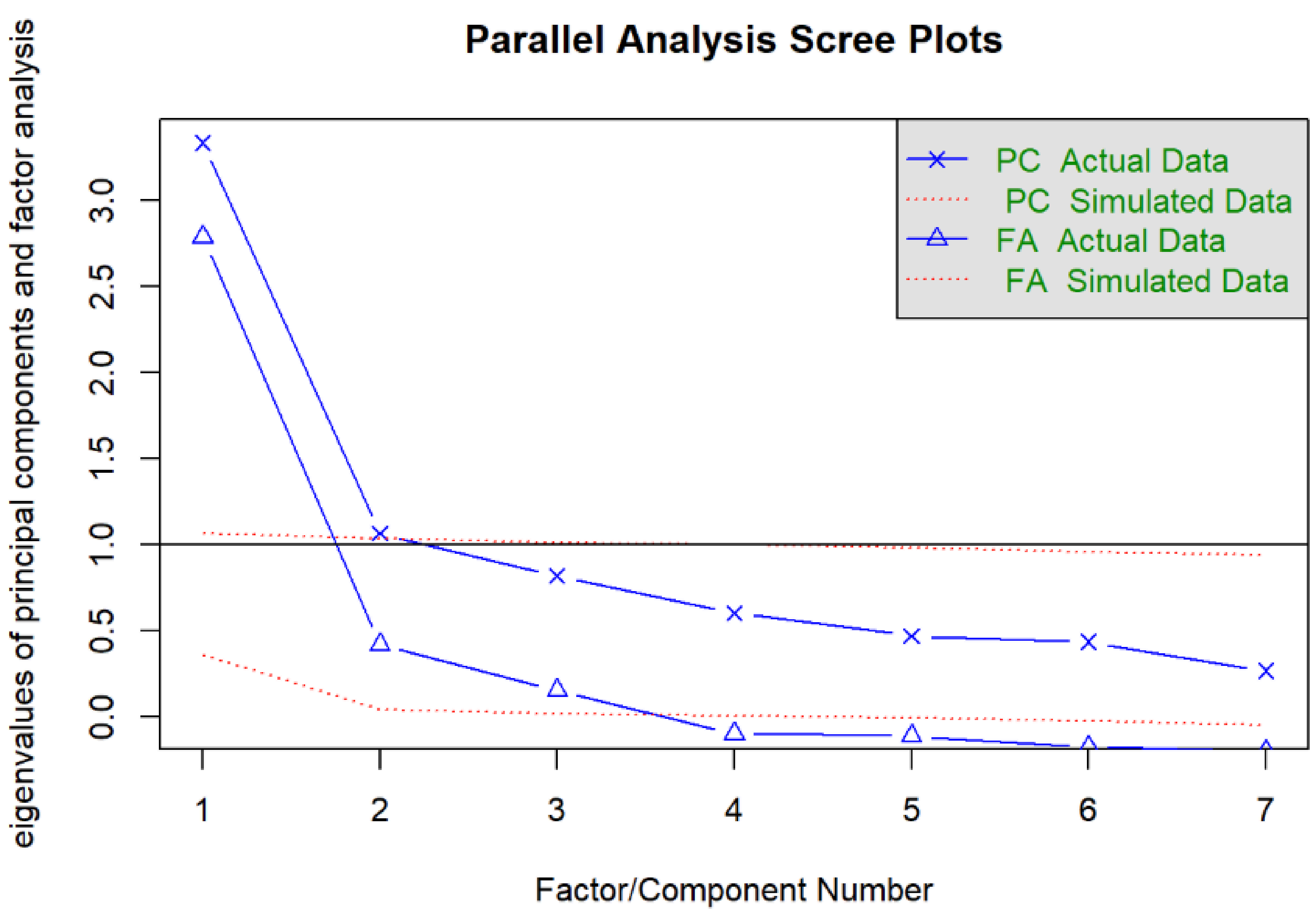
¿Qué nos dice este método?

