

Unidad 6: Construcción de Indicadores Complejos

La Unión Europea está constituida como un organismo supranacional debido a que tiene potestad para decidir en materia pública sobre la totalidad de países que la conforman. Las decisiones son tomadas de manera conjunta a través del Consejo Europeo y el Parlamento Europeo, que son órganos donde están representados los países miembros. De manera paralela, los miembros de la UE manejan cifras que les permiten adoptar disposiciones individuales haciendo uso pleno de su soberanía.

La siguiente base de datos presenta la recolección de algunos índices macroeconómicos de los países miembros de la Unión Europea utilizando como fuente la base de datos del Fondo Monetario Internacional. Esta información es relevante para los organismos encargados de establecer políticas públicas en materia económica a nivel interestatal.

Construcción de un Indicador y un Índice

En la estadística los *indicadores e índices* nos permiten crear una medición certera sobre conceptos cuyo grado de complejidad requiere un esfuerzo adicional para su sistematización. El primero es la versión cuantificada de un concepto; en otras palabras, los términos de medición de una variable compleja. Por otro lado, el índice equivale al conjunto de indicadores que nos permitirán cuantificar un concepto. Este concepto puede ser de tal complejidad que no siempre se va a poder medir con una sola variable. Esta variable requiere de un conjunto de indicadores para generar una medición acertada puesto que ninguna variable por sí sola es capaz satisfacer sus distintas dimensiones conceptuales. Para crear *indicadores e índices* existen diversos métodos como el de máxima verosimilitud o el de análisis alfa. En esta oportunidad nos concentraremos en el *Método de Componentes Principales* y nos enfocaremos en la técnica del *Índice Aritmético* y del *Análisis Factorial*.

Podrían resumir mi clase:

http://prezi.com/4ae3t2vnbf7i/?utm_campaign=share&utm_medium=copy&rc=ex0share

Análisis Factorial

Para esto por favor copien, peguen y editen lo que está aquí:

<https://www.dropbox.com/s/gug3zu9a74j3ko7/QU%C3%89%20ES%20EL%20AN%C3%81LISIS%20FACTORIAL.docx?dl=0>

El análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Esos grupos homogéneos se forman con las variables que se correlacionan mucho entre sí y que los grupos sean independientes de otros.

Cuando recogemos un gran número de variables de forma simultánea, podemos estar interesados en averiguar si las preguntas del cuestionario se agrupan de alguna forma característica. Aplicando un análisis factorial a las respuestas de los individuos podemos encontrar grupos de variables con significado común y conseguir de esta manera reducir el número de dimensiones necesarias para explicar las respuestas de los sujetos.

En el análisis factorial todas las variables del análisis cumplen el mismo papel: todas ellas son independientes en el sentido de que no existe a priori una dependencia conceptual de unas variables sobre otras. Vale recordar por tanto que con el análisis factorial no estamos planteando un modelo explicativo o causal entre variables independientes y dependientes.

En este caso hipotético, este análisis nos permitiría formar ya un primer par de indicadores bajo los nombres de *gastos de élites conservadores* y *gastos de élites progresistas* en nuestro camino a construir el indicador sobre la influencia de las élites. Ahora veamos cómo funciona el Análisis Factorial en el software con la base de datos UEunidad2.

Esta base de datos presenta como casos a los países de la Unión Europea y algunos de sus índices financieros internacionales como *the Government Bonds* y *the Value of Exports and Imports* como variables.

Recordemos que para abrir la base de datos de indicadores financieros de la Unión Europea que está guardado en SPSS utilizamos el el paquete *foreign*.

```
library(foreign)
```

```
data<-read.spss("UE.sav",use.value.labels=TRUE, max.value.labels=Inf, to.data.frame=TRUE)  
#importamos la base de datos en SPSS
```

Para trabajar el Análisis Factorial las variables tienen que ser escalares y deben de encontrarse, de preferencia, en una misma escala. es que las variables a agrupar deben ser escalares y deben de encontrarse tipificadas .. Veamos un par de ejemplos sobre estandarización.

```
ZCPIBASED<- scale(CPIBASED)#convertir la variable a la tipificación Z
```

```
summary(ZCPIBASED)#describir la nueva variable tipificada
```

V1

Min. :-1.5311

1st Qu.: -0.5973

Median :-0.2908

Mean : 0.0000

3rd Qu.: 0.2100

Max. : 3.1197

NA's :9

```
ZGovernmentbonds<- scale(Governmentbonds)#convertir la variable a la tipificación Z
```

summary(ZGovernmentbonds)

