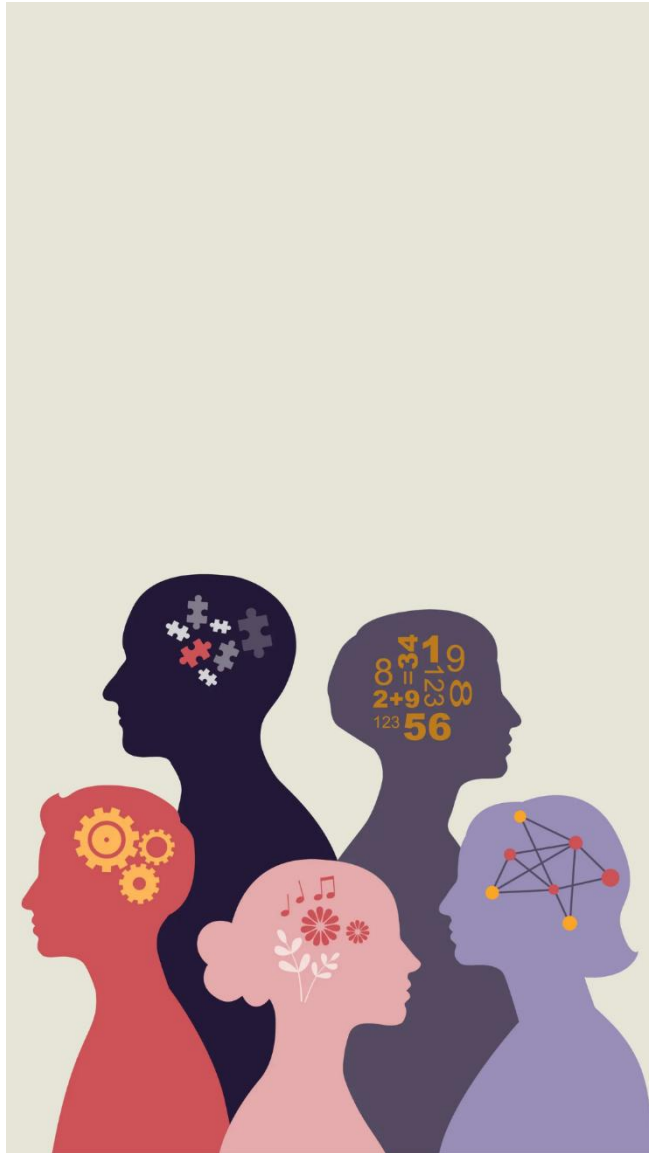




Práctica 1 : análisis factorial

ENUNCIADO 116 – PERSONALIDAD 16
TÉCNICAS MULTIVARIANTES 1

Manuela Ibáñez | Ángel Vega | octubre 2025

Contenido

Enunciado de la práctica.....	2
1. Preparación de los datos	3
a) Carga de datos, examinar correlaciones / covarianzas, detectar variables problemáticas y decidir qué variables entran al modelo y cuáles se dejan como suplementarias.	4
a.1) Variables que se incluyen en el modelo	4
a.2) Análisis de la matriz de correlaciones o de covarianzas	4
a.3) Determinar la estructura y si se elimina alguna de las variables se debe utilizar como suplementarias	6
a.4) Depuración y separación de suplementarias.....	8
2. Análisis factorial.....	9
b) Llevar a cabo análisis factorial por el método de las componentes principales, determinando y justificando la elección del número de factores elegido.”.	9
b.1) Aplicar el método de las componentes principales	9
b.2) Reestimación fijando $k = 5$ y reporte del modelo sin rotar	11
3. Rotación e interpretación de factores	13
c) Rotación de la solución y determinación de las variables más relacionadas con cada factor.....	13
4. Creación de nuevos conjuntos de datos.....	14
d) Puntuaciones factoriales (proyección de las observaciones).....	14
e) Proyección de las variables sobre los ejes factoriales (coordenadas factoriales)	15
5. Representaciones gráficas	16
f) Representación gráfica.....	16
f.1) Factores 1 y 2.....	17
f.2) Factores 3 y 4.....	18
f.3) Factores 1 y 5.....	19
6. Bondad del ajuste	20
g) Matriz de correlaciones residuales	20
7. Variables suplementarias.....	21
h) Proyectar las variables suplementarias generadas en el apartado a sobre los ejes factoriales. Representarlas junto al resto de variables en los planos generados en el apartado f.....	21

Enunciado de la práctica

El conjunto de datos Personalidad16 contiene respuestas a cuestionario de personalidad de 20000 participantes. Se trabaja con 50 variables: Se pide Realizar análisis Factorial siguiendo la sistemática identificada en clase, determinando:

- a) Variables que se incluyen en el modelo, previo análisis de la matriz de correlaciones o de covarianzas (elegir). Determinar la estructura y si eliminamos alguna de las variables debéis luego utilizarlas como suplementarias.
- b) Llevar a cabo análisis factorial por el método de las componentes principales, determinando y justificando la elección del número de factores elegido.
- c) Rotar la solución y determinar el conjunto de variables que se encuentran más relacionadas con cada factor. A partir de ahí, proponer un nombre para cada uno de los factores.
- d) Crear un conjunto de datos que contenga las proyecciones de todas las observaciones sobre los factores elegidos.
- e) Crear otro conjunto de datos que contenga las proyecciones de las variables sobre los ejes determinados por cada factor
- f) Si hemos retenido k factores y k es par, Representarlos gráficamente para las variables en los planos generados por los factores 1-2 3-4,..., $(k-1)$ - k . Si es impar el último cruzarlo con el 1º
- g) Examinar la matriz de correlaciones residuales, si veis que el modelo no es suficientemente bueno rectificarlo. Explicar por qué la solución final es buena o no.
- h) Proyectar las variables suplementarias generadas en el apartado a sobre los ejes factoriales. Representarlas junto al resto de variables en los planos generados en el apartado f.

1. Preparación de los datos

En este ejercicio vamos a trabajar con una base de datos llamada personalidad16, que contiene un total de 51 variables y 20.000 observaciones. Los datos que contiene corresponden a un cuestionario de personalidad en el que cada punto refleja un comportamiento en una escala del 0 al 5.

Las variables son de tipo numérico y recogen y agrupan las respuestas. Algunas agrupaciones son:

- **Rasgos de extroversión:** alma_fiesta, hablo_muchas_personas, empiezo_conversaciones, comodo_con_gente, nohablo, callado, poco_que_decir, etc.
- **Rasgos de amabilidad y empatía:** simpatizo_sentimientos, corazon_blando, emocioono_con_otros, interes_personas, preocupo_por_demas, hago_gente_sienta_gusto, frente a ítems como insulto_gente, no_interes_demas, no_interesan_problemas_otros.
- **Responsabilidad y orden:** atento_detalle, gusta_orden, sigo_horario, hago_tareas_rapido, estoy_preparado, en contraposición con dejo_pertenencias_porahi, desastre_cosas, olvido_donde_dejocosas, eludo_deberes.
- **Estabilidad emocional/neuroticismo:** estresofacil, cambio_estado_animo, cambio_humor, irritado_facil, me_enojo, me_perturban, siento_triste, preocupado, junto a ítems de carácter opuesto como no_triste o relajado.
- **Apertura a la experiencia:** imaginacion_vivida, tengo_grandes_ideas, uso_palabras_dificiles, vocabulario_rico, reflexiono_cosas, estoy_lleno_ideas, entiendo_rapido; frente a variables como dificultades_ideas_abstractas, no_interes_ideas_abstractas, no_soy_imaginativo.
- Finalmente, la **variable N** representa un índice global y es borrado por ello.