Carrasco, Olivares y Nuñez, 2024

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

▼ Código

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




```
Parallel analysis suggests that the number of factors =  3  and the number of components =
```

¿Qué nos dice este otro método?

¿Con cuántos factores nos quedamos?

Carrasco, Olivares y Nuñez, 2024

Paso 4. Extracción de los factores con el método de componentes o factores principales. Considerando que se considerarán 3

▼ Código

```
fac_pa <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "pa")
#summary(fac_pa)
fac_pa
```

Paso 4. Extracción de los factores con el método de componentes o factores principales. Considerando que se considerarán 3

```
Factor Analysis using method =  pa
Call: fa(r = ., nfactors = 3, fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
             PA1   PA2   PA3   h2   u2 com
conf_gob    0.77  0.14  0.00 0.72 0.28 1.1
conf_part   0.07  0.72 -0.04 0.55 0.45 1.0
conf_carab  0.52 -0.13  0.32 0.45 0.55 1.8
conf_sind  -0.19  0.29  0.42 0.27 0.73 2.2
conf_empre  0.14  0.07  0.59 0.50 0.50 1.1
conf_cong   0.12  0.53  0.23 0.55 0.45 1.5
conf_pres   0.87  0.00 -0.02 0.74 0.26 1.0

                       PA1  PA2  PA3
SS loadings           1.84 1.08 0.87
Proportion Var        0.26 0.15 0.12
Cumulative Var        0.26 0.42 0.54
Proportion Explained  0.48 0.29 0.23
Cumulative Proportion 0.48 0.77 1.00
```

Carrasco, Olivares y Nuñez, 2024

Paso 4. Extracción de los factores con el método de Maximum Likelihood. Considerando que se considerarán 3

▼ Código

```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml")
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.77  0.13  0.02 0.72 0.28 1.1
conf_part   0.05  0.78 -0.01 0.64 0.36 1.0
conf_carab  0.51 -0.12  0.29 0.43 0.57 1.7
conf_sind  -0.19  0.24  0.46 0.27 0.73 1.9
conf_empre  0.11  0.00  0.66 0.53 0.47 1.1
conf_cong   0.14  0.44  0.30 0.52 0.48 2.0
conf_pres   0.88 -0.01 -0.03 0.75 0.25 1.0

                       ML1  ML2  ML3
SS loadings           1.83 1.02 1.00
Proportion Var        0.26 0.15 0.14
Cumulative Var        0.26 0.41 0.55
Proportion Explained  0.47 0.27 0.26
Cumulative Proportion 0.47 0.74 1.00
```

Carrasco, Olivares y Nuñez, 2024

Paso 5. Rotación ortogonal (Varimax)

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "varimax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
            ML1  ML2  ML3   h2   u2 com
conf_gob   0.76 0.32 0.17 0.72 0.28 1.4
conf_part  0.21 0.74 0.22 0.64 0.36 1.3
conf_carab 0.57 0.05 0.31 0.43 0.57 1.6
conf_sind  0.04 0.23 0.46 0.27 0.73 1.5
conf_empre 0.35 0.11 0.63 0.53 0.47 1.6
conf_cong  0.34 0.48 0.42 0.52 0.48 2.8
conf_pres  0.83 0.22 0.10 0.75 0.25 1.2

                       ML1  ML2  ML3
SS loadings           1.89 1.00 0.97
Proportion Var        0.27 0.14 0.14
Cumulative Var        0.27 0.41 0.55
Proportion Explained  0.49 0.26 0.25
Cumulative Proportion 0.49 0.75 1.00
```

```
fac_ml_var <- data %>%
select(conf_gob:conf_pres) %>% fa(nfactors = 3,
fm= "ml", rotate="varimax")
```

Carrasco, Olivares y Nuñez, 2024

Paso 5. Rotación oblicua (proman)

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.80  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03  0.78  0.01 0.64 0.36 1.0
conf_carab  0.57 -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13  0.51 0.27 0.73 1.3
conf_empre  0.19 -0.12  0.67 0.53 0.47 1.2
conf_cong   0.17  0.39  0.30 0.52 0.48 2.3
conf_pres   0.91  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```