V1

Min. :-1.0066

1st Qu.: -0.8078

Median :-0.2021

Mean : 0.0000

3rd Qu.: 0.2943

Max. : 2.5030

NA's :8

str(data)# nos muestra las variables tipificadas de la base de datos.

```
$ ZCPIBASED      : num -0.0821 1.1711 -1.5311 NA -0.3333 ...
$ ZTreasuryBills : num NA -0.556 NA NA NA ...
$ ZGovernmentbonds : num -0.67 0.258 NA NA -0.823 ...
$ ZTotalReserves : num -0.21 -0.208 -0.209 -0.213 -0.21 ...
$ ZTotalReservesMinusGold : num -0.671 -0.523 -0.55 -1.047 -0.78 ...
$ ZGold          : num -0.208 -0.465 NA -0.507 -0.455 ...
$ ZReservePositionintheFund : num -0.271 -0.688 -0.706 -0.706 -0.432 ...
$ ZSDRHoldings   : num -0.25 -0.567 -0.657 -0.728 -0.418 ...
$ ZForeignExchange : num -0.675 -0.393 -0.402 -0.95 -0.745 ...
$ ZConsumerPrices : num -0.00713 0.92438 -1.00119 1.66528 0.13949 ...
$ ZIndustrialProduction : num 0.49 0.429 -0.952 3.197 -0.381 ...
$ ZGoodsValueofexports : num -0.222 -0.655 -0.701 -0.69 -0.499 ...
$ ZGoodsValueofImports : num -0.208 -0.699 -0.731 -0.747 -0.529 ...
$ ZGDPDeflator    : num -0.0923 1.9759 -0.1848 0.9583 0.6848 ...
- attr(*, "variable.labels")= Named chr "PaÃ-s " "CPI BASED" "Treasury Bills" "Government bonds" ...
..- attr(*, "names")= chr "PaÃ-s" "CPIBASED" "TreasuryBills" "Governmentbonds" ...
- attr(*, "codepage")= int 65001
```

Es necesario crear un subconjunto de variables que utilizaremos en el Análisis Factorial.

```
data2<-subset(data,select=c(16,19,23,27,28), na.value=NULL) #seleccionar las variables a trabajar en el análisis factorial
```

```
names(data2)#verificar la selección
```

```
[1] "ZCPIBASED"      "ZTotalReserves"  "ZSDRHoldings"
```

```
[4] "ZGoodsValueofexports" "ZGoodsValueofImports"
```

Para poder ejecutar el análisis factorial, las variables deben estar posicionadas en una matriz; por eso, es necesario hacer esta conversión y verificarla con el comando *is.matrix*.

```
is.matrix(data2)#analizar si el grupo filtrado es una matriz
```

```
[1] FALSE
```

```
data3<-as.matrix(data2)#convertir la subdata en una matriz
```

```
is.matrix(data3)#verificar si ya es una matriz
```

```
[1] TRUE
```

Luego de esto, es necesario probar la fiabilidad de nuestras variables. Este concepto se mide con el Alfa de CronBach y lo que esperamos es un resultado entre el 0 y el 1 mayor a 0.7 o mayor a 0.5. Para esto, instalamos el paquete *cocron* y utilizamos el comando *cronbach.alpha* con la matriz que creamos en el paso anterior. Al analizar la fiabilidad no estamos generando aún ninguna agrupación entre las variables, además de la matriz.

```
install.packages("cocron")#instalar el paquete cocron para pedir el alfa de CronBach
```

```
library(cocron)#ejecutar el paquete cocron
```

```
cronbach.alpha(data3, standardized = FALSE)# pedir el alfa de CronBach para data3
```

```
alpha
```

```
0.6531036
```

El *alfa de cronbach* para “data3” confirma el cumplimiento de las características que necesitamos (mayor a 0.5). Por lo tanto, podemos asegurar que las variables que hemos

seleccionado son confiables. Luego de completar esto y el requisito de la estandarización, el análisis se dividirá en dos grandes ejes: el Análisis Exploratorio y el Análisis Confirmatorio. El primero se realiza con el paquete *psych* y los comandos *KMO*, *pairs.panels*, *vss* y *efacp*, *print* y *fa.diagram*.

```
library(psych)# ejecutar el paquete psych
```

