

Análisis de Factores (FA)

José A. Perusquía Cortés

Análisis Multivariado Semestre 2023-2



▸ Orígenes:

Spearman (1904). "General-Intelligence," Objectively Determined and Measured. *American Journal of Psychology*. **15** (2): 201–293.

Thurstone (1931). Multiple factor analysis. *Psychological Review*. **38** (5): 406–427.

Thurstone (1934). The Vectors of Mind. *The Psychological Review*. **41**: 1–32.

- Orígenes:

Spearman (1904). "General-Intelligence," Objectively Determined and Measured. *American Journal of Psychology*. **15** (2): 201–293.

Thurstone (1931). Multiple factor analysis. *Psychological Review*. **38** (5): 406–427.

Thurstone (1934). The Vectors of Mind. *The Psychological Review*. **41**: 1–32.

- El análisis de factores es un modelo matemático que busca:

Explicar la correlación de un conjunto de p variables a través de m factores latentes.

Ejemplo (Spearman)

▸ Correlaciones entre el aprovechamiento académico de:

1. Estudios clásicos

2. Francés

3. Inglés

4. Matemáticas

$$R = \begin{pmatrix} 1 & .83 & .78 & .70 \\ & 1 & .67 & .67 \\ & & 1 & .64 \\ & & & 1 \end{pmatrix}$$

Ejemplo (Spearman)

- Podemos modelarlo como:

$$x_1 = \lambda_1 f + u_1$$

$$x_2 = \lambda_2 f + u_2$$

$$x_3 = \lambda_3 f + u_3$$

$$x_4 = \lambda_4 f + u_4$$

Donde, f es un factor común ("general intelligence ability")

Ejemplo (Spearman)

- Podemos modelarlo como:

$$\begin{array}{l} x_1 = \lambda_1 f + u_1 \\ x_2 = \lambda_2 f + u_2 \\ x_3 = \lambda_3 f + u_3 \\ x_4 = \lambda_4 f + u_4 \end{array} \longrightarrow \begin{array}{l} x_1 = .955f + u_1 \\ x_2 = .875f + u_2 \\ x_3 = .811f + u_3 \\ x_4 = .748f + u_4 \end{array}$$

Donde, f es un factor común ("general intelligence ability") y

$$\text{Var}(u_1) = .088, \text{Var}(u_2) = .251, \text{Var}(u_3) = .342, \text{Var}(u_4) = .440$$

- Sea $\mathbf{x}_{p \times 1}$ un vector aleatorio con $\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Var}(\mathbf{x}) = \boldsymbol{\Sigma}$ entonces

$$\mathbf{x} = \boldsymbol{\Lambda} \mathbf{f} + \mathbf{u} + \boldsymbol{\mu}$$

- Sea $\mathbf{x}_{p \times 1}$ un vector aleatorio con $\mathbb{E}(\mathbf{x}) = \mu$ y $\text{Var}(\mathbf{x}) = \Sigma$ entonces

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u} + \mu$$

Donde

1. $\Lambda_{p \times k}$ es una matriz de constantes (factor loadings)
2. $\mathbf{f}_{k \times 1}$ es un vector aleatorio de factores comunes
3. $\mathbf{u}_{p \times 1}$ es un vector aleatorio de "errores"