```
fac_ml_pro <- data %>%
select(conf_gob:conf_pres) %>% fa(nfactors = 3,
fm= "ml", rotate="promax")fac_ml_pro
```

Carrasco, Olivares y Nuñez, 2024

Paso 6. Puntajes factoriales (con 3 factores y rotación ortogonal)

Acá tenemos la posibilidad de encontrar una interpretación!!!

▼ Código

```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml", rotate = "promax'
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.80  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03  0.78  0.01 0.64 0.36 1.0
conf_carab  0.57 -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13  0.51 0.27 0.73 1.3
conf_empre  0.19 -0.12  0.67 0.53 0.47 1.2
conf_cong   0.17  0.39  0.30 0.52 0.48 2.3
conf_pres   0.91  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```

ML1 (Factor 1) = confianza en el ejecutivo (gobierno + presidente + carabineros)

ML2 (Factor 2) = confianza en los partidos (partidos + congreso)

ML3 (Factor 3) = confianza en instituciones del trabajo (empresas + sindicatos)

Carrasco, Olivares y Nuñez, 2024

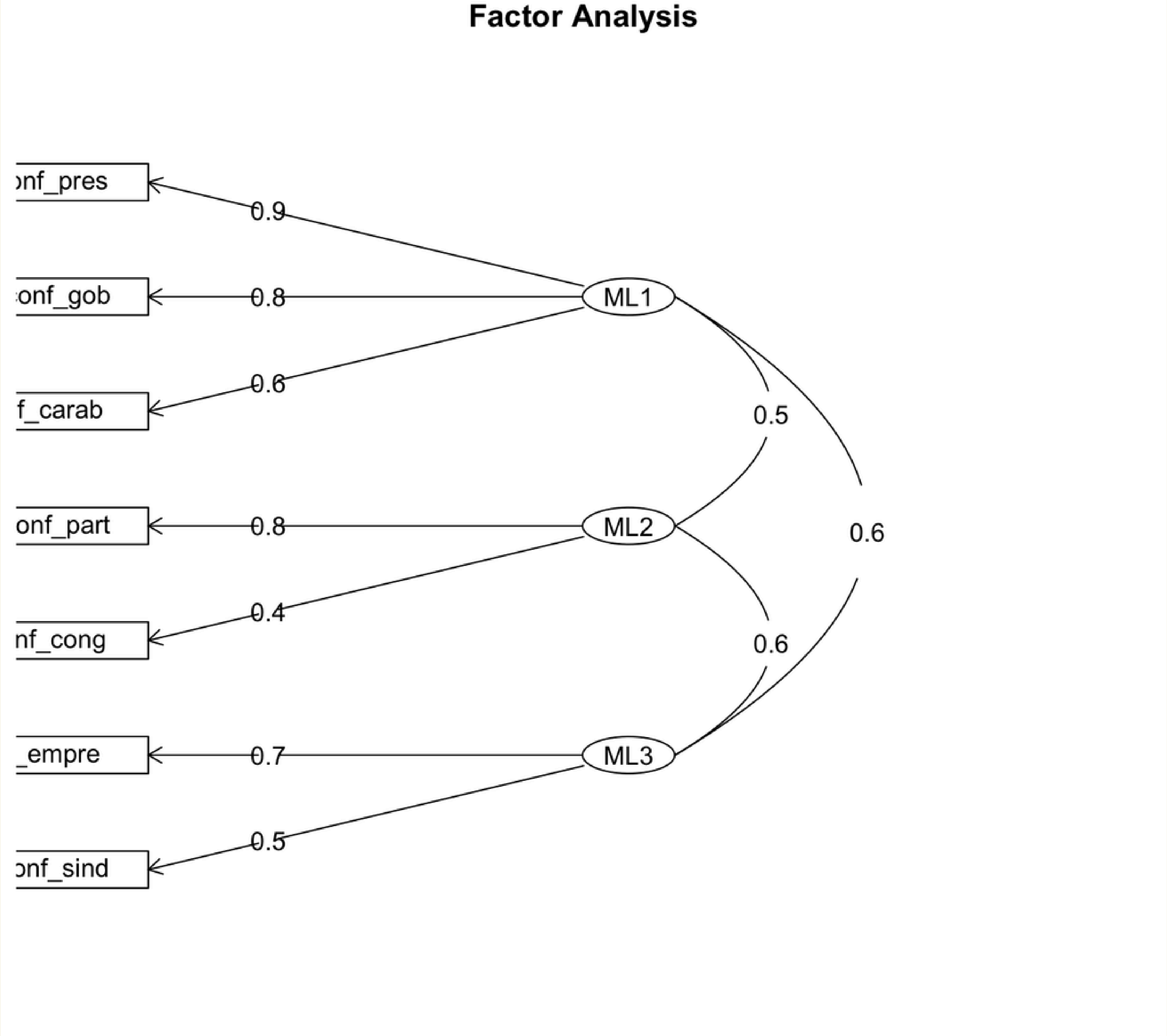
Paso 6. Puntajes factoriales

▼ Código

```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml", rotate = "promax"
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.80  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03  0.78  0.01 0.64 0.36 1.0
conf_carab  0.57 -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13  0.51 0.27 0.73 1.3
conf_empre  0.19 -0.12  0.67 0.53 0.47 1.2
conf_cong   0.17  0.39  0.30 0.52 0.48 2.3
conf_pres   0.91  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```




```
fa.diagram(fac_ml)
```

Paso 6. Puntajes factoriales

▼ Código

