

ANALISIS FACTORIAL

1.- Lectura de la matriz de datos

```
x<-as.data.frame(state.x77)
x
```

##	Population	Income	Illiteracy	Life Exp	Murder	HS	Grad	Frost
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20	
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	
## Arkansas	2110	3378	1.9	70.66	10.1	39.9	65	
## California	21198	5114	1.1	71.71	10.3	62.6	20	
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139	
## Delaware	579	4809	0.9	70.06	6.2	54.6	103	
## Florida	8277	4815	1.3	70.66	10.7	52.6	11	
## Georgia	4931	4091	2.0	68.54	13.9	40.6	60	
## Hawaii	868	4963	1.9	73.60	6.2	61.9	0	
## Idaho	813	4119	0.6	71.87	5.3	59.5	126	
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127	
## Indiana	5313	4458	0.7	70.88	7.1	52.9	122	
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140	
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114	
## Kentucky	3387	3712	1.6	70.10	10.6	38.5	95	
## Louisiana	3806	3545	2.8	68.76	13.2	42.2	12	
## Maine	1058	3694	0.7	70.39	2.7	54.7	161	
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101	
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103	
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125	
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	
## Mississippi	2341	3098	2.4	68.09	12.5	41.0	50	
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108	
## Montana	746	4347	0.6	70.56	5.0	59.2	155	
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139	
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	
## New Hampshire	812	4281	0.7	71.23	3.3	57.6	174	
## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115	
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120	
## New York	18076	4903	1.4	70.55	10.9	52.7	82	
## North Carolina	5441	3875	1.8	69.21	11.1	38.5	80	
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186	
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124	
## Oklahoma	2715	3983	1.1	71.42	6.4	51.6	82	
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44	
## Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126	
## Rhode Island	931	4558	1.3	71.90	2.4	46.4	127	
## South Carolina	2816	3635	2.3	67.96	11.6	37.8	65	
## South Dakota	681	4167	0.5	72.08	1.7	53.3	172	
## Tennessee	4173	3821	1.7	70.11	11.0	41.8	70	
## Texas	12237	4188	2.2	70.90	12.2	47.4	35	
## Utah	1203	4022	0.6	72.90	4.5	67.3	137	
## Vermont	472	3907	0.6	71.64	5.5	57.1	168	
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85	
## Washington	3559	4864	0.6	71.72	4.3	63.5	32	
## West Virginia	1799	3617	1.4	69.48	6.7	41.6	100	
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	
##	Area							

## Alabama	50708
## Alaska	566432
## Arizona	113417
## Arkansas	51945
## California	156361
## Colorado	103766
## Connecticut	4862
## Delaware	1982
## Florida	54090
## Georgia	58073
## Hawaii	6425
## Idaho	82677
## Illinois	55748
## Indiana	36097
## Iowa	55941
## Kansas	81787
## Kentucky	39650
## Louisiana	44930
## Maine	30920
## Maryland	9891
## Massachusetts	7826
## Michigan	56817
## Minnesota	79289
## Mississippi	47296
## Missouri	68995
## Montana	145587
## Nebraska	76483
## Nevada	109889
## New Hampshire	9027
## New Jersey	7521
## New Mexico	121412
## New York	47831
## North Carolina	48798
## North Dakota	69273
## Ohio	40975
## Oklahoma	68782
## Oregon	96184
## Pennsylvania	44966
## Rhode Island	1049
## South Carolina	30225
## South Dakota	75955
## Tennessee	41328
## Texas	262134
## Utah	82096
## Vermont	9267
## Virginia	39780
## Washington	66570
## West Virginia	24070
## Wisconsin	54464
## Wyoming	97203