- Supuestos

$$\mathbb{E}(\mathbf{f}) = \mathbf{0}$$

$$\text{Var}(\mathbf{f}) = \mathbf{I}$$

- Supuestos

$$\mathbb{E}(\mathbf{f}) = \mathbf{0}$$

$$\text{Var}(\mathbf{f}) = \mathbf{I}$$

$$\mathbb{E}(\mathbf{u}) = \mathbf{0}$$

$$\text{Var}(\mathbf{u}) = \Psi = \text{diag}(\Psi_{11}, \dots, \Psi_{pp})$$

- Supuestos

$$\mathbb{E}(\mathbf{f}) = \mathbf{0}$$

$$\text{Var}(\mathbf{f}) = \mathbf{I}$$

$$\mathbb{E}(\mathbf{u}) = \mathbf{0}$$

$$\text{Var}(\mathbf{u}) = \Psi = \text{diag}(\Psi_{11}, \dots, \Psi_{pp})$$

$$\text{Cov}(\mathbf{f}, \mathbf{u}) = 0$$

▸ Así

$$x_i = \sum_{j=1}^k \lambda_{ij} f_j + u_i + \mu_i$$

$\text{Var}(x_i)$

$\text{Cov}(x_i, x_j)$

▸ Así

$$x_i = \sum_{j=1}^k \lambda_{ij} f_j + u_i + \mu_i$$

$$\text{Var}(x_i) = \sum_{j=1}^k \lambda_{ij}^2 + \Psi_{ii} = h_i^2 + \Psi_{ii}$$

$$\text{Cov}(x_i, x_k) = \sum_{j=1}^k \lambda_{ij} \lambda_{jk}$$

▸ Por lo que

$$\Sigma = \Lambda \Lambda^T + \Psi$$

- Invariante ante cambios de escala

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

$$\mathbf{C} = \text{diag}(c_1, \dots, c_p)$$

- Invariante ante cambios de escala

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

$$\mathbf{C} = \text{diag}(c_1, \dots, c_p)$$



$$\mathbf{y} = \mathbf{C}\mathbf{\Lambda}\mathbf{f} + \mathbf{C}\mathbf{u} + \mathbf{C}\boldsymbol{\mu}$$

- Invariante ante cambios de escala

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

$$\mathbf{C} = \text{diag}(c_1, \dots, c_p)$$



$$\mathbf{y} = \mathbf{C}\mathbf{\Lambda}\mathbf{f} + \mathbf{C}\mathbf{u} + \mathbf{C}\boldsymbol{\mu}$$



$$\text{Var}(\mathbf{y}) = \mathbf{C}\mathbf{\Lambda}\mathbf{\Lambda}^T\mathbf{C} + \mathbf{C}\boldsymbol{\Psi}\mathbf{C} = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}$$

- Λ no es única

$$\mathbf{x} = (\Lambda \mathbf{G})(\mathbf{G}^T \mathbf{f}) + \mathbf{u} + \mu$$

- Λ no es única

$$\mathbf{x} = (\Lambda \mathbf{G})(\mathbf{G}^T \mathbf{f}) + \mathbf{u} + \mu$$



$$\Sigma = (\Lambda \mathbf{G})(\mathbf{G}^T \Lambda^T) + \Psi$$

- Λ no es única

$$\mathbf{x} = (\Lambda \mathbf{G})(\mathbf{G}^T \mathbf{f}) + \mathbf{u} + \mu$$



$$\Sigma = (\Lambda \mathbf{G})(\mathbf{G}^T \Lambda^T) + \Psi$$

- Añadimos una restricción

$$G = \Lambda^T \Psi^{-1} \Lambda = \text{diag}(g_{11}, \dots, g_{pp})$$

$$g_{11} > g_{22} > \dots > g_{pp}$$

- La solución dependerá de

$$\begin{aligned}s &= \frac{p(p+1)}{2} - \left[p + pk - \frac{k(k-1)}{2} \right] \\ &= \frac{1}{2}[(p-k)^2 - (p+k)]\end{aligned}$$

- La solución dependerá de

$$\begin{aligned}s &= \frac{p(p+1)}{2} - \left[p + pk - \frac{k(k-1)}{2} \right] \\ &= \frac{1}{2}[(p-k)^2 - (p+k)]\end{aligned}$$

1. Si $s < 0$ hay una infinidad de soluciones (no interesa este caso)

- La solución dependerá de

$$\begin{aligned}s &= \frac{p(p+1)}{2} - \left[p + pk - \frac{k(k-1)}{2} \right] \\ &= \frac{1}{2}[(p-k)^2 - (p+k)]\end{aligned}$$

1. Si $s < 0$ hay una infinidad de soluciones (no interesa este caso)
2. Si $s = 0$ una única solución (no siempre viable y no interesa)

- La solución dependerá de

$$\begin{aligned}s &= \frac{p(p+1)}{2} - \left[p + pk - \frac{k(k-1)}{2} \right] \\ &= \frac{1}{2}[(p-k)^2 - (p+k)]\end{aligned}$$

1. Si $s < 0$ hay una infinidad de soluciones (no interesa este caso)
2. Si $s = 0$ una única solución (no siempre viable y no interesa)
3. Si $s > 0$ no hay solución exacta y se estima (caso interesante)

- ¿Cómo estimar Λ, Ψ a partir de \mathbf{S} ?