Todo esto son escalas que miden cuanta relación tiene los individuos con cada punto, permitiendo aplicar un análisis factorial para detectar dimensiones que estructuren la personalidad

a) Carga de datos, examinar correlaciones / covarianzas, detectar variables problemáticas y decidir qué variables entran al modelo y cuáles se dejan como suplementarias.

a.1) Variables que se incluyen en el modelo

```
/* DATASET */
proc contents data=mydata.personalidad16;
run;
```

Alphabetic List of Variables and Attributes			
#	Variable	Type	Len
11	Estreso_facil	Num	8
51	N	Num	8
24	Simpatizo_sentimientos	Num	8
1	alma_fiesta	Num	8
33	atento_detalle	Num	8
10	callado	Num	8
17	cambio_estado_animo	Num	8
18	cambio_humor	Num	8
3	comodo_con_gente	Num	8
26	corazon_blando	Num	8
32	dejo_pertenencias_porahi	Num	8
34	desastre_cosas	Num	8
42	dificultades_ideas_abstractas	Num	8
38	eludo_deberes	Num	8
29	emociono_con_otros	Num	8
5	empiezo_conversaciones	Num	8
47	entiendo_rapido	Num	8
50	estoy_lleno_ideas	Num	8

50	estoy_lleno_ideas	Num	8
31	estoy_preparado	Num	8
40	exigente_trabajo	Num	8
37	gusta_orden	Num	8
7	hablo_muchas_personas	Num	8
30	hago_gente_sienta_gusto	Num	8
35	hago_tareas_rapido	Num	8
43	imaginacion_vivida	Num	8
23	insulto_gente	Num	8
22	interes_personas	Num	8
19	irrito_facil	Num	8
4	mantego_2_plano	Num	8
16	me_enojo	Num	8
15	me_perturban	Num	8
27	no_interes_demas	Num	8
44	no_interes_ideas_abstractas	Num	8
25	no_interesan_problemas_otros	Num	8
8	no_llamo_atencion	Num	8
9	no_soy_centro_atencion	Num	8
46	no_soy_imaginativo	Num	8
14	no_triste	Num	8

Dado que la variable N se interpreta como un total, hemos decidido eliminarla de la base de ya que puede inducir colinealidad y sesgo.

```
data personalidad16_sinN;
set mydata.personalidad16;
drop N;
run;
```

a.2) Análisis de la matriz de correlaciones o de covarianzas

Analizando la matriz de correlaciones estamos buscando correlaciones altas entre variables (que podrían pertenecer a un factor común), problemas de multicolinealidad cuando las

correlaciones son muy elevadas o variables con correlaciones muy bajas que aportan poca información y que acabaremos tratando como suplementarias.

```
/*a.2) Matriz de correlaciones y covarianzas */
proc corr data=personalidadl6_sinN cov plots=matrix(histogram);
    var _all_;
run;
```

Extracto de la matriz de covarianzas: la diagonal contiene las varianzas de cada variable y los valores fuera de la diagonal son covarianzas que miden cómo varían conjuntamente dos variables en sus unidades originales:

$$S = \frac{1}{n-1} (X - \bar{X})' (X - \bar{X})$$

	alma_fiesta	nohablo	comodo_con_gente	mantego_2_plano	empiezo_conversaciones	poco_que_decir	hablo_muchas_personas	no_llamo_atencion
alma_fiesta	1.596302855	-0.710279114	0.705338547	-0.706631732	0.750377159	-0.511611301	1.024313656	-0.554498
nohablo	-0.710279114	1.785288264	-0.657193060	0.851818591	-0.850586379	0.875762338	-0.826074654	0.638147
comodo_con_gente	0.705338547	-0.657193060	1.514167668	-0.619275764	0.862473854	-0.483248952	0.905863123	-0.420543
mantego_2_plano	-0.706631732	0.851818591	-0.619275764	1.528700435	-0.679636382	0.665428471	-0.751587979	0.728112
empiezo_conversaciones	0.750377159	-0.850586379	0.862473854	-0.679636382	1.685190437	-0.641983182	1.058459525	-0.461987
poco_que_decir	-0.511611301	0.875762338	-0.483248952	0.665428471	-0.641983182	1.586661231	-0.593420529	0.515285
hablo_muchas_personas	1.024313656	-0.826074654	0.905863123	-0.751587979	1.058459525	-0.593420529	1.975595557	-0.527215
no_llamo_atencion	-0.554498885	0.638147557	-0.420543397	0.728112006	-0.461987847	0.515285957	-0.527215433	1.677007
no_soy_centro_atencion	0.729055413	-0.579590380	0.644997970	-0.643205760	0.673734797	-0.442073634	0.765169068	-0.736827
callado	-0.632290895	0.775538977	-0.625265223	0.796894645	-0.730975779	0.616251103	-0.806961678	0.658491
Estreso_facil	-0.155205000	0.110758388	-0.308504355	0.255131157	-0.178032929	0.174980382	-0.218104858	0.201896
relajado	0.220327696	0.011728136	0.408907455	-0.068342217	0.236327884	0.009047900	0.269170251	-0.012783
preocupado	-0.139530617	0.100221361	-0.169889724	0.273156058	-0.072113158	0.123184867	-0.142503353	0.206882
no_triste	0.181567278	-0.059434722	0.277250013	-0.108067403	0.203885457	-0.023577216	0.228945085	-0.049113
me_perturban	-0.075422291	0.075674334	-0.208754828	0.193102855	-0.093925579	0.190726834	-0.117812548	0.114302
me_enojo	-0.107702585	0.050588029	-0.264541127	0.216462823	-0.126169633	0.178485399	-0.159144032	0.126480
cambio_estado_animo	-0.030691255	0.033442722	-0.242677424	0.199245162	-0.099741945	0.149114478	-0.076941830	0.084514
cambio_humor	-0.037029331	0.040383969	-0.290704645	0.203376969	-0.130606023	0.164059530	-0.106945315	0.070237

Extracto de la matriz de correlaciones: las correlaciones son coeficientes de Pearson estandarizados, la diagonal vale siempre 1 (correlación perfecta consigo misma), los valores negativos indican relación inversa, los positivos relación directa, el p-valor (<.0001 en casi todos los casos) indica si la correlación es estadísticamente significativa.

$$R = D^{-\frac{1}{2}} S D^{-\frac{1}{2}}$$

SAS obtiene las correlaciones individuales con el coeficiente de correlación de Pearson:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