2.- Quitar los espacios de los nombres

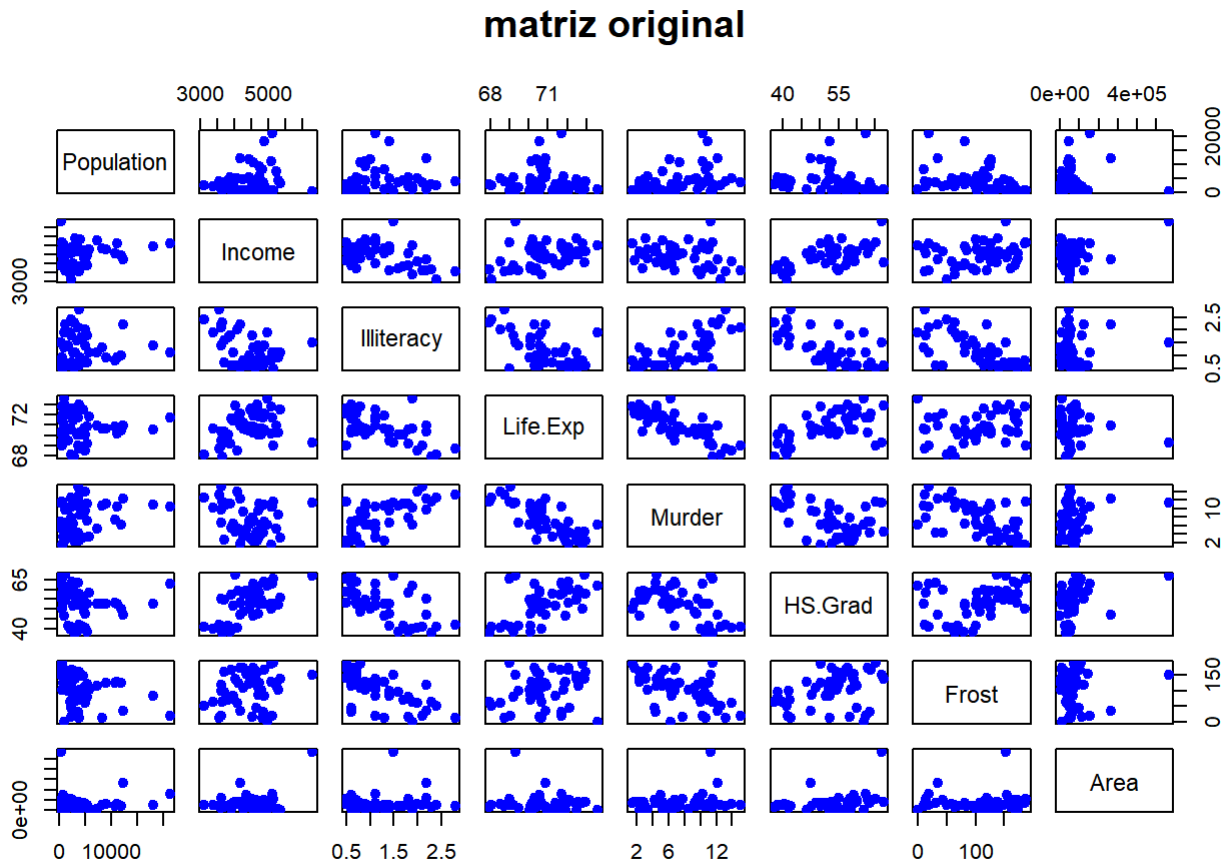
```
colnames(x)[4]="Life.Exp"
colnames(x)[6]="HS.Grad"
```

3.- Separa n (estados) y p (variables)

```
n<-dim(x)[1]
p<-dim(x)[2]
```

4.- Generacion de un scatter plot para la visualización de variables originales

```
pairs(x, col="blue", pch=19, main="matriz original")
```