- ¿Cómo estimar Λ, Ψ a partir de \mathbf{S} ?
- Buscar $\hat{\Lambda}, \hat{\Psi}$ tales que $\mathbf{S} = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$

- ¿Cómo estimar Λ, Ψ a partir de \mathbf{S} ?
- Buscar $\hat{\Lambda}, \hat{\Psi}$ tales que $\mathbf{S} = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$
- Dado un estimador $\hat{\Lambda}$ podemos hacer

$$\hat{\Psi}_{ii} = \mathbf{S}_{ii} - \sum_{j=1}^k \hat{\lambda}_{ij}^2$$

- ¿Cómo estimar Λ, Ψ a partir de \mathbf{S} ?
- Buscar $\hat{\Lambda}, \hat{\Psi}$ tales que $\mathbf{S} = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$
- Dado un estimador $\hat{\Lambda}$ podemos hacer

$$\hat{\Psi}_{ii} = S_{ii} - \sum_{j=1}^k \hat{\lambda}_{ij}^2$$

- O equivalentemente,

$$\hat{\sigma}_{ii} = S_{ii}$$

- Para trabajar con **R**

- Para trabajar con **R**

$$\mathbf{Y} = \mathbf{HXD}^{-\frac{1}{2}}$$

- Para trabajar con **R**

$$\mathbf{Y} = \mathbf{H}\mathbf{X}\mathbf{D}^{-\frac{1}{2}}$$



$$\widehat{\Lambda}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Lambda}_x$$

$$\widehat{\Psi}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Psi}_x$$

- Para trabajar con **R**

$$\mathbf{Y} = \mathbf{H}\mathbf{X}\mathbf{D}^{-\frac{1}{2}}$$



$$\widehat{\Lambda}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Lambda}_x$$

$$\widehat{\Psi}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Psi}_x$$



$$\mathbf{R} = \widehat{\Lambda}_y \widehat{\Lambda}_y^T + \widehat{\Psi}_y$$

- Para trabajar con **R**

$$\mathbf{Y} = \mathbf{H}\mathbf{X}\mathbf{D}^{-\frac{1}{2}}$$



$$\widehat{\Lambda}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Lambda}_x$$

$$\widehat{\Psi}_y = \mathbf{D}^{-\frac{1}{2}} \widehat{\Psi}_x$$



$$\mathbf{R} = \widehat{\Lambda}_y \widehat{\Lambda}_y^T + \widehat{\Psi}_y$$



$$\widehat{\Psi}_{ii} = 1 - \sum_{j=1}^k \widehat{\lambda}_{ij}^2$$

Análisis de Factores Principales

- Técnica de descomposición con valores propios para la matriz de correlación reducida

$$\mathbf{R} - \widehat{\Psi} \quad \text{diag} \left(\mathbf{R} - \widehat{\Psi} \right) = \left(\hat{h}_1^2, \dots, \hat{h}_p^2 \right)$$

Análisis de Factores Principales

- Técnica de descomposición con valores propios para la matriz de correlación reducida

$$\mathbf{R} - \widehat{\Psi} \quad \text{diag} \left(\mathbf{R} - \widehat{\Psi} \right) = \left(\hat{h}_1^2, \dots, \hat{h}_p^2 \right)$$

- Primero hay que estimar \hat{h}_i^2

Análisis de Factores Principales

- Técnica de descomposición con valores propios para la matriz de correlación reducida

$$\mathbf{R} - \widehat{\Psi} \quad \text{diag} \left(\mathbf{R} - \widehat{\Psi} \right) = \left(\hat{h}_1^2, \dots, \hat{h}_p^2 \right)$$