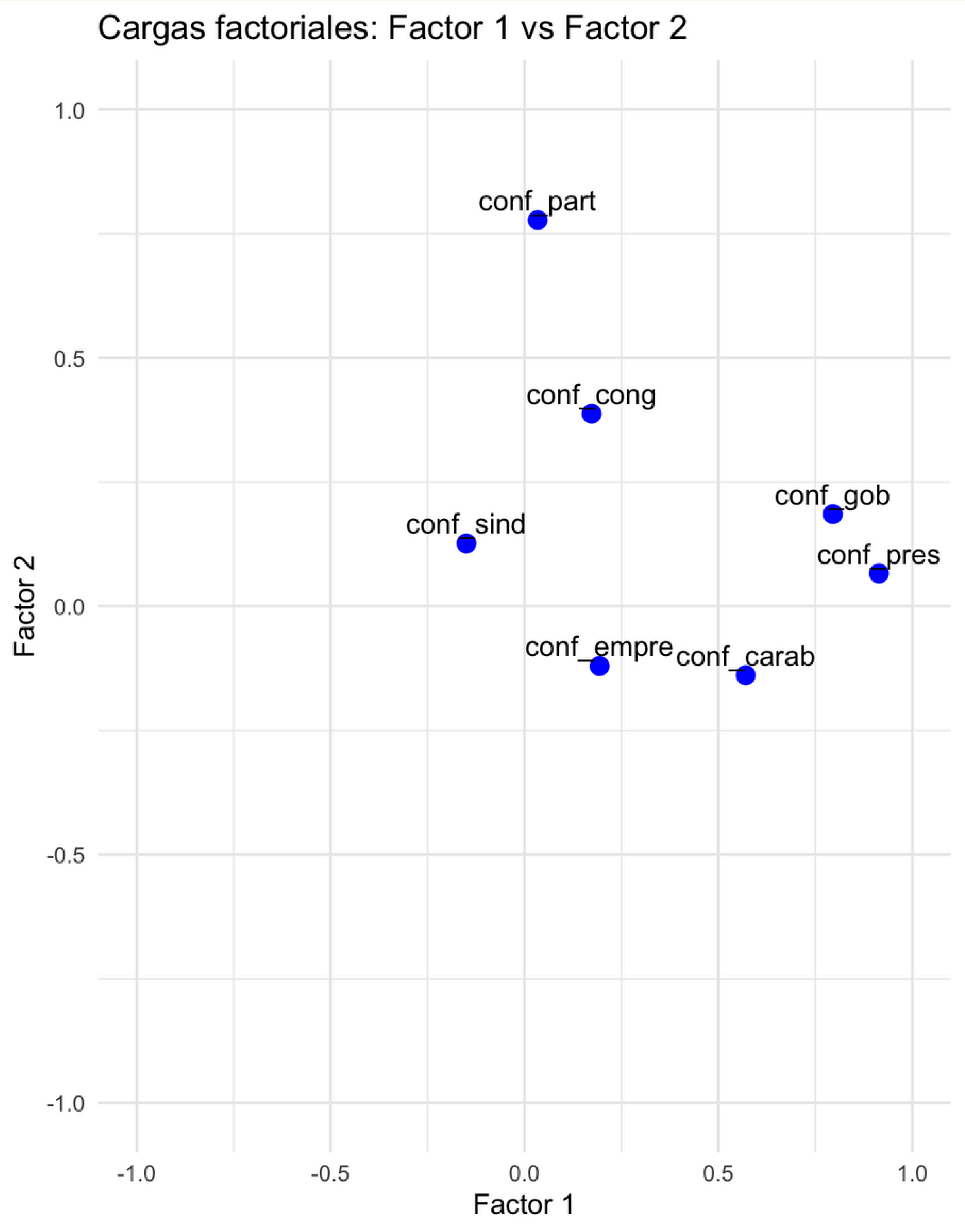
```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml", rotate = "promax
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.80  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03  0.78  0.01 0.64 0.36 1.0
conf_carab  0.57 -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13  0.51 0.27 0.73 1.3
conf_empre  0.19 -0.12  0.67 0.53 0.47 1.2
conf_cong   0.17  0.39  0.30 0.52 0.48 2.3
conf_pres   0.91  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```

```
loadings$variable <- rownames(loadings)

ggplot(loadings, aes(x = ML1, y = ML2, label = variable)) +
  geom_point(color = "blue", size = 3) +
  geom_text(vjust = -0.5) +
  xlim(-1, 1) + ylim(-1, 1) +
  labs(title = "Cargas factoriales: Factor 1 vs Factor 2",
       x = "Factor 1", y = "Factor 2") +
  theme_minimal()
```