Transformación de alguna variables

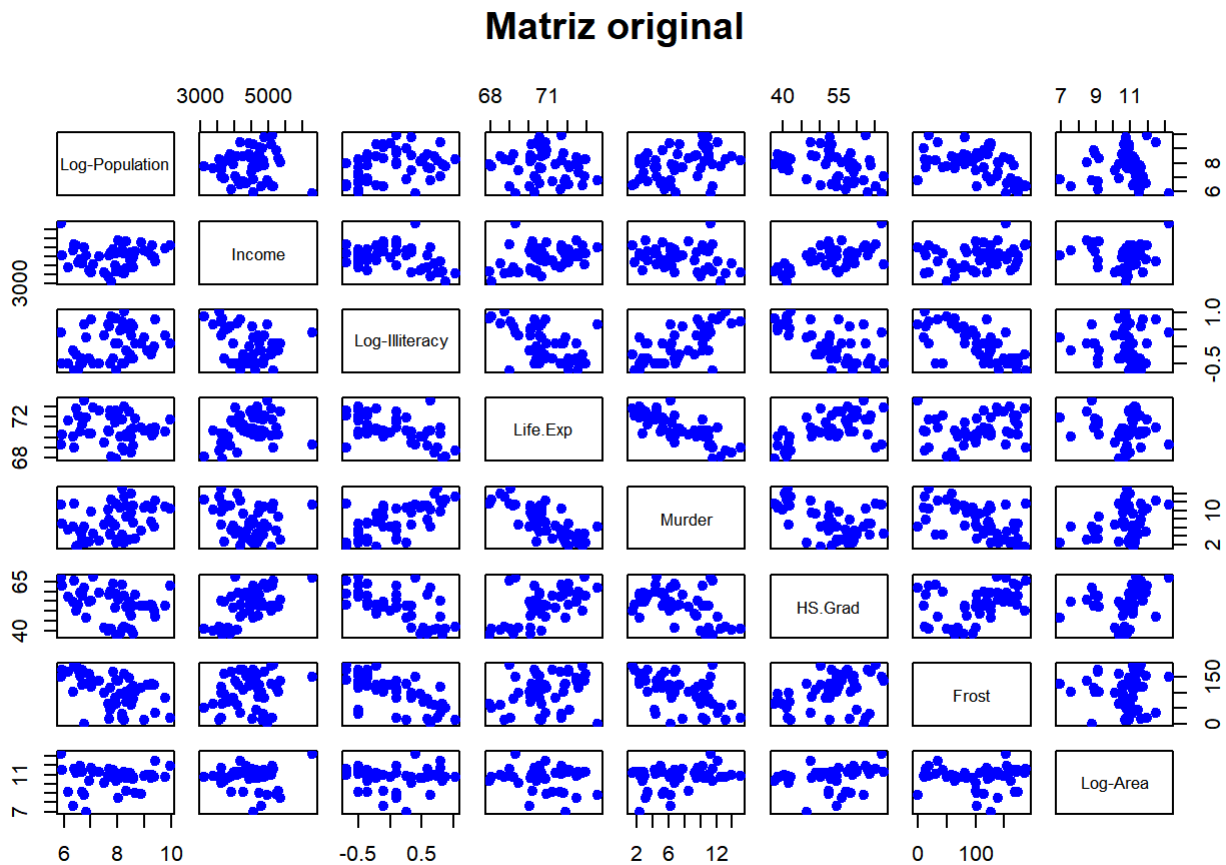
1.- Aplicamos logaritmo para las columnas 1,3 y 8

```
x[,1]<-log(x[,1])
colnames(x)[1]<-"Log-Population"
```

```
x[,3]<-log(x[,3])
colnames(x)[3]<-"Log-Illiteracy"
```

```
x[,8]<-log(x[,8])
colnames(x)[8]<-"Log-Area"
```

```
pairs(x,col="blue", pch=19, main="Matriz original")
```



Nota: Como las variables tiene diferentes unidades de medida, se va a implementar la matriz de correlaciones para estimar la matriz de carga

Reduccion de la dimensionalidad Análisis Factorial de componentes principales (PCFA)

1.- Calcular la matriz de medias y de correlaciones

```
mu<-colMeans(x)
```

```
R<-cor(x)
R
```