- Primero hay que estimar \hat{h}_i^2
 1. El cuadrado del coeficiente de correlación múltiple de la i-ésima variable con el resto de las variables

Análisis de Factores Principales

- Técnica de descomposición con valores propios para la matriz de correlación reducida

$$\mathbf{R} - \widehat{\Psi} \quad \text{diag} \left(\mathbf{R} - \widehat{\Psi} \right) = \left(\hat{h}_1^2, \dots, \hat{h}_p^2 \right)$$

- Primero hay que estimar \hat{h}_i^2
 1. El cuadrado del coeficiente de correlación múltiple de la i-ésima variable con el resto de las variables
 2. El coeficiente de correlación más grande entre la i-ésima variable y alguna de las otras, i.e., $\max_{j \neq i} |r_{ij}|$

- Por el teorema de descomposición espectral se tiene que

$$\mathbf{R} - \widehat{\Psi} = \sum_{i=1}^p \alpha_i \gamma_{(i)} \gamma_{(i)}^T$$

- Por el teorema de descomposición espectral se tiene que

$$\mathbf{R} - \widehat{\Psi} = \sum_{i=1}^p \alpha_i \gamma_{(i)} \gamma_{(i)}^T$$

- Donde
 1. $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p$ son los eigenvalores
 2. $\gamma_{(1)}, \dots, \gamma_{(p)}$ los eigenvectores

- Por el teorema de descomposición espectral se tiene que

$$\mathbf{R} - \widehat{\Psi} = \sum_{i=1}^p \alpha_i \gamma_{(i)} \gamma_{(i)}^T$$

- Donde
 1. $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p$ son los eigenvalores
 2. $\gamma_{(1)}, \dots, \gamma_{(p)}$ los eigenvectores

$$\text{Así } \widehat{\Lambda} = \Gamma_k \mathbf{A}_k^{\frac{1}{2}} \text{ donde } \Gamma_k = \left(\gamma_{(1)}, \dots, \gamma_{(k)} \right), \mathbf{A}_k = \text{diag}(\alpha_1, \dots, \alpha_k) \text{ y } \widehat{\Psi}_{ii} = 1 - \sum_{j=1}^k \widehat{\lambda}_{ij}^2$$

- 88 calificaciones de 5 exámenes a libro abierto o cerrado.

Lineal (C)	Estadística (C)	Probabilidad(A)	Finanzas (A)	Cálculo (A)
97	92	77	72	96
83	88	90	75	96
95	83	81	71	96
75	82	73	75	83
83	73	75	75	78
73	71	82	69	88
71	77	75	70	83

- La matriz de correlación es

$$R = \begin{pmatrix} 1 & .546 & .545 & .410 & .390 \\ & 1 & .613 & .489 & .449 \\ & & 1 & .712 & .666 \\ & & & 1 & .666 \\ & & & & 1 \end{pmatrix}$$

- ¿Hasta cuántos factores podemos considerar?

- Para $k = 1$

Variable	λ_i	h_i^2	Ψ_{ii}
Lineal	0.61	0.37	0.63
Estadística	0.69	0.48	0.52
Probabilidad	0.91	0.83	0.17
Finanzas	0.76	0.58	0.42
Cálculo	0.72	0.51	0.49

- ¿Qué interpretación le pueden dar?

- Para $k = 2$

Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.64	0.33	0.51	0.49
Estadística	0.71	0.28	0.59	0.41
Probabilidad	0.90	-0.08	0.81	0.19
Finanzas	0.77	-0.23	0.65	0.35
Cálculo	0.72	-0.22	0.57	0.43

- ¿Qué interpretación le pueden dar?