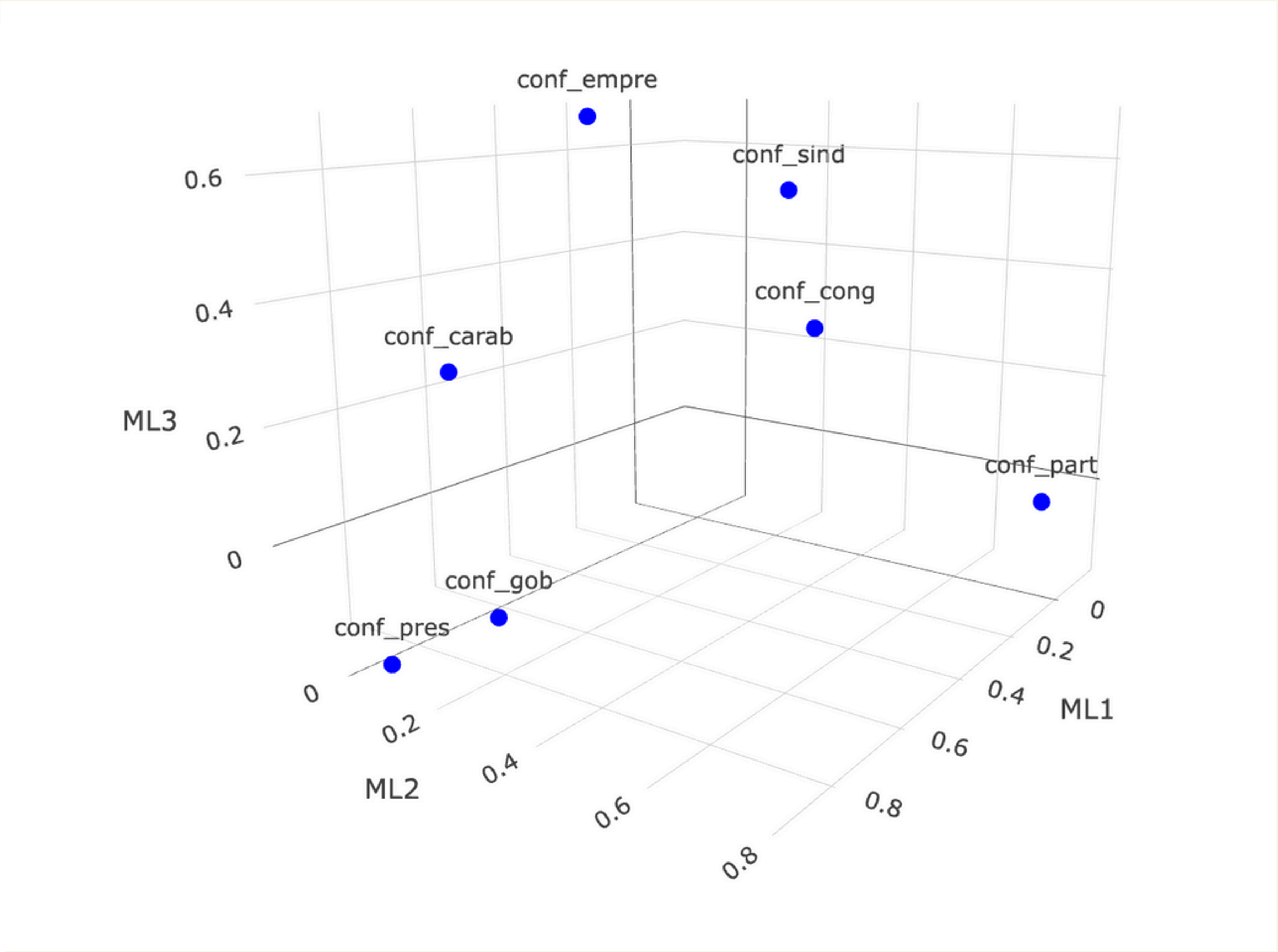
Paso 6. Puntajes factoriales

▼ Código

```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml", rotate = "promax"
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob    0.80  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03  0.78  0.01 0.64 0.36 1.0
conf_carab  0.57 -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13  0.51 0.27 0.73 1.3
conf_empre  0.19 -0.12  0.67 0.53 0.47 1.2
conf_cong   0.17  0.39  0.30 0.52 0.48 2.3
conf_pres   0.91  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```




```
plot_ly(loadings, x = ~ML1, y = ~ML2, z = ~ML3,
        text = ~variable, type = "scatter3d", mode =
"markers+text",
        textposition = "top center",
        marker = list(size = 5, color = "blue"))
```

Paso 6. Puntajes factoriales

Paso 6. Puntajes factoriales

▼ Código

```
fac_ml <- data %>% select(conf_gob:conf_pres) %>% fa(nfactors = 3, fm= "ml", rotate = "promax
fac_ml
```

```
Factor Analysis using method =  ml
Call: fa(r = ., nfactors = 3, rotate = "promax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
             ML1   ML2   ML3   h2   u2 com
conf_gob   [illegible]  0.19 -0.09 0.72 0.28 1.1
conf_part   0.03 [illegible]  0.01 0.64 0.36 1.0
conf_carab [illegible] -0.14  0.23 0.43 0.57 1.4
conf_sind  -0.15  0.13 [illegible] 0.27 0.73 1.3
conf_empre  0.19 -0.12 [illegible] 0.53 0.47 1.2
conf_cong   0.17 [illegible]  0.30 0.52 0.48 2.3
conf_pres  [illegible]  0.07 -0.16 0.75 0.25 1.1