	alma_fiesta	nohablo	comodo_con_gente	mantego_2_plano	empiezo_conversaciones	poco_que_decir	hablo_muchas_personas	no_llamo_atencion
alma_fiesta	1.00000	-0.42074 <.0001	0.45368 <.0001	-0.45235 <.0001	0.45751 <.0001	-0.32147 <.0001	0.57680 <.0001	-0.33890 <.0001
nohablo	-0.42074 <.0001	1.00000	-0.39972 <.0001	0.51562 <.0001	-0.49039 <.0001	0.52034 <.0001	-0.43986 <.0001	0.36881 <.0001
comodo_con_gente	0.45368 <.0001	-0.39972 <.0001	1.00000	-0.40704 <.0001	0.53993 <.0001	-0.31178 <.0001	0.52375 <.0001	-0.26391 <.0001
mantego_2_plano	-0.45235 <.0001	0.51562 <.0001	-0.40704 <.0001	1.00000	-0.42344 <.0001	0.42727 <.0001	-0.43248 <.0001	0.45475 <.0001
empiezo_conversaciones	0.45751 <.0001	-0.49039 <.0001	0.53993 <.0001	-0.42344 <.0001	1.00000	-0.39261 <.0001	0.58010 <.0001	-0.27481 <.0001
poco_que_decir	-0.32147 <.0001	0.52034 <.0001	-0.31178 <.0001	0.42727 <.0001	-0.39261 <.0001	1.00000	-0.33517 <.0001	0.31589 <.0001
hablo_muchas_personas	0.57680 <.0001	-0.43986 <.0001	0.52375 <.0001	-0.43248 <.0001	0.58010 <.0001	-0.33517 <.0001	1.00000	-0.28965 <.0001
no_llamo_atencion	-0.33890 <.0001	0.36881 <.0001	-0.26391 <.0001	0.45475 <.0001	-0.27481 <.0001	0.31589 <.0001	-0.28965 <.0001	1.00000
no_soy_centro_atencion	0.42392 <.0001	-0.31868 <.0001	0.38508 <.0001	-0.38219 <.0001	0.38128 <.0001	-0.25783 <.0001	0.39994 <.0001	-0.41801 <.0001
callado	-0.37982 <.0001	0.44052 <.0001	-0.38565 <.0001	0.48917 <.0001	-0.42736 <.0001	0.37131 <.0001	-0.43573 <.0001	0.38592 <.0001
Estreso_facil	-0.09084 <.0001	0.06130 <.0001	-0.18540 <.0001	0.15259 <.0001	-0.10142 <.0001	0.10273 <.0001	-0.11475 <.0001	0.11529 <.0001
relajado	0.14041 <.0001	0.00707 <.0001	0.26755 <.0001	-0.04450 <.0001	0.14658 <.0001	0.00578 <.0001	0.15419 <.0001	-0.00795 <.0001
preocupado	-0.09678	0.06573	-0.12099	0.19360	-0.04868	0.08570	-0.08885	0.14000

Extracto de estadísticos descriptivos: número de observaciones disponibles (20.000 en todas), la media, la desviación típica, la suma total de valores y los mínimos y máximos observados, que en este caso oscilan entre 0 y 5 al tratarse de una escala tipo Likert.

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
alma_fiesta	20000	2.62640	1.26345	52528	0	5.00000
nohablo	20000	2.79900	1.33615	55980	0	5.00000
comodo_con_gente	20000	3.22980	1.23052	64596	0	5.00000
mantego_2_plano	20000	3.17600	1.23641	63520	0	5.00000
empiezo_conversaciones	20000	3.22615	1.29815	64523	0	5.00000
poco_que_decir	20000	2.45855	1.25963	49171	0	5.00000
hablo_muchas_personas	20000	2.73665	1.40556	54733	0	5.00000
no_llamo_atencion	20000	3.38565	1.29499	67713	0	5.00000
no_soy_centro_atencion	20000	2.96860	1.36118	59372	0	5.00000
callado	20000	3.56020	1.31760	71204	0	5.00000
E stresofacil	20000	3.33285	1.35228	66657	0	5.00000
relajado	20000	3.11755	1.24202	62351	0	5.00000
preocupado	20000	3.89635	1.14113	77927	0	5.00000
no_triste	20000	2.66825	1.24558	53365	0	5.00000
me_perturban	20000	2.91055	1.27911	58211	0	5.00000
me_enojo	20000	2.93550	1.33945	58710	0	5.00000
cambio_estado_animo	20000	3.13105	1.30206	62621	0	5.00000
cambio_humor	20000	2.79195	1.35024	55839	0	5.00000
irrito_facil	20000	3.13490	1.31293	62698	0	5.00000

Por ejemplo, *alma_fiesta* correlaciona fuertemente con *hablo_muchas_personas* ($r \approx 0.65$), lo que sugiere que ambas variables miden un mismo rasgo de extroversión. En cambio, *estresofacil* presenta correlación negativa con *relajado* ($r \approx -0.52$), coherente con que representan polos opuestos de estabilidad emocional.

a.3) Determinar la estructura y si se elimina alguna de las variables se debe utilizar como suplementarias

Vamos a utilizar la **matriz de correlaciones** en lugar de la de covarianzas, ya que, entre otras cosas, las variables del ejercicio presentan varianzas distintas. De este modo se identifica de manera más fiable qué variables están asociadas entre sí y, por tanto, son adecuadas para extraer factores comunes en el análisis factorial.

Código y fórmulas empleadas para hacer el análisis discriminatorio: test de adecuación KMO, MSA, Bartlett.

Índice KMO de Kaiser-Meyer-Olkin:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2}$$

Medida de adecuación individual (MSA)

$$MSA = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2}$$

Test de esfericidad de Bartlett

$$\chi^2 = - \left[n - 1 - \frac{2p + 5}{6} \right] \ln|R|$$

```
/* =====
Test de adecuaci+n: KMO, MSA, Bartlett
===== */
proc factor data=personalidad16_sinN
  method=principal
  priors=one /* usa la matriz de correlaciones original
  rotate=none
  msa
  plots=screes;
  var _all_;
run;

/* =====
Test de esfericidad de Bartlett
===== */
proc iml;
  use corr_matrix where(TYPE="CORR");
  read all var _NUM_ into R;
  close corr_matrix;

  use corr_matrix where(TYPE="N");
  read all var _NUM_ into Nobs;
  close corr_matrix;

  n = Nobs[1,1];
  p = ncol(R);
  detR = det(R);
  ChiSq = -(n - 1 - (2*p + 5)/6)*log(detR);
  df = p*(p-1)/2;
  pval = 1 - probchi(ChiSq, df);
  print "TEST DE ESFERICIDAD DE BARTLETT" ChiSq df pval;
quit;
```

El código imprime los índices de adecuación (**KMO** y **MSA**) y el Test de Esfericidad de Bartlett.

The SAS System	
KMO	
ÍNDICE KMO GLOBAL	1
MSA	
MSA por variable	1
	1.8901395
	1.900626
	1.5470466
	1.7850446
	1.5830646
	2.1338332
	1.6440751
	3.9918322

The SAS System					
TEST DE ESFERICIDAD DE BARTLETT					
	ChiSq		df	pval	
Chi-cuadrado	351590.67	gl	1225	p-valor	0

El KMO global y el MSA por variable evalúan si se puede realizar un análisis factorial y qué variables son aceptables o no; el **scree plot** muestra graficas que te ayudan en la decisión de cuantos factores elegir para el estudio. Por último, el **test de Bartlett** actúa de forma que se evalúa una hipótesis nula: matriz de correlaciones esférica, donde no hay estructura común y no se podría hacer el análisis factorial, y una matriz no esférica con correlaciones, **tras el test se habremos si aceptar o rechazar la H0 según un p-valor (normalmente, p=0.05)**. Este ha sido nuestro caso y por lo tanto continuamos.

a.4) Depuración y separación de suplementarias

Hemos empleado un análisis exploratorio con rotación para detectar ítems débiles.

```
/* Análisis factorial con la base depurada */
proc factor data=personalidad16_sinN
  method=principal
  priors=one
  rotate=varimax
  reorder
  msa
  plots=scree;
  var _all_;
run;
```