- Asumiendo que $\mathbf{x} \sim N(\mu, \Sigma)$ podemos pensar en maximizar la log-verosimilitud

$$L = \frac{n}{2}(\log |2\pi\Sigma|) - \frac{n}{2}\text{tr}(\Sigma^{-1}\mathbf{S})$$

- Asumiendo que $\mathbf{x} \sim N(\mu, \Sigma)$ podemos pensar en maximizar la log-verosimilitud

$$L = \frac{n}{2}(\log |2\pi\Sigma|) - \frac{n}{2}\text{tr}(\Sigma^{-1}\mathbf{S})$$

- O equivalentemente minimizar

$$F = \text{tr}(\Sigma^{-1}\mathbf{S}) - \log |\Sigma^{-1}\mathbf{S}| - p$$

- Podemos hacer pruebas de hipótesis para el número de factores

$$U = n' \min(F) \qquad n' = n - 1 - \frac{1}{6}(2p + 5) - \frac{2}{3}k$$

- Podemos hacer pruebas de hipótesis para el número de factores

$$U = n' \min(F)$$

$$n' = n - 1 - \frac{1}{6}(2p + 5) - \frac{2}{3}k$$

- Si k son suficientes entonces

$$U \sim \chi^2_\nu$$

$$\nu = \frac{1}{2}(p - k)^2 - \frac{1}{2}(p + k)$$

▸ Para $k = 1$

MLE

Variable	λ_i	h_i^2	Ψ_{ii}
Lineal	0.60	0.36	0.64
Estadística	0.67	0.45	0.55
Probabilidad	0.92	0.84	0.16
Finanzas	0.77	0.60	0.40
Cálculo	0.73	0.53	0.47

PA

Variable	λ_i	h_i^2	Ψ_{ii}
Lineal	0.61	0.37	0.63
Estadística	0.69	0.48	0.52
Probabilidad	0.91	0.83	0.17
Finanzas	0.76	0.58	0.42
Cálculo	0.72	0.51	0.49

▸ Para $k = 2$

MLE

Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.62	0.38	0.53	0.47
Estadística	0.70	0.29	0.57	0.43
Probabilidad	0.90	-0.05	0.81	0.19
Finanzas	0.78	-0.20	0.65	0.35
Cálculo	0.73	-0.19	0.57	0.43

PA

Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.64	0.33	0.51	0.49
Estadística	0.71	0.28	0.59	0.41
Probabilidad	0.90	-0.08	0.81	0.19
Finanzas	0.77	-0.23	0.65	0.35
Cálculo	0.72	-0.22	0.57	0.43

- Hacemos la prueba de hipótesis para $k = 1$

$$U = n' \min(F) = 7.742228$$

$$\chi^2_{5,.95} = 11.0705$$

- Por lo que no rechazamos la hipótesis nula de que un factor es suficiente

- Para tener unicidad en los loadings metimos una restricción ortogonal

$$G = \Lambda^T \Psi^{-1} \Lambda = \text{diag}(g_{11}, \dots, g_{pp})$$

$$g_{11} > g_{22} > \dots > g_{pp}$$

- Para tener unicidad en los loadings metimos una restricción ortogonal

$$G = \Lambda^T \Psi^{-1} \Lambda = \text{diag}(g_{11}, \dots, g_{pp}) \quad g_{11} > g_{22} > \dots > g_{pp}$$

- Para la interpretación de los factores es preferible que:
 1. Cada variable esté asociada fuertemente a lo más a un factor
 2. Los loadings son muy grandes y positivos o cercanos al cero con algunos valores intermedios

Rotación Varimax

- Propuesta por Kaiser (1958).

- Propuesta por Kaiser (1958).
- Sea Λ la matriz de loadings y \mathbf{G} matriz ortogonal entonces la matriz de loadings rotados es:

$$\Delta = \Lambda \mathbf{G}$$

- Propuesta por Kaiser (1958).
- Sea Λ la matriz de loadings y \mathbf{G} matriz ortogonal entonces la matriz de loadings rotados es:

$$\Delta = \Lambda \mathbf{G}$$

- Objetivo: Maximizar ϕ

$$\phi = \sum_{j=1}^k \sum_{i=1}^p (d_{ij}^2 - \bar{d}_j)^2 \quad d_{ij} = \frac{\delta_{ij}}{h_i} \quad \bar{d}_j = p^{-1} \sum_{i=1}^p d_{ij}^2$$