                       ML1  ML2  ML3
SS loadings           1.98 0.96 0.92
Proportion Var        0.28 0.14 0.13
Cumulative Var        0.28 0.42 0.55
Proportion Explained  0.51 0.25 0.24
Cumulative Proportion 0.51 0.76 1.00
```

Factor 1 (ML1)

$$F1=0.80 \cdot conf\_gob+0.03 \cdot conf\_part+0.57 \cdot conf\_carab-0.15 \cdot conf\_sind+0.19 \cdot conf\_empre+0.17 \cdot conf\_cong+0.91 \cdot conf\_pres$$

Factor 2 (ML2)

$$F2=0.19 \cdot conf\_gob+0.78 \cdot conf\_part-0.14 \cdot conf\_carab+0.13 \cdot conf\_sind-0.12 \cdot conf\_empre+0.39 \cdot conf\_cong+0.07 \cdot conf\_pres$$

Factor 3 (ML3)

$$F3=-0.09 \cdot conf\_gob+0.01 \cdot conf\_part+0.23 \cdot conf\_carab+0.51 \cdot conf\_sind+0.67 \cdot conf\_empre+0.30 \cdot conf\_cong-0.16 \cdot conf\_pres$$

Ejercicio Módulo 2

Replicar análisis presentado en clases utilizando algún otro módulo de la encuesta ELSOC. Seguir los códigos utilizados en la versión de 2024 del curso; https://estadisticaiv.netlify.app/practicos/06-content
Entregar el script interpretación de los siguientes pasos

1. Generar la matriz de correlaciones entre las variables observadas
2. Realizar el test de adecuación de la matriz de correlaciones de variables observadas e interpretar el resultado
3. Realizar el test de esfericidad de Barlett, interpretar
4. Seleccionar el número de factores por ambos método (interpretar)
5. Realizar el cálculo de los factores con los métodos componentes principales y de Maximum likelihood e interpretar
6. Calcular los puntajes factoriales e interpretar interpretando los factores

```
## Factor Analysis using method =  ml
## Call: fa(r = fa, nfactors = 2, scores = "regression", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                         ML1   ML2   h2   u2 com
## son_refinadas          0.01  0.62 0.38 0.62   1
## ser_protegidas        -0.01  0.58 0.34 0.66   1
## consiguen_privilegios  0.67  0.03 0.47 0.53   1
## quejan_discriminacion  0.68 -0.03 0.44 0.56   1
##
##                        ML1  ML2
## SS loadings           0.90 0.72
```

```
##    benevolente_prom hostil_prom         ML1         ML2
## 1               4.5         4.0  0.62586434  0.65986183
## 2               4.5         3.5  0.25445615  0.67328242
## 3               4.5         4.0  0.62498033  0.69225511
## 4               3.5         3.5 -0.05380749 -0.36360884
## 5               4.0         4.0  0.53314791  0.22298473
## 6               4.0         4.0  0.53314791  0.22298473
## 7               4.0         3.0 -0.33161523  0.05429558
## 8               4.0         3.0 -0.33161523  0.05429558
## 9               4.0         4.0  0.53314791  0.22298473
## 10              4.0         3.0 -0.39347262 -0.01107631
```

Factor 1

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.80036 -0.45533  0.19260  0.00396  0.53315  1.64432
```

Factor 2

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.19764 -0.36361  0.13979  0.00049  0.45705  1.36319
```

Análisis de componentes principales

02

Clase hoy

- lógica matmética con ejemplo simple

Caso simulado de notas en diferentes ramos

Ejemplo medición inteligencia

Matriz de correlaciones de las notas de 220 estudiantes en 6 asignaturas

	Español	Inglés	Historia	Cálculo	Algebra	Geometría
Español	1					
Inglés	0,439	1				
Historia	0,41	0,351	1			
Cálculo	0,288	0,354	0,164	1		
Algebra	0,329	0,32	0,19	0,595	1	
Geometría	0,248	0,329	0,181	0,47	0,464	1

Esas notas pueden estar causadas por un único factor subyacente, que sería lo que podría denominarse inteligencia general, o existen varios factores que pueden explicar distintos tipos de inteligencia.

Fuente: Lawley y Maxwell (1971) en Aldás y Uriel, 2017 https://estadisticaiv.netlify.app/practicos/06-content