Variables que se han detectado como débil (y su respectiva justificación):

no_triste:

no_interes_ideas_abstractas	0.02457	0.15042	-0.05880	0.10691	-0.22017
no_triste	-0.09501	-0.28100	0.02770	0.08584	0.00245

lo	no_triste	m
36	0.26518892	

hago_gente_sienta_gusto

hago_gente_sienta_gusto	-0.25768	-0.09477	0.46621	0.14061	0.10734
-------------------------	----------	----------	---------	---------	---------

hago_gente_sienta_gusto	0.41865103
-------------------------	------------

Para seleccionar las variables suplementarias, no nos hemos guiado por un 0.5 de relación, ya que hay variables con valores muy próximos que de ser eliminados podrían suponer un deterioro del proyecto. Dicho esto, solo hemos seleccionado 2 variables: no_triste y hago_gente_sienta_gusto porque no tienen ningún valor significativo.

```

/* Separar base principal (sin las suplementarias) */
data principal;
    set personalidad16_sinN;
    drop no_triste hago_gente_sienta_gusto relajado
        preocupo_por_demas no_soy_imaginativo reflexiono_cosas;
run;

/* Crear base de variables suplementarias */
data suplementarias;
    set personalidad16_sinN;
    keep no_triste hago_gente_sienta_gusto relajado
        preocupo_por_demas no_soy_imaginativo reflexiono_cosas;
run;

```

2. Análisis factorial inicial

El objetivo consiste en aplicar el método de Componentes Principales, calcular autovalores y varianza explicada, y justificar el número de factores retenidos con criterios estándar (Kaiser, proporción acumulada y scree plot).

b) Llevar a cabo análisis factorial por el método de las componentes principales, determinando y justificando la elección del número de factores elegido.”.

b.1) Aplicar el método de las componentes principales

```

/* b.1) Aplicar el método de las Componentes Principales */
proc factor data=principal
    method=principal
    priors=one
    rotate=none
    reorder
    n=5
    plots=(scree);
var _all_;
run;

```

En esta etapa no se aplica rotación, puesto que el objetivo es exclusivamente determinar k .

Matriz de partida y estandarización:

$$S = \frac{1}{n-1} (X - \bar{X})^\top (X - \bar{X}), \quad R = \frac{1}{n-1} Z^\top Z, \quad Z = \text{estandarización de } X.$$

Formulación de máxima varianza:

$$z = Xu, \quad \text{Var}(z) = \frac{1}{n-1} u^\top X^\top Xu, \quad \max_{\|u\|=1} u^\top X^\top Xu,$$

$$R = \frac{1}{n-1} X^\top X, \quad Ru = \lambda u.$$

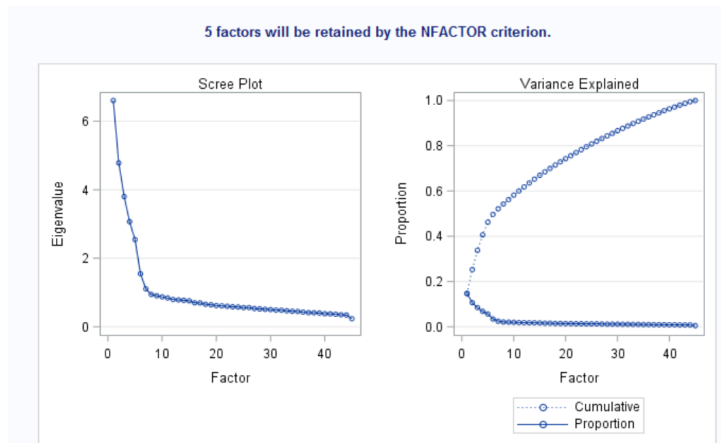
Descomposición espectral y propiedades:

$$R = P\Lambda P^\top, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p), \quad P^\top P = I, \quad \text{tr}(R) = p.$$

Varianza explicada y regla de Kaiser:

$$\text{prop}(j) = \frac{\lambda_j}{p}, \quad \text{acum}(m) = \frac{\sum_{j=1}^m \lambda_j}{p}, \quad \lambda_j > 1.$$

Rotated Factor Pattern							
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
mantego_2_plano	0.74094	0.12097	0.00424	-0.03577	0.11406	0.03389	0.12622
nohablo	0.72820	-0.02960	-0.10842	0.01194	0.07974	-0.02125	0.12967
callado	0.67637	0.16265	0.00793	-0.02291	0.10179	0.04449	0.11505
no_llamo_atencion	0.61683	0.03616	0.09357	0.06751	0.04754	0.05976	0.18341
poco_que_decir	0.59332	0.06042	-0.11167	-0.02726	0.00538	-0.15870	0.32854
no_soy_centro_atencion	-0.60865	-0.00374	-0.00888	-0.02037	0.22501	0.07113	0.13782
comodo_con_gente	-0.61306	-0.22334	0.30841	0.10886	0.10103	0.02841	0.21334
empiezo_conversaciones	-0.68286	-0.05540	0.22692	0.11094	0.16338	0.07481	0.11853
alma_fiesta	-0.69787	-0.03990	0.06804	-0.00546	0.17314	-0.02765	0.16334
hablo_muchas_personas	-0.70315	-0.07497	0.17525	0.03605	0.15612	0.02398	0.16821
cambio_humor	0.00296	0.77590	-0.04690	-0.15347	0.06213	-0.01433	0.01497
me_enojo	0.04375	0.76424	0.04698	-0.04658	-0.03534	-0.01385	0.05412
cambio_estado_animo	-0.00286	0.75916	-0.03272	-0.13663	0.08788	-0.02899	0.03642
irrito_facil	0.03588	0.75147	-0.14912	-0.00877	-0.01145	0.06098	0.08806
Estreso_facil	0.11889	0.71005	0.11237	0.01157	-0.08115	0.00845	-0.00581
preocupado	0.16352	0.61384	0.25685	0.06590	0.06298	0.03595	0.01824
siento_triste	0.25761	0.60779	0.00558	-0.17681	0.06749	0.10195	-0.07359
me_perturban	0.04588	0.59698	0.02614	-0.04628	0.01418	-0.07592	0.10732
Simpatizo_sentimientos	-0.02550	0.06738	0.79408	0.06163	0.09413	0.00882	0.01708
emociono_con_otros	-0.06602	0.11226	0.72313	0.07817	0.16170	0.00687	0.01292
corazon_blando	0.04658	0.16192	0.64865	0.02482	0.11565	-0.09534	0.09492
tomo_tiempo_demas	-0.09173	-0.00750	0.63011	0.10298	0.13215	0.05972	0.13451
interes_personas	-0.31084	-0.04263	0.59736	0.00215	0.07195	0.12982	0.02719
insulto_gente	-0.12820	0.29191	-0.40477	-0.20226	0.07092	0.20055	0.20776
no_interes_demas	0.32323	0.11962	-0.61996	-0.00536	0.14969	0.01292	0.25842
no_interesan_problemas_otros	0.15964	0.02729	-0.65355	-0.00261	0.19191	0.00520	0.30477
sigo_horario	-0.03619	0.03048	0.14280	0.65675	0.04215	0.05422	0.17795
hago_tareas_rapido	-0.06928	-0.07134	0.07106	0.64282	0.09996	-0.08403	0.11832
estoy_preparado	-0.02559	-0.08177	0.04747	0.63169	0.11198	0.15773	0.10261
gusta_orden	0.05666	0.11714	0.07497	0.60592	0.07538	0.14080	0.14501
exigente_trabajo	-0.02155	0.01373	0.10926	0.45325	0.29783	0.24123	0.11111
atento_detalle	0.05251	0.04686	0.12150	0.41867	0.36064	0.16204	0.04414
eludo_deberes	0.05934	0.24744	-0.10151	-0.50695	0.13726	0.01723	0.22279
desastre_cosas	0.06880	0.38568	0.01729	-0.58669	0.07908	0.08853	0.17309
dejo_pertenencias_porahi	-0.02546	0.09516	0.12312	-0.61197	0.11673	0.22768	0.20352
olvido_donde_dejocosas	0.01694	0.18049	0.06327	-0.65943	0.13770	0.11324	0.22771
estoy_lleno_ideas	-0.16160	0.00485	0.05872	0.01395	0.76081	0.14574	-0.14680
tengo_grandes_ideas	-0.19453	-0.04884	-0.00021	0.13204	0.68249	0.18166	-0.03299
imaginacion_vivida	-0.00120	0.13840	0.10092	-0.09270	0.66862	0.01794	-0.16603
reflesiono_cosas	0.16137	0.17931	0.25887	0.06351	0.38879	0.21260	-0.12020
vocabulario_rico	-0.03304	-0.01250	0.01289	0.04086	0.16629	0.80041	-0.13803
uso_palabras_dificiles	-0.02660	0.09361	-0.05785	-0.04138	0.16135	0.79596	-0.11327
entiendo_rapido	-0.05046	-0.11817	0.05580	0.16826	0.39907	0.46815	0.02321
no_interes_ideas_abstractas	0.01668	0.08861	-0.06188	0.09782	-0.27360	-0.08526	0.68796
dificultades_ideas_abstractas	0.04014	0.19201	0.01689	0.01374	-0.27615	-0.23848	0.61544