##	Log-Population	Income	Log-Illiteracy	Life.Exp	Murder
## Log-Population	1.00000000	0.034963788	0.28371749	-0.1092630	0.3596542
## Income	0.03496379	1.00000000	-0.35147773	0.3402553	-0.2300776
## Log-Illiteracy	0.28371749	-0.351477726	1.00000000	-0.5699943	0.6947320
## Life.Exp	-0.10926301	0.340255339	-0.56999432	1.0000000	-0.7808458
## Murder	0.35965424	-0.230077610	0.69473198	-0.7808458	1.0000000
## HS.Grad	-0.32211720	0.619932323	-0.66880911	0.5822162	-0.4879710
## Frost	-0.45809012	0.226282179	-0.67656232	0.2620680	-0.5388834
## Log-Area	0.08541473	-0.007462068	-0.05830524	-0.1086351	0.2963133

##	HS.Grad	Frost	Log-Area
## Log-Population	-0.3221172	-0.45809012	0.085414734
## Income	0.6199323	0.22628218	-0.007462068
## Log-Illiteracy	-0.6688091	-0.67656232	-0.058305240
## Life.Exp	0.5822162	0.26206801	-0.108635052
## Murder	-0.4879710	-0.53888344	0.296313252
## HS.Grad	1.0000000	0.36677970	0.196743429
## Frost	0.3667797	1.00000000	-0.021211992
## Log-Area	0.1967434	-0.02121199	1.000000000

2.- Reducción de la dimensionalidad mediante Análisis factorial de componentes principales.

1.- Calcular los valores y vectores propios.

```
eR<-eigen(R)
```

2.- Valores propios

```
eigen.val<-eR$values
```

3.- Vectores propios

```
eigen.vec<-eR$vectors
```

4.- Calcular la proporcion de variabilidad

```
prop.var<-eigen.val/sum(eigen.val)
```

5.- Calcular la proporcion de variabilidad acumulada

```
prop.var.acum<-cumsum(eigen.val)/sum(eigen.val)
```

Estimacion de la matriz de carga

Nota: se estima la matriz de carga usando los autovalores y autovectores. se aplica la rotación varimax

Primera estimación de Lamda mayuscula se calcula multiplicando la matriz de los 3 primeros autovectores por la matriz diagonal formada por la raiz cuadrada de los primeros 3 autovalores.

```
L.est.1<-eigen.vec[,1:3] %%% diag(sqrt(eigen.val[1:3]))
```

Rotación varimax

```
L.est.1.var<-varimax(L.est.1)
```

Estimación de la matriz de los errores

1.- Estimación de la matriz de perturbaciones

```
Psi.est.1<-diag(diag(R-as.matrix(L.est.1.var$loadings)%*% t(as.matrix(L.est.1.var$loadings))))
```

2.- Se utiliza el método Análisis de factor principal (PFA) para estimación de autovalores y autovectores

```
RP<-R-Psi.est.1
```

```
eRP<-eigen(RP)
```

Autovalores

```
eigen.val.RP<-eRP$values
```

Autovectores

```
eigen.vec.RP<-eRP$vectors
```

Proporcion de variabilidad

```
prop.var.RP<-eigen.val.RP/ sum(eigen.val.RP)
```

Proporcion de variabilidad acumulada

```
prop.var.RP.acum<-cumsum(eigen.val.RP)/ sum(eigen.val.RP)
```

Estimación de la matriz de cargas con rotación varimax

```
L.est.2<-eigen.vec.RP[,1:3] %*% diag(sqrt(eigen.val.RP[1:3]))
```

Rotacion varimax

```
L.est.2.var<-varimax(L.est.2)
```

Estimación de la matriz de covarianzas de los errores.

```
Psi.est.2<-diag(diag(R-as.matrix(L.est.2.var$loadings)%*% t(as.matrix(L.est.2.var$loadings))))
```

Obtencion de los scores de ambos métodos

PCFA

```
FS.est.1<-scale(x)%*% as.matrix(L.est.1.var$loadings)
```

PFA

```
FS.est.2<-scale(x)%% as.matrix (L.est.2.var$loadings)
```

graficamos ambos scores

```
par(mfrow=c(2,1))  
plot(FS.est.1[,1], FS.est.1[,2], xlab="primer factor",  
      ylab="segundo factor", main="scores con factor I y II con PCFA",  
      pch=19, col="blue")
```