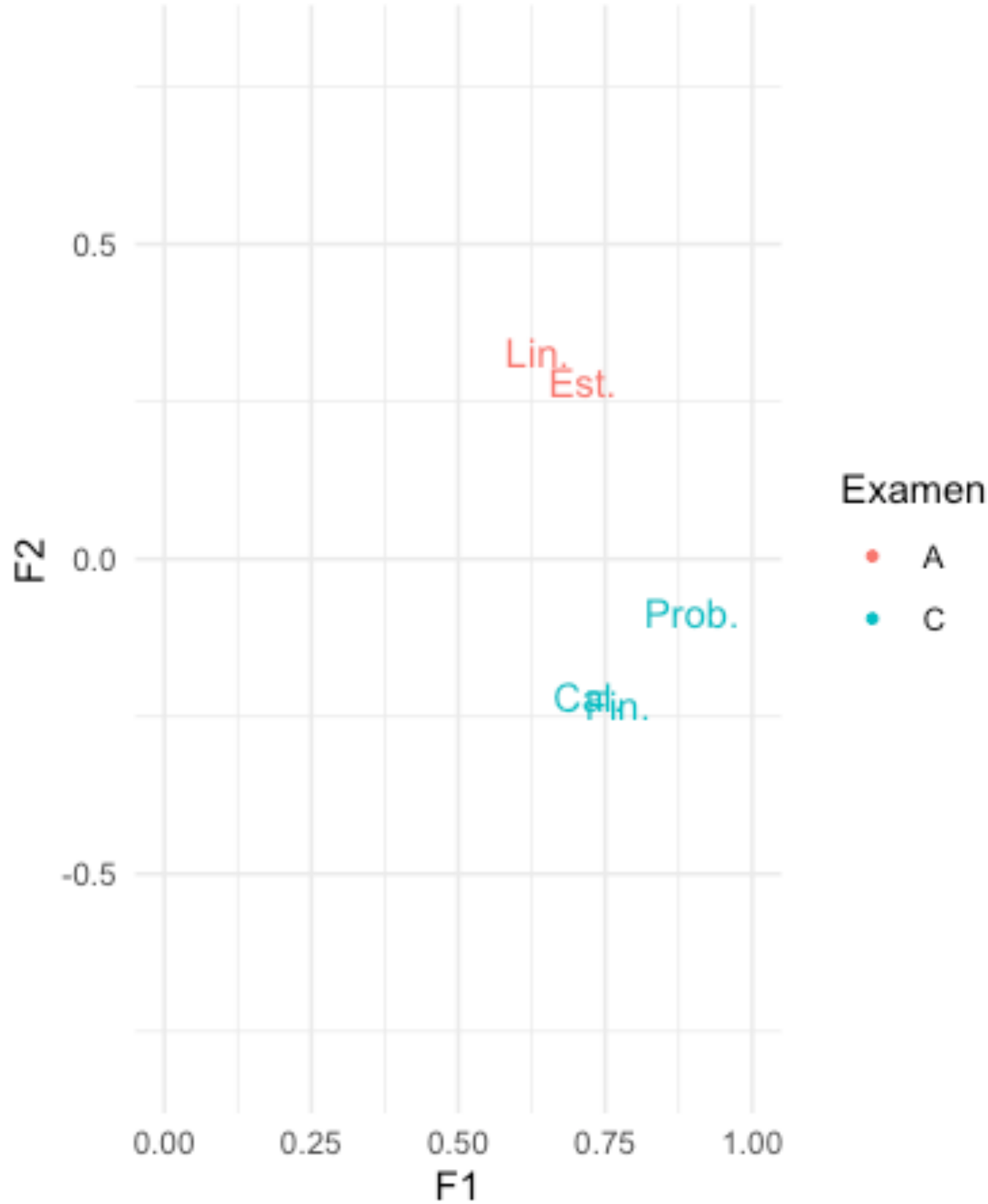
▸ Para $k = 2$

PA con rotación varimax

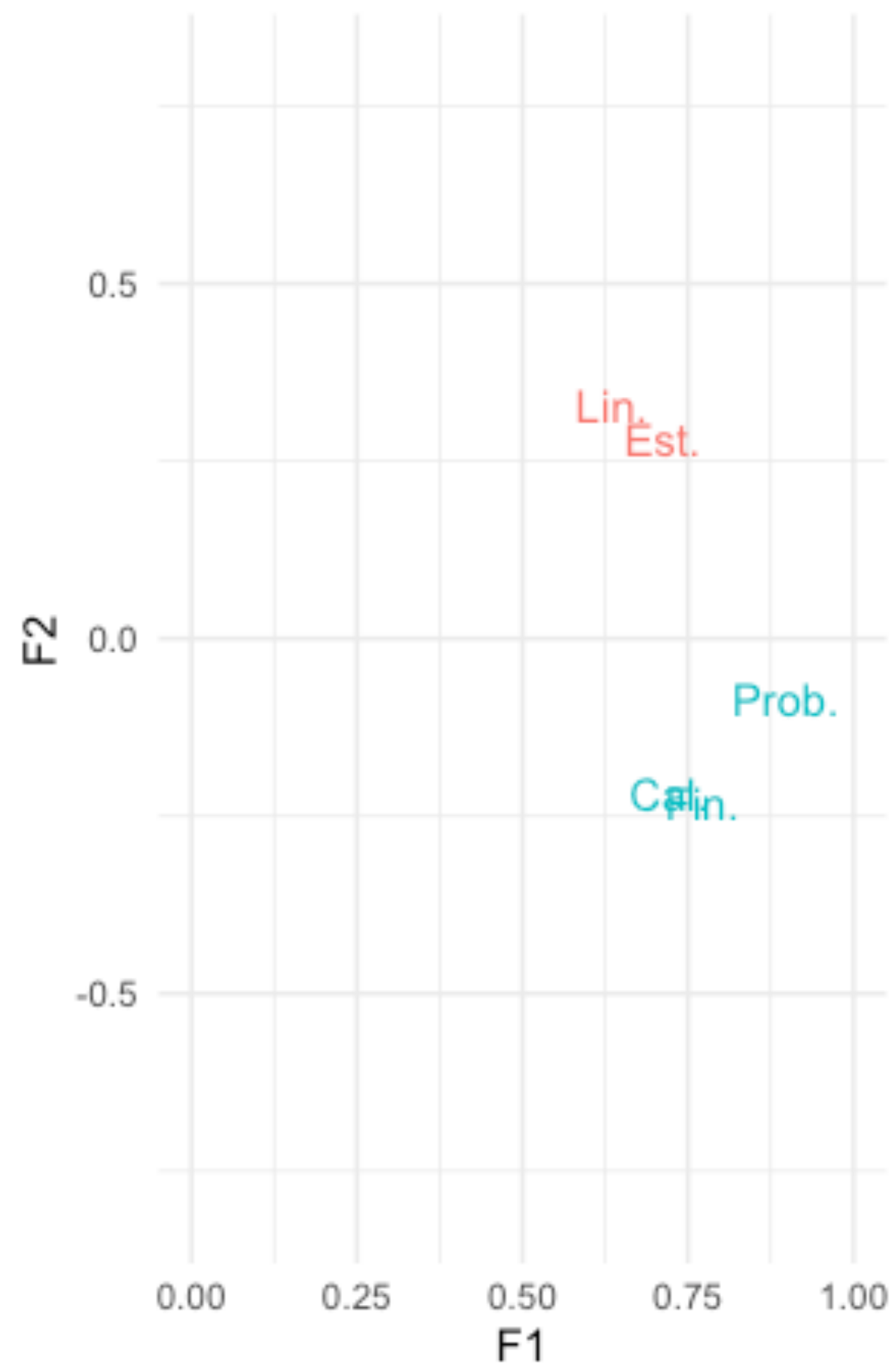
Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.27	0.66	0.51	0.49
Estadística	0.36	0.68	0.59	0.41
Probabilidad	0.74	0.52	0.81	0.19
Finanzas	0.74	0.32	0.65	0.35
Cálculo	0.69	0.30	0.57	0.43

PA sin rotación

Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.64	0.33	0.51	0.49
Estadística	0.71	0.28	0.59	0.41
Probabilidad	0.90	-0.08	0.81	0.19
Finanzas	0.77	-0.23	0.65	0.35
Cálculo	0.72	-0.22	0.57	0.43