Luego de la prueba de fiabilidad, es necesario establecer si la matriz tiene validez. Para esto se utiliza la prueba de KMO, donde se espera la misma respuesta que con la fiabilidad: en una escala de 0 a 1, el resultado debe ser mayor a 0.5 para aceptarla como válida e indica una mayor validez si el resultado es mayor a 0.7.

```
KMO(data3)#verificar KMO de la subdata convertida en matriz, data3
```

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = data3)

Overall MSA = 0.54

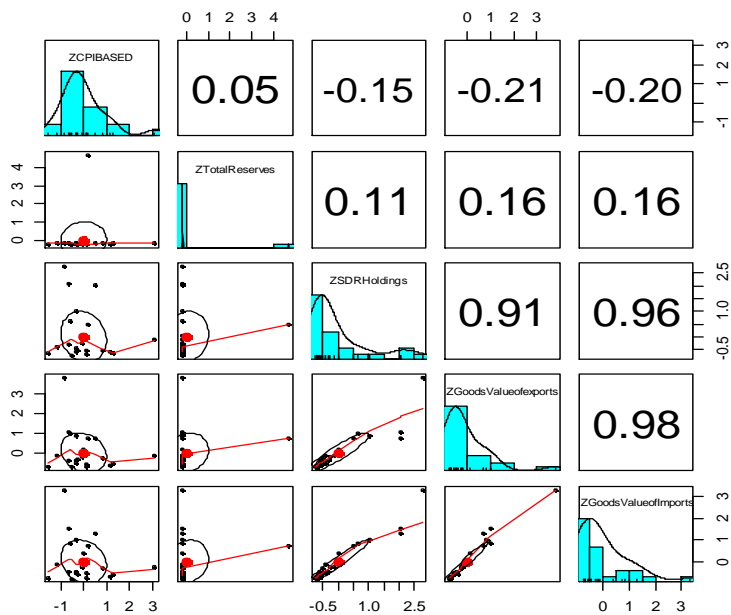
MSA for each item =

ZCPIBASED	ZTotalReserves	ZSDRHoldings
0.65	0.30	0.56

ZGoodsValueofexports	ZGoodsValueofImports
0.56	0.52

El KMO nos indica que la matriz cumple el requisito de validez. Además, vemos qué tanto se extrae el factor de cada variable. Siguiendo la misma lógica, en una escala del 0 al 1, la cercanía al 0 indica que no explica nada, mientras que la cercanía al 1 indica una explicación total. El paso siguiente es la conformación de los factores.

```
pairs.panels(data3)
```



```
data3<-na.omit(data3)
```

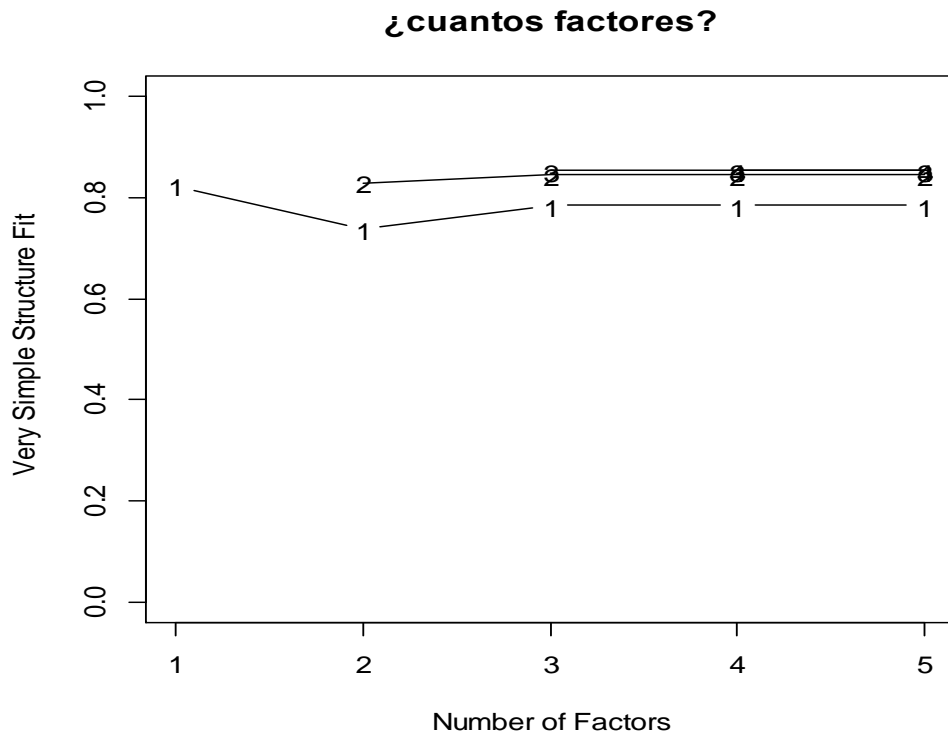
```
fit <- princomp(data3, cor=TRUE)#guardar los componentes en un nuevo objeto
```

```
summary(fit)#descripción de los componentes
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.72	1.02	0.94	0.31	0.07
Proportion of Variance	0.59	0.20	0.17	0.01	0.00
Cumulative Proportion	0.59	0.80	0.97	0.99	1.00

El cuadro de la Varianza Total Explicada nos indica que el segundo componente acumula alrededor del 80% de los datos.



A grandes rasgos, el cuadro titulado “cuántos factores” nos indica que debemos formar es 2. Este a su vez, concentra el 84% de los valores. Para determinar qué variables componen esos factores, sigamos con el análisis.

