

MAT-269: Sesión 17, Modelo de Análisis Factorial II

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Esperanzas de vida en años organizados por país, edad y sexo. Los datos fueron extraídos desde Keyfitz y Flieger (1971)¹ y hace referencia la esperanza de vida en los años 60s.

La base de datos `life` tiene 31 Observaciones (países) y 8 variables, que corresponden a las esperanzas de vida, para:

- ▶ `m0`, `m25`, `m50`, `m75`: hombres a las edades de 0, 25, 50 y 75 años.
- ▶ `w0`, `w25`, `w50`, `w75`: mujeres a las edades de 0, 25, 50 y 75 años.

¹Population: Facts and Methods of Demography.



Datos de esperanza de vida

```
> source("lifeexp.R") # lee datos desde fuentes R

> life
```

	m0	m25	m50	m75	w0	w25	w50	w75
Algeria	63	51	30	13	67	54	34	15
Cameroon	34	29	13	5	38	32	17	6
Madagascar	38	30	17	7	38	34	20	7
Mauritius	59	42	20	6	64	46	25	8
Reunion	56	38	18	7	62	46	25	10
Seychelles	62	44	24	7	69	50	28	14
South Africa(C)	50	39	20	7	55	43	23	8
South Africa(W)	65	44	22	7	72	50	27	9
Tunisia	56	46	24	11	63	54	33	19
Canada	69	47	24	8	75	53	29	10
...								
Argentina	65	46	24	9	71	51	28	10
Chile	59	43	23	10	66	49	27	12
Columbia	58	44	24	9	62	47	25	10
Ecuador	57	46	28	9	60	49	28	11



Es posible llevar a cabo un test de bondad de ajuste probando la hipótesis:

$$H_0 : \Sigma = \Gamma\Gamma^\top + \Psi,$$

basado en k factores, versus la alternativa $H_1 : \Sigma$ es no restringida. Evaluando la log-verosimilitud marginal en $\Sigma = S$ y $\Sigma = \hat{\Gamma}\hat{\Gamma}^\top + \hat{\Psi}$, tenemos

$$LR = n \{ \log |\hat{\Sigma}| + \text{tr}(S\hat{\Sigma}^{-1}) - \log |S| - p \} = n \log \left(\frac{\hat{\Sigma}}{|S|} \right),$$

que tiene una distribución aproximadamente $\chi^2(\nu)$ con $\nu = \frac{1}{2}[(p-k)^2 - (p+k)]$.

Observación:

Debido a este test de hipótesis, el [algoritmo de estimación de Jöreskog](#) minimiza:

$$\log |\Sigma| + \text{tr}(S\Sigma^{-1}) - \log |S| - p.$$



Datos de esperanza de vida

```
# ajuste un modelo de análisis factorial
> fa1 <- factanal(life, factors = 1, method = "mle")
> fa2 <- factanal(life, factors = 2, method = "mle")
> fa3 <- factanal(life, factors = 3, method = "mle")

# elementos en un objeto de clase 'factanal'
> attributes(fa3)
$names
 [1] "converged"      "loadings"      "uniquenesses"  "correlation"
 [5] "criteria"      "factors"      "dof"           "method"
 [9] "rotmat"       "STATISTIC"    "PVAL"         "n.obs"
[13] "call"

$class
[1] "factanal"

# extrayendo algunos objetos
> fa3[c("STATISTIC", "PVAL")]
$STATISTIC
objective
 6.727248

$PVAL
objective
0.4578204
```



Los resultados del test de bondad de ajuste en la siguiente Tabla, permiten notar que la solución usando $k = 3$ factores es adecuado para los datos

k	Estadístico	gl	valor- p
1	163.1121	20	< 0.0001
2	45.2357	13	< 0.0001
3	6.7272	7	0.4578



Datos de esperanza de vida

```
> fa3
```

```
Call:
```

```
factanal(x = life, factors = 3, method = "mle")
```

```
Uniquenesses:
```

m0	m25	m50	m75	w0	w25	w50	w75
0.005	0.362	0.066	0.288	0.005	0.011	0.020	0.146

```
Loadings:
```

	Factor1	Factor2	Factor3
m0	0.964	0.122	0.226
m25	0.646	0.169	0.438
m50	0.430	0.354	0.790
m75		0.525	0.656
w0	0.970	0.217	
w25	0.764	0.556	0.310
w50	0.536	0.729	0.401
w75	0.156	0.867	0.280

	Factor1	Factor2	Factor3
SS loadings	3.375	2.082	1.640
Proportion Var	0.422	0.260	0.205
Cumulative Var	0.422	0.682	0.887

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458



Datos de esperanza de vida

Recuerde que $\widehat{\Sigma} = \widehat{\Gamma}\widehat{\Gamma}^\top + \widehat{\Psi}$, y las comunalidades² fueron definidas como:

$$h_i^2 = \sum_{j=1}^k \gamma_{ij}^2.$$

```
> Gamma <- unclass(fa3$loadings)
> Gamma
      Factor1 Factor2 Factor3
m0  0.96395  0.1216  0.22623
m25 0.64627  0.1689  0.43830
m50 0.42955  0.3536  0.79004
m75 0.07999  0.5246  0.65627
w0  0.97027  0.2171  0.08093
w25 0.76414  0.5559  0.30988
w50 0.53584  0.7293  0.40115
w75 0.15615  0.8666  0.28044

> communality <- apply(Gamma^2, 1, sum)
> communality
      m0      m25      m50      m75      w0      w25      w50      w75
0.9952 0.6383 0.9337 0.7122 0.9951 0.9889 0.9799 0.8540
```

²Note que $h^2 = \text{diag}(\widehat{\Gamma}\widehat{\Gamma}^\top)$



Interpretación:

- ▶ El 1er factor es dominado por la “esperanza de vida al nacer”.
- ▶ El 2do factor refleja la “esperanza de vida de la vejez”.
- ▶ El 3er factor tiene una mayor carga para hombres entre 50 y 75 años y puede entenderse como “esperanza de vida en hombres mayores”.