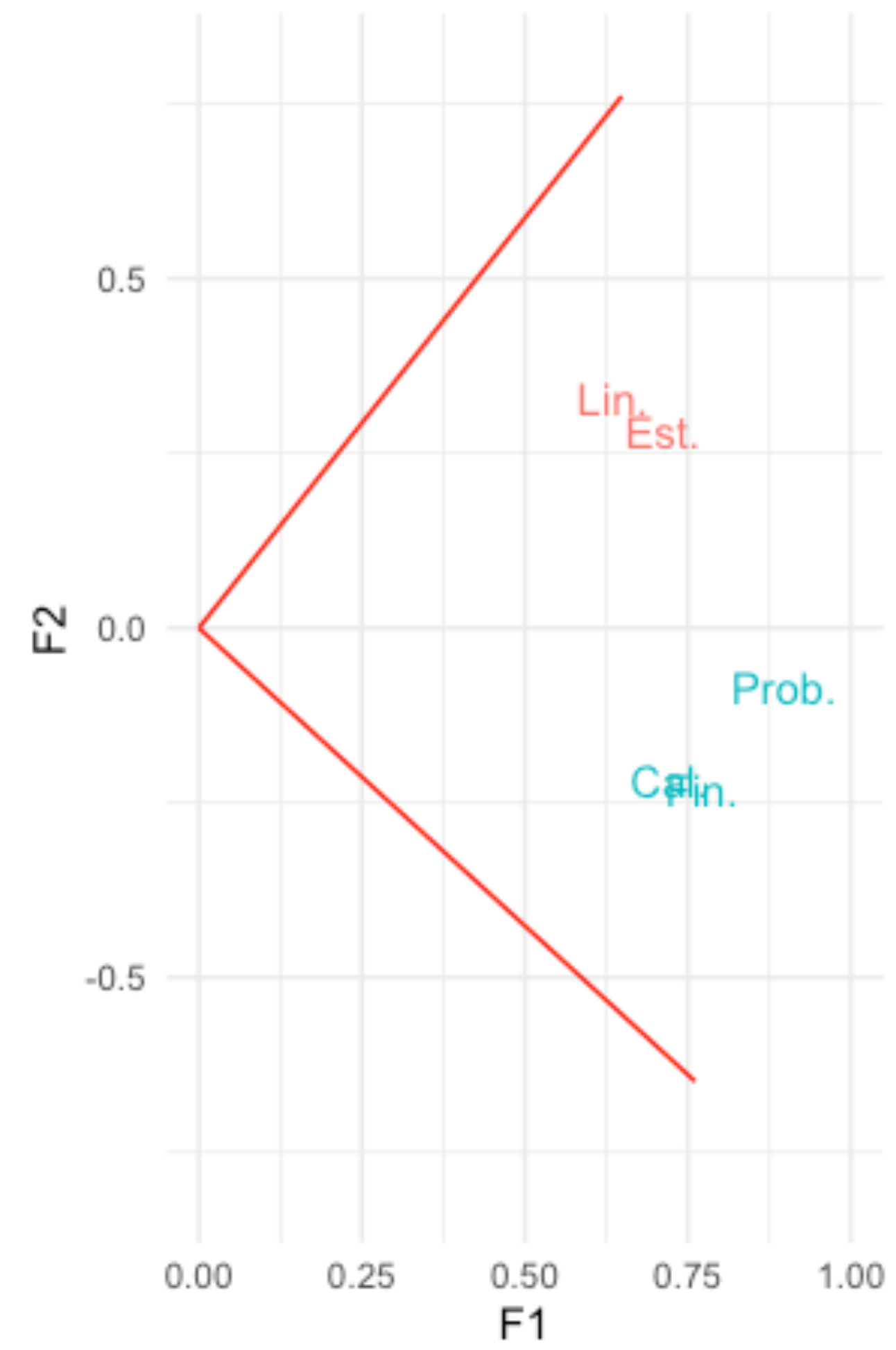


Ejemplo



Examen

- A
- C



Examen

- A
- C

Ejemplo