```
library(GPArotation)#ejecutar el paquete GPArotation
```

```
efacp<- principal(data2, nfactors=2, scores=TRUE, rotate="varimax")#análisis exploratorio
```

```
summary(efacp)#descripción del nuevo elemento
```

Factor analysis with Call: principal(r = data3, nfactors = 2, rotate = "varimax", scores = TRUE)

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the model is 1 and the objective function was 1.49

The number of observations was 30 with Chi Square = 37.62 with prob < 8.6e-10

```
print(efacp$loadings,cutoff=.33, sort=TRUE)#agrupación de las variables
```

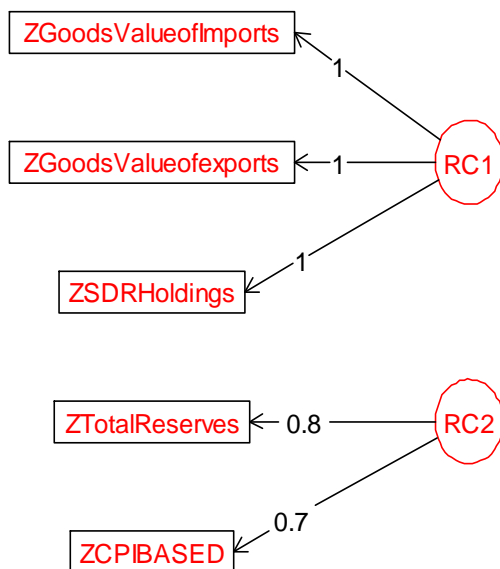
Loadings:

	RC1	RC2
ZSDRHoldings	0.962	
ZGoodsValueofexports	0.976	
ZGoodsValueofImports	0.994	
ZCPIBASED		0.684
ZTotalReserves		0.761

RC1 RC2
SS loadings 2.979 1.048
Proportion Var 0.596 0.210
Cumulative Var 0.596 0.805

`fa.diagram(efacp, col="red", main="Proyecto 2.0")` #grafico de los componentes agrupados

Proyecto 2.0



El análisis exploratorio nos indica primero que el primer componente se compone de las variables *ZGoodsValueofImports*, *ZGoodsofexports* y *ZSDRHoldings*, de estas es extraído el

0.96%, 0.97% y 0.99%, respectivamente. El segundo componente lo componen las variables *ZTotalReserves* y *ZCPIBASED* de donde es extraído el 0.68% y el 0.76% de la segunda. Asimismo, el último gráfico nos confirma estos componentes.

De esta manera, el análisis factorial nos confirma que se pueden armar dos índices que agrupen estas 5 variables con criterios aceptables de confiabilidad y validez. Esta misma estructura se puede ejecutar con otras variables y con otras bases de datos.

Ahora es tu turno...

Ejercicios

- Para nuestro trabajo de investigación hemos decidido estudiar el tema de políticas financieras en micro y pequeña empresa. Para esto, usted considera determinante reducir la cantidad de variables debido a que algunas de ellas responden a una lógica muy similar. Este es el caso de las variables relacionadas a la duración del servicio en las capacitaciones hechas por el MEF y las variables relacionadas a los ámbitos de capacitación. Para esto, utilice las variables *P4_4_5_1*, *P4_4_5_2*, *P4_4_5_3*, *P8_1_1_8*, *P8_1_1_9*, *P8_1_1_10* y *P8_1_1_11* de la data guardada con el nombre “Ejercicios U6” del INEI sobre retroalimentación de la gestión de recursos humanos. Luego, interprete los resultados