Observación:

Al rotular los “factores” se puede ser bastante creativo. De modo que, tendrían que tener algún sentido.



Datos de esperanza de vida

Es posible **estimar** (predecir) los factores usando la siguiente distribución condicional,

$$z|x \sim N_m(\Gamma^T \Sigma^{-1}(x - \mu), (\Gamma^T \Psi^{-1} \Gamma + I)^{-1}),$$

de este modo,

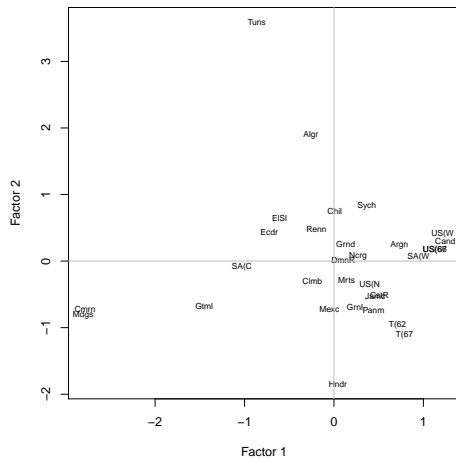
$$\hat{z}_i = \hat{\Gamma}^T (\hat{\Gamma} \hat{\Gamma}^T + \Psi)^{-1} (x_i - \bar{x}), \quad i = 1, \dots, n.$$

```
# objeto 'fa3' contendrá elemento 'scores'
> fa3 <- factanal(life, factors = 3, method = "mle",
+   scores = "regression")
> scores <- fa3$scores
> scores
```

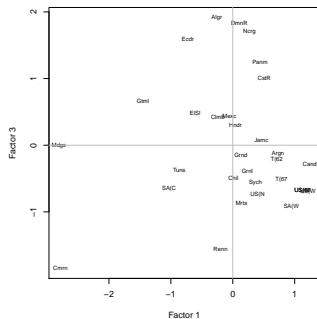
	Factor1	Factor2	Factor3
Algeria	-0.25806	1.9010	1.9158
Cameroon	-2.78250	-0.7234	-1.8477
Madagascar	-2.80643	-0.8116	-0.0121
...			
Argentina	0.73134	0.2481	-0.1282
Chile	0.00975	0.7522	-0.4920
Columbia	-0.24060	-0.2954	0.4292
Ecuador	-0.72345	0.4425	1.5916



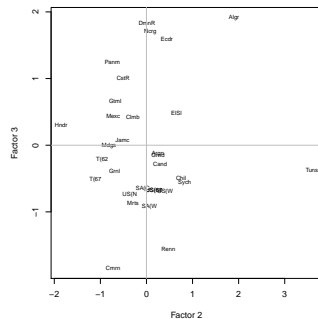
Datos de esperanza de vida



Datos de esperanza de vida



(a)



(b)



Gráficos anteriores se realizaron mediante los comandos:

```
> par(pty = "s")
> plot(scores[,1], scores[,2], type = "n", xlab = "Factor 1",
+      ylab = "Factor 2")
> text(scores[,1], scores[,2], labels = abbreviate(row.names(scores)),
+      cex = 0.7, lwd = 2)
> abline(h = 0, col = "gray")
> abline(v = 0, col = "gray")
```



Datos de esperanza de vida

```
# objetos con parámetro de 'rotación'
> fa3.none <- factanal(life, factors = 3, method = "mle",
+   rotation = "none")
> fa3.varimax <- factanal(life, factors = 3, method = "mle",
+   rotation = "varimax")
> fa3.promax <- factanal(life, factors = 3, method = "mle",
+   rotation = "promax")

# factores sin rotar
> par(pty = "s")
> plot(fa3.none$loadings[,1], fa3.none$loadings[,2], xlab = "Factor 1",
+   ylab = "Factor 2", xlim = c(-1,1), ylim = c(-1,1))
> text(fa3.none$loadings[,1], fa3.none$loadings[,2], pos = 3,
+   lab = colnames(life), col = "blue")
> abline(h = 0, v = 0, col = "gray")

# rotación varimax
> par(pty = "s")
> plot(fa3.varimax$loadings[,1], fa3.varimax$loadings[,2],
+   xlab = "Factor 1", ylab = "Factor 2", xlim = c(-1,1),
+   ylim = c(-1,1))
> text(fa3.varimax$loadings[,1], fa3.varimax$loadings[,2], pos = 3,
+   lab = colnames(life), col = "blue")
> abline(h = 0, v = 0, col = "gray")

# y análogamente para rotación 'promax'
```