Elegimos 5 factores porque, aunque haya 7 que tienen un autovalor mayor que 1 (Kaiser) son los que mayor representación tienen, ya que al observar las cargas del sexto y séptimo factor representan a unas pocas variables. Añadir los factores 6–7 aporta **muy poca varianza** y tampoco mejora el ajuste (residuales/RMSR) además de complicar la estructura con cargas cruzadas. Con 5 obtenemos una solución más **sencilla e interpretable** además coherente con el llamado **Big Five** de la personalidad/emociones que son los rasgos con los que posteriormente nombraremos a los factores.

Además, en los próximos apartados se puede apreciar como la puntuación de cada factor utilizando $k=5$ es muy alta (entre 0.8 y 0.9) mientras que si usamos $k=7$, los factores 5, 6 y 7 tienen mala puntuación (menor que 0.7).

Squared Multiple Correlations of the Variables with Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
0.87485607	0.88119735	0.83362141	0.80516733	0.81656601

Correlaciones múltiples cuadradas de las variables con cada factor						
Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
0.87401947	0.88081587	0.83921423	0.80665875	0.68681449	0.68154019	0.67323599

b.2) Reestimación fijando $k = 5$ y reporte del modelo sin rotar

Volvemos a justar el análisis pero esta vez reteniendo el número de factores que hemos decidido:

```

/* b.2) Volvemos a ejecutar el análisis reteniendo el número óptimo de factores */

proc factor data=principal
    method=principal
    priors=one
    rotate=none
    reorder
    nfact=5
    msa
    plots=(scree pattern);
var _all_;
run;

```

Cargas de ACP y communalidades:

$$L = P_5 \Lambda_5^{1/2}, \quad h_i^2 = \sum_{j=1}^5 l_{ij}^2, \quad \psi_i = 1 - h_i^2.$$

Puntuaciones factoriales:

$$T = X P_5.$$

Final Communality Estimates: Total = 20.798720		
n_blando	no_interes_demas	tomo_tiempo_demas
.45891391	0.54310565	0.41697356

El porcentaje de varianza común explicada por el modelo es

$$\frac{\sum_{j=1}^5 \lambda_j}{p} = \frac{20.7987}{p},$$

Eigenvalues of the Correlation Matrix: Total = 45 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	6.60650120	1.82351186	0.1468	0.1468
2	4.78298934	0.98415218	0.1063	0.2531
3	3.79883716	0.73261756	0.0844	0.3375
4	3.06621960	0.52204661	0.0681	0.4057
5	2.54417299	0.99731173	0.0565	0.4622
6	1.54686126	0.44022636	0.0344	0.4966
7	1.10663490	0.16540198	0.0246	0.5212
8	0.94123292	0.03835024	0.0209	0.5421
9	0.90288268	0.03172929	0.0201	0.5621
10	0.87115339	0.03096328	0.0194	0.5815
11	0.84019010	0.04356035	0.0187	0.6002
12	0.79662975	0.01235801	0.0177	0.6179
13	0.78427174	0.01335069	0.0174	0.6353
14	0.77092105	0.01414120	0.0171	0.6524
15	0.75677985	0.05875171	0.0168	0.6693

Autovalores y criterios de retención					
Factor	Autovalor	Diferencia	%	% Acum.	Marcas
1	6.607	1.824	14.7%	14.7%	[Kaiser][Codo]
2	4.783	0.984	10.6%	25.3%	[Kaiser][Codo]
3	3.799	0.733	8.4%	33.8%	[Kaiser][Codo]
4	3.066	0.522	6.8%	40.6%	[Kaiser][Codo]
5	2.544	0.997	5.7%	46.2%	[Kaiser][Codo]
6	1.547	0.440	3.4%	49.7%	[Kaiser]
7	1.107	0.165	2.5%	52.1%	[Kaiser]
8	0.941	0.038	2.1%	54.2%	
9	0.903	0.032	2.0%	56.2%	
10	0.871	0.031	1.9%	58.1%	
11	0.840	0.044	1.9%	60.0%	
12	0.797	0.012	1.8%	61.8%	
13	0.784	0.013	1.7%	63.5%	
14	0.771	0.014	1.7%	65.2%	
15	0.757	0.059	1.7%	66.9%	

3. Rotación e interpretación de factores

c) *Rotación de la solución y determinación de las variables más relacionadas con cada factor*

Aplicamos una **rotación ortogonal VARIMAX** a la solución con $k = 5$ factores con el fin de hacer más simple la estructura de cargas. Tras la rotación, para cada factor, se identifican las **cargas** y en base a esas agrupaciones hemos propuesto nombres para cada factor.

Rotación ortogonal:

$$L^* = LT, \quad T^T T = I, \quad h_i^2 = \sum_{j=1}^k (l_{ij}^*)^2 = \sum_{j=1}^k l_{ij}^2, \quad \Phi = I_k.$$

Criterio VARIMAX (forma clásica, maximizando el contraste de las cargas al cuadrado):

$$\max_T \sum_{j=1}^k \left[\frac{1}{p} \sum_{i=1}^p (l_{ij}^*)^4 - \left(\frac{1}{p} \sum_{i=1}^p (l_{ij}^*)^2 \right)^2 \right].$$

(Con rotación **oblicua**, por comparación: $L^* = LT$, pero $\Phi = T^T T \neq I$ y la **matriz de estructura** sería $S = L^* \Phi$.)

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix					
	1	2	3	4	5
1	-0.68864	0.50455	-0.35664	-0.32197	-0.20086
2	0.21654	0.74005	0.46207	-0.08354	0.43005
3	-0.58135	-0.05190	0.24332	0.72896	0.26221
4	0.01184	-0.12845	-0.56842	-0.10003	0.80638
5	0.37521	0.42256	-0.52630	0.58989	-0.23601

(La tabla rotada con los factores subrayados está en el [apartado b.1](#))

Tal y como he mencionado anteriormente, la estructura obtenida se alinea con el **modelo Big Five** de personalidad el que se distinguen 5 grandes rasgos de la personalidad: **extraversión, amabilidad, responsabilidad (conciencia), estabilidad emocional y apertura a la experiencia.**

Por lo tanto, nuestros factores quedan:

FACTOR 1: **Extraversión**

FACTOR 2: **Estabilidad emocional**

FACTOR 3: **Amabilidad**

FACTOR 4: **Responsabilidad**

FACTOR 5: **Intelecto**

4. Creación de nuevos conjuntos de datos

d) Puntuaciones factoriales (proyección de las observaciones)

Se obtienen las **puntuaciones factoriales** de cada observación sobre los 5 factores rotados (VARIMAX). Estas puntuaciones son combinaciones lineales de las variables ya estandarizadas y permiten **posicionar** cada una de las observaciones.

Esta tabla indica **qué tan bien puede predecirse o definirse** cada factor a partir del conjunto de datos, los valores altos implican factores bien determinados (entre 0.8-0.9).

Squared Multiple Correlations of the Variables with Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
0.87485607	0.88119735	0.83362141	0.80516733	0.81656601

La matriz de **pesos** para calcular las **puntuaciones factoriales por regresión**. Cada celda es el peso de una variable (estandarizada) en un factor. Para un sujeto i y el factor k :

$$\text{Score}_{ik} = \sum_j w_{jk} z_{ij}$$

	Factor1	Factor2	Factor3	Factor4	Factor5
hablo_muchas_personas	0.15984	0.03256	-0.00463	0.01726	-0.00078
alma_fiesta	0.14347	0.03400	-0.03033	0.00681	-0.00251
empiezo_conversaciones	0.14598	0.03599	0.01059	0.03808	0.01494
comodo_con_gente	0.12125	0.00009	0.03920	0.03288	-0.01495
no_soy_centro_atencion	0.11186	0.03148	-0.04113	0.00367	0.02984
poco_que_decir	-0.00919	0.01581	-0.00342	0.02527	-0.03245
no_llamo_atencion	-0.10784	-0.00027	0.03719	0.03015	0.01649
callado	-0.11982	0.01476	0.02683	0.01167	0.03770
nohablo	-0.15133	-0.02212	0.00518	0.02153	0.03069
mantengo_2_plano	-0.16239	0.00589	0.03478	0.00916	0.04878
cambio_humor	0.04362	0.20849	-0.02962	-0.00840	-0.00691
cambio_estado_animo	0.03785	0.17919	-0.01880	0.00348	-0.00378
me_enojo	0.01809	0.15897	0.01126	0.02643	-0.03468
irrito_facil	0.03159	0.16865	-0.06274	0.04924	-0.01082
Estreso_facil	-0.00734	0.13056	0.03247	0.02999	-0.03288
preocupado	-0.02660	0.11059	0.06161	0.04033	0.00787
siento_triste	-0.03882	0.07649	0.01472	-0.03110	0.03642
me_perturban	0.00915	0.08690	0.00460	0.01296	-0.02290
Simpatizo_sentimientos	-0.04431	0.00876	0.25826	-0.00572	-0.00189
emociono_con_otros	-0.02136	0.02058	0.19193	0.00570	0.01593
corazon_blando	-0.02668	0.02612	0.13733	0.00244	-0.01461
tomo_tiempo_demas	-0.01007	0.00822	0.11547	0.01880	0.00979
interes_personas	0.02792	0.00192	0.11897	-0.01807	0.01748
insulto_gente	0.04274	0.05297	-0.08881	-0.01602	0.02773
no_interes_demas	-0.02747	0.04682	-0.18594	0.05531	0.03922
no_interesan_problemas_otros	0.00777	0.03363	-0.16775	0.04648	0.03458
sigio_horario	0.00263	0.03982	0.00123	0.18348	-0.01498
hago_tareas_rapido	0.00687	0.01471	-0.00846	0.16427	-0.01958
estoy_preparado	-0.00691	0.01323	-0.01935	0.15897	0.02319
gusta_orden	-0.00975	0.04289	-0.00814	0.15133	0.01197
exigente_trabajo	-0.00437	0.02153	-0.00425	0.10775	0.06241
atento_detalle	-0.01567	0.01966	0.00060	0.09440	0.06862
eludo_deberes	0.00629	0.03358	-0.01518	-0.09078	0.01881
dejo_pertenencias_porahi	0.00789	0.01007	0.02705	-0.13848	0.05093
desastre_cosas	0.00596	0.06272	0.01117	-0.14711	0.02823
olvido_donde_dejocosas	0.00616	0.02436	0.02557	-0.16952	0.04405
estoy_lleno_ideas	0.01282	0.00132	-0.01179	-0.00680	0.22405
vocabulario_rico	-0.01068	-0.01211	-0.02971	-0.00428	0.18773
uso_palabras_dificiles	-0.00363	0.00439	-0.04638	-0.01986	0.17990
tengo_grandes_ideas	0.02600	0.00559	-0.03809	0.03911	0.16448
entiendo_rapido	-0.00668	-0.01245	-0.01625	0.03747	0.12792
imaginacion_vivida	-0.01138	0.00966	0.01456	-0.02530	0.11448
reflesiono_cosas	-0.03658	0.01574	0.04238	0.00864	0.08843
no_interes_ideas_abstractas	0.02640	0.05876	-0.02981	0.06918	-0.10534
dificultades_ideas_abstractas	0.02406	0.07465	-0.00281	0.05390	-0.14047

Los signos indican si la variable suma o resta al factor (son correlaciones variables–factor).

e) Proyección de las variables sobre los ejes factoriales (coordenadas factoriales)

Hemos obtenido las **coordenadas factoriales de las variables** (cargas rotadas sobre los factores) para representar la estructura en los **planos factoriales**. Lo interpretamos como la **correlación** entre la variable i y el factor j .

Coordenadas factoriales (variables):

$$\text{Coord}(x_i \text{ en } F_j) = l_{ij}^*.$$

Comunalidad (calidad de representación de x_i):

$$h_i^2 = \sum_{j=1}^k (l_{ij}^*)^2.$$

Rotated Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
hablo_muchas_personas	0.69033	-0.01777	0.15045	0.07521	0.09143
alma_fiesta	0.66992	0.00990	0.05816	0.03190	0.07237
empiezo_conversaciones	0.66660	-0.01459	0.20084	0.13811	0.13799
comodo_con_gente	0.60968	-0.14651	0.26635	0.15262	0.04999
no_soy_centro_atencion	0.57048	0.03858	-0.01034	0.01532	0.16509
poco_que_decir	-0.53309	0.13935	-0.12597	0.02800	-0.16028
no_llamo_atencion	-0.54576	0.09164	0.05412	0.10062	0.02607
callado	-0.61953	0.19409	-0.01041	0.00812	0.06776
nohablo	-0.67457	0.01915	-0.11582	0.04259	0.01221
mantego_2_plano	-0.68858	0.16258	-0.01288	0.00097	0.07147
cambio_humor	-0.02024	0.74649	-0.03543	-0.15020	0.02410
cambio_estado_animo	-0.01287	0.73267	-0.02200	-0.12788	0.02959
me_enojo	-0.06562	0.70375	0.05221	-0.05010	-0.04923
irrito_facil	-0.05979	0.69886	-0.13655	-0.01159	-0.00289
Estreso_facil	-0.13888	0.62740	0.11391	-0.00833	-0.05154
preocupado	-0.16607	0.55594	0.23987	0.06093	0.06298

Final Communality Estimates: Total = 17.838568

5. Representaciones gráficas

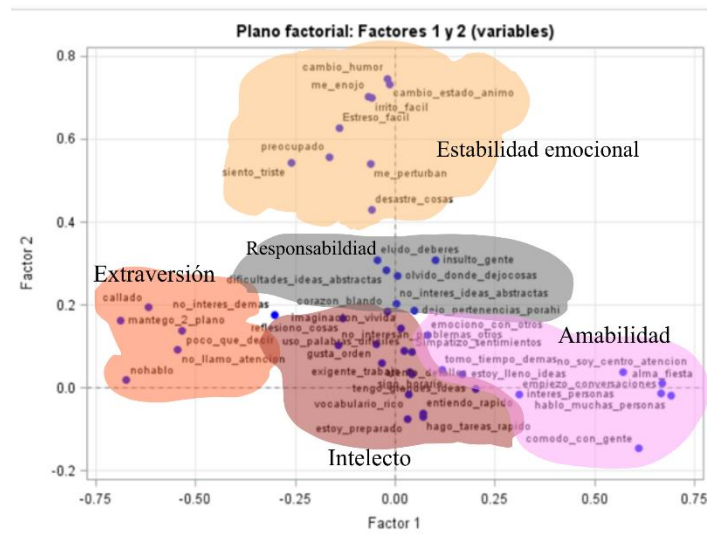
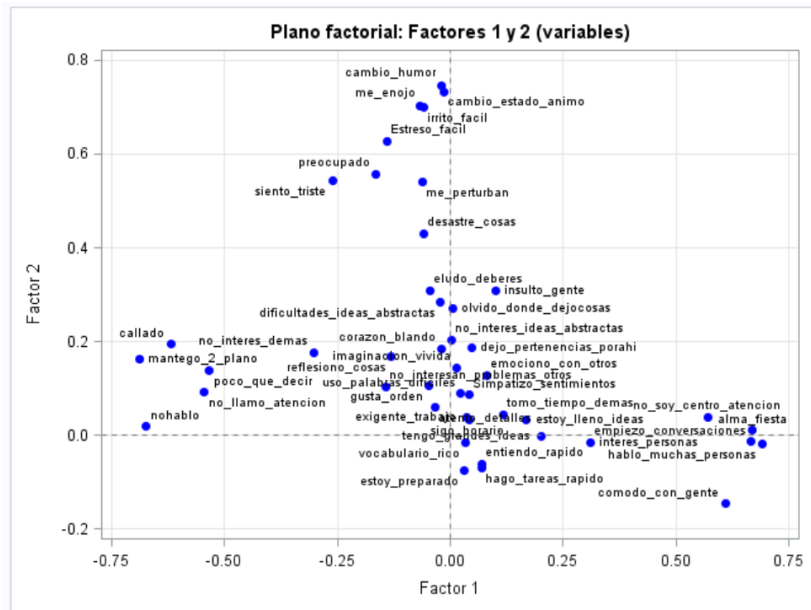
f) Representación gráfica

Visualización de la estructura factorial retenida con cinco factores.

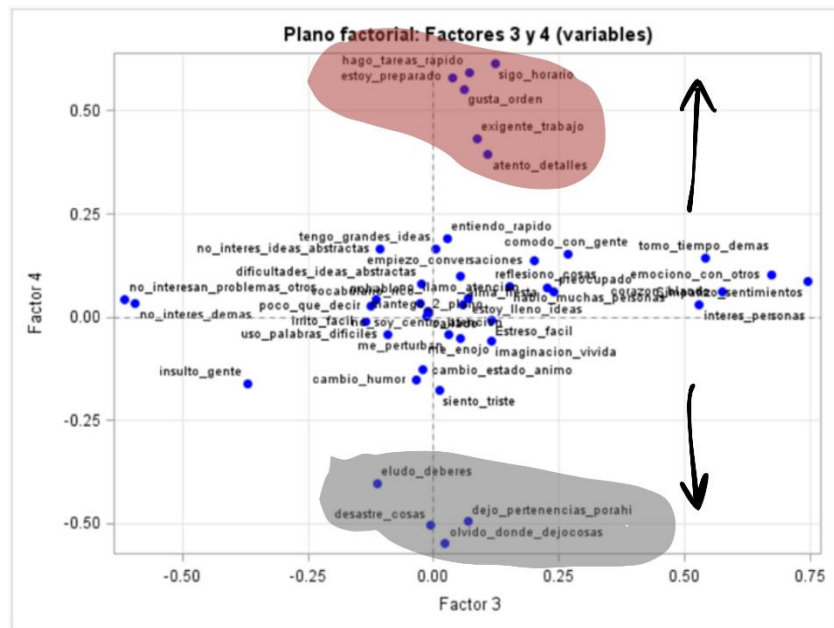
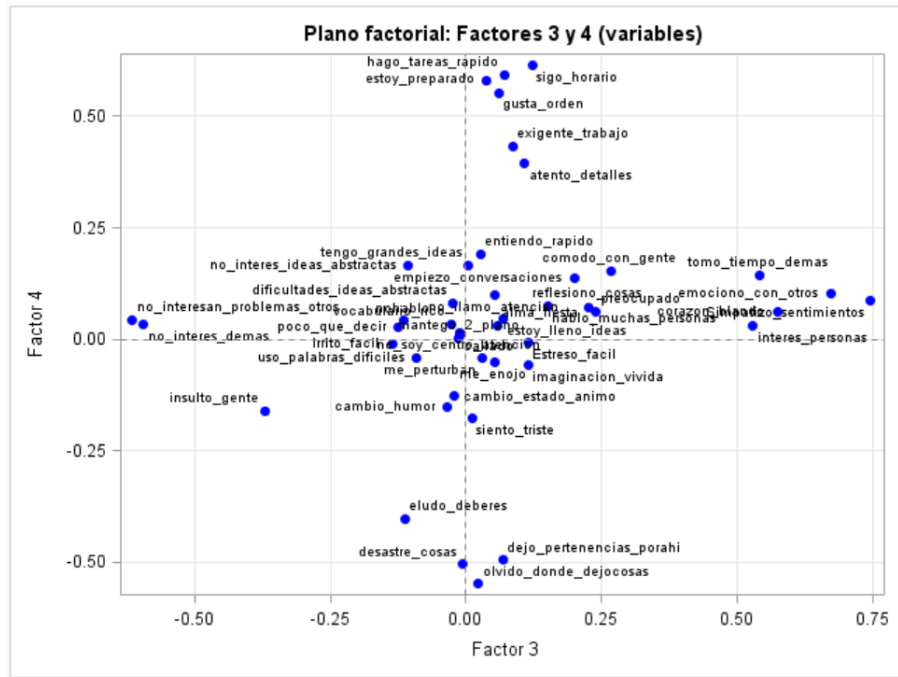
Círculo unidad:

Las variables que están cercanas al círculo unidad están bien representadas en ese plano, tienen una alta proyección, mientras que las cercanas al origen aportan poca información (en ese plano). Como se puede apreciar en los gráficos, los agrupamientos de variables sugieren dimensiones comunes mientras que las opuestas sugieren relaciones negativas.

f.1) Factores 1 y 2

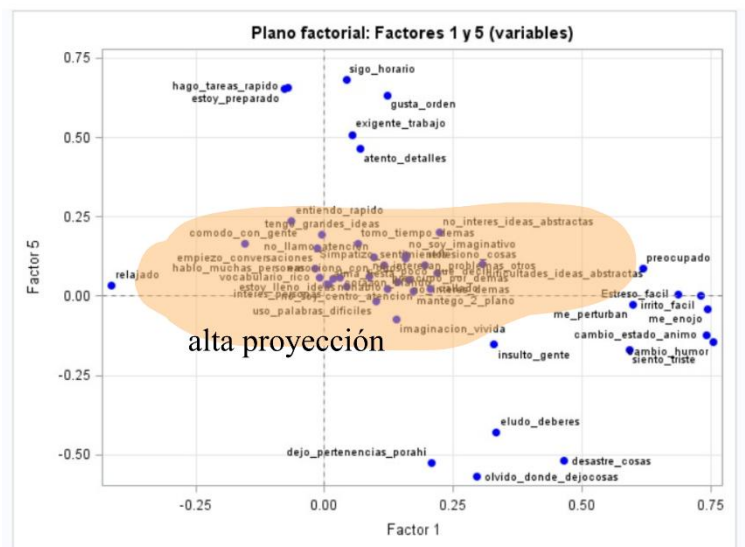
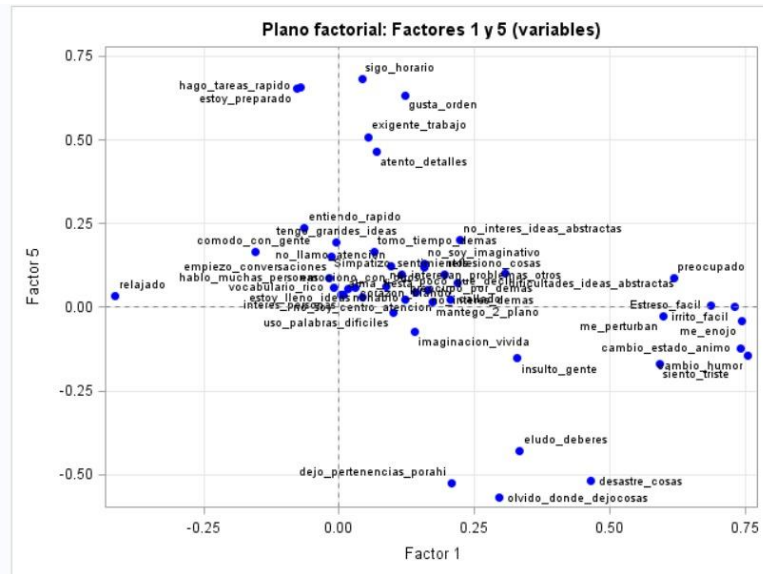


f.2) Factores 3 y 4



(se aprecian las direcciones opuestas)

f.3) Factores 1 y 5



6. Bondad del ajuste

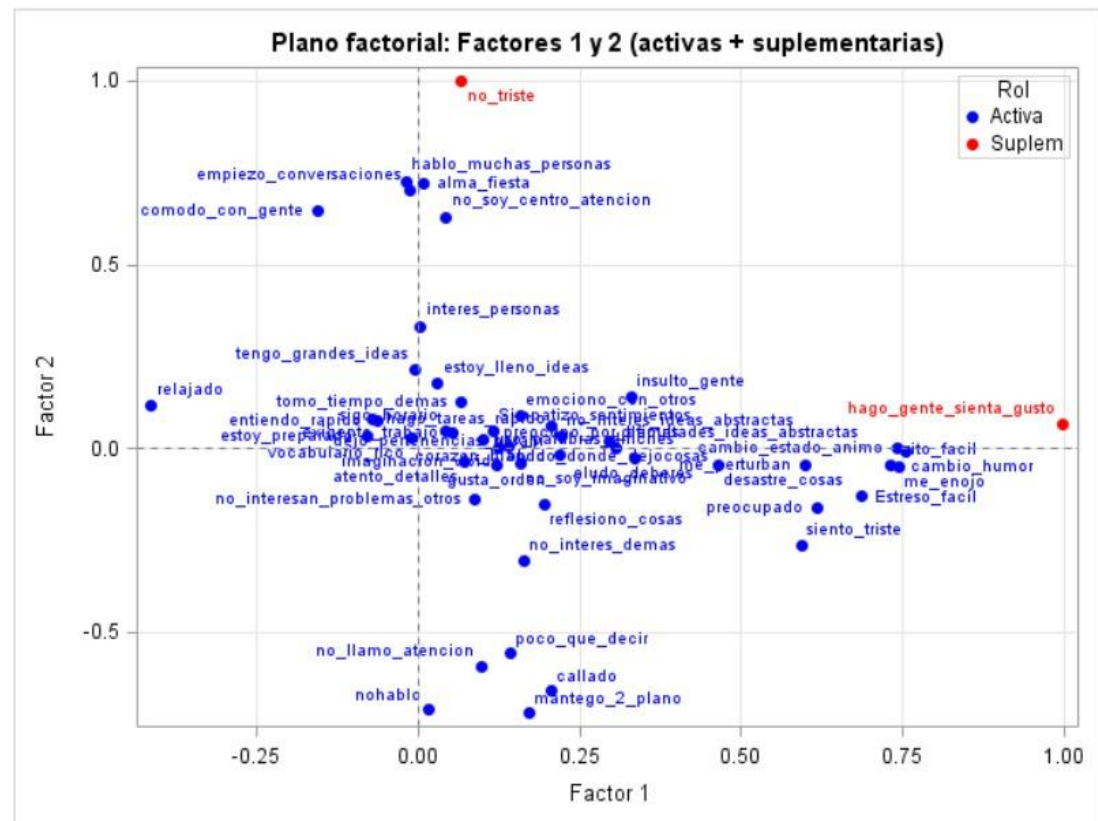
g) Matriz de correlaciones residuales

Residual Correlations With Uniqueness on the Diagonal							
Simpatizo_sentimientos	no_interesan_problemas_otros	corazon_blando	no_interes_demas	tomo_tiempo_demas	emociono_con_otros	estoy_preparado	
0.02161	0.03531	0.03093	0.04136	0.01411	0.01919	-0.00948	
0.00832	0.02015	0.00931	-0.00277	0.02596	0.01966	0.00047	
-0.00114	0.07166	0.00015	0.05515	0.01965	-0.03142	0.00032	
0.00027	0.03201	-0.01410	-0.00415	0.01344	-0.01145	0.00328	
-0.00478	0.03969	-0.00645	0.06428	0.00501	-0.01923	-0.02460	
0.02204	0.05188	0.03881	0.03481	0.05167	0.03486	0.00158	
0.00286	0.05999	0.00057	0.04260	0.02130	-0.00343	-0.01572	
-0.01650	0.04823	-0.03406	0.02107	0.04561	-0.02949	-0.02831	
0.02633	0.02514	0.04216	0.03153	0.01845	0.05415	-0.01134	
0.00068	0.01719	-0.00948	-0.00102	0.02144	-0.00677	0.00103	
-0.03314	-0.02495	-0.05249	-0.03453	-0.05778	-0.05547	0.01269	
-0.04813	0.00072	-0.05619	-0.00832	-0.05364	-0.07105	-0.01321	
-0.00817	-0.01219	-0.00677	-0.00870	-0.01941	-0.00908	-0.01463	
-0.01802	-0.03193	-0.00250	-0.03673	-0.03596	-0.01493	0.00354	
-0.00910	-0.04464	-0.02780	-0.03716	-0.00669	-0.00669	0.00600	

Root Mean Square Off-Diagonal Residuals: Overall = 0.04290449			
corazon_blando	no_interes_demas	tomo_tiempo_demas	emociono_con_otros

La matriz de correlaciones residuales sin contar la diagonal tiene una media de 0.04290449 lo cual es muy bajo y denota un ajuste excelente.

h) Proyectar las variables suplementarias generadas en el apartado a sobre los ejes factoriales. Representarlas junto al resto de variables en los planos generados en el apartado f.



Tal y como era de esperar, las variables suplementarias están extremadamente lejos del origen.