Análisis de Factores (FA)

José A. Perusquía Cortés

Análisis Multivariado, Semestre 2025-II

Motivación

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

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

- Orígenes en psicología:

  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.

Formulación

- 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
  2. $\mathbf{f}_{k\times 1}$ es un vector aleatorio de factores comunes
  3. $\mathbf{u}_{p\times 1}$ es un vector aleatorio de factores únicos

- Así, para cada variable se tiene la representación

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

Formulación

- **Supuestos**

1. $\mathbb{E}(\mathbf{f}) = \mathbf{0}$
2. $\mathrm{Var}(\mathbf{f}) = \mathbf{I}$
3. $\mathbb{E}(\mathbf{u}) = \mathbf{0}$
4. $\mathrm{Var}(\mathbf{u}) = \boldsymbol{\Psi} = \mathrm{diag}\left(\Psi_{11}, \ldots, \Psi_{pp}\right)$
5. $\mathrm{Cov}(\mathbf{f}, \mathbf{u}) = \mathbf{O}$

Propiedades

Proposición 1

Dado el modelo de análisis de factores, se tiene que la varianza de cada x_i es

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

y la covarianza de x_i y x_l es

$$\text{Cov}\left(x_i, x_l\right) = \sum_{j=1}^{k} \lambda_{ij}\lambda_{jl}\,.$$

Por lo que, $\Sigma = \Lambda\Lambda^T + \Psi$

Propiedades

- **Invariante ante cambios de escala**, i.e. para $\mathbf{y} = \mathbf{C}\mathbf{x}$ con, $\mathbf{C} = \text{diag}(c_1, \ldots, c_p)$, se cumple

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

- Λ **no es única**, ya que si $\mathbf{G} \in \mathcal{O}(k)$ entonces $\mathbf{x} = (\Lambda\mathbf{G})(\mathbf{G}^T\mathbf{f}) + \mathbf{u} + \mu$ y así

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

- Se necesita una restricción en términos de una rotación, $\Lambda^T\Psi^{-1}\Lambda$ sea diagonal con los elementos de Ψ^{-1} ordenados de mayor a menor

Estimación

Dada una matriz de datos $\mathbf{X}$ ¿Cómo se puede estimar a Λ y a Ψ a partir de $\mathbf{S}$?

- Buscar $\widehat{\Lambda}$, $\widehat{\Psi}$ tales que $\mathbf{S} = \widehat{\Lambda}\,\widehat{\Lambda}^T + \widehat{\Psi}$ y $\widehat{\Psi}_{ii} \geq 0$

- Dado un estimador $\widehat{\Lambda}$ podemos hacer

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

Estimación

- La solución dependerá de la diferencia entre los grados de libertad de Σ y de Λ, Ψ

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

Observación 1

- Si $s < 0$: Hay una infinidad de soluciones (no es interesante)
- Si $s = 0$: existe una única solución (no siempre es viable)
- Si $s > 0$: no hay solución exacta y debe aproximarse (caso interesante)

Estimación

Proposición 2

Si $\mathbf{Y} = \mathbf{HXD}^{-\frac{1}{2}}$ entonces considerando el modelo de análisis de factores se cumple

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

donde

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

Observación 2

Con la matriz de correlaciones se tiene que, $\widehat{\Psi}_{ii} = 1 - \sum_{j=1}^{k} \widehat{\lambda}_{ij}^2$

Estimación por factores principales

- Aplicar una descomposición de valores propios para la matriz de correlación reducida

$$\mathbf{R} - \widehat{\mathbf{\Psi}} \qquad\qquad \mathrm{diag}\left(\mathbf{R} - \widehat{\mathbf{\Psi}}\right) = \left(\hat{h}_1^2, \ldots, \hat{h}_p^2\right)$$

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

Estimación por factores principales

- 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, $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_p$ son los eigenvalores y $\gamma_1, \ldots, \gamma_p$ los eigenvectores

- Así $\widehat{\Lambda} = \Gamma_k \mathbf{A}_k^{\frac{1}{2}}$, donde $A_k = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$

- Finalmente, volver a estimar a $\widehat{\Psi}_{ii}$ como

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

- En R: la función fa() de la librería psych

Ejemplo 1 (Spearman)

- Correlaciones entre el aprovechamiento académico de:

  - Estudios clásicos

  - Francés

  - Inglés

$$R = \begin{pmatrix} 1 & 0.83 & 0.78 \\ & 1 & 0.67 \\ & & 1 \end{pmatrix}$$

- Considerando un solo factor se tiene una solución exacta

Ejemplo 1 (Spearman)

- El modelo queda definido como:

$$x_1 = 0.983 \cdot f + u_1$$
$$x_2 = 0.844 \cdot f + u_2$$
$$x_3 = 0.794 \cdot f + u_3$$

- ¿Quién es f?

- f es un factor común llamado “general intelligence ability” y donde

$$\Psi_{11} = \text{Var}(u_1) = 0.034$$
$$\Psi_{22} = \text{Var}(u_2) = 0.288$$
$$\Psi_{33} = \text{Var}(u_3) = 0.370$$

Ejemplo 2

- Considerar la matriz de correlaciones

$$R = \begin{pmatrix} 1 & 0.84 & 0.6 \\ & 1 & 0.35 \\ & & 1 \end{pmatrix}$$

- La solución al sistema es

$$\begin{aligned} \lambda_1 &= 1.2 \\ \lambda_2 &= 0.7 \\ \lambda_3 &= 0.5 \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} \Psi_{11} &= \text{Var}(u_1) = -0.44 \\ \Psi_{22} &= \text{Var}(u_2) = 0.51 \\ \Psi_{33} &= \text{Var}(u_3) = 0.75 \end{aligned}$$

- No es admisible y una posible solución es fijar $\lambda_1 = 1$ (caso de Heywood)

Estimación por máxima verosimilitud

- Suponiendo que $\mathbf{x}_1, \ldots, \mathbf{x}_{n+1} \sim N(\mu, \Sigma)$ (iid) donde $\Sigma = \Lambda\Lambda^T + \Psi$ y $\hat{\mu} = \bar{\mathbf{x}}$ podemos pensar en maximizar la log-verosimilitud

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

- O equivalentemente minimizar (Jörekog, 1967)

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

1. Minimizar F analíticamente con respecto a Λ para una Ψ fija
2. Minimizar numéricamente F con respecto a Ψ

Prueba de hipótesis para el número de factores

- Bajo el supuesto de normalidad se pueden hacer pruebas de hipótesis para el número de factores

- El estadístico de prueba estará dado por (Bartlett, 1954)

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

- Si k factores son suficientes entonces,

$$U \sim \chi^2_\nu \qquad \nu = \frac{1}{2}(p - k)^2 - \frac{1}{2}(p + k)$$

- Se puede probar de forma secuencial para $k = 1,2,\ldots$ para encontrar el número de factores a considerar

Ejemplo 3

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

Alumno	Lineal (C)	Estadística (C)	Probabilidad (A)	Finanzas (A)	Cálculo (A)
1	97	92	77	72	96
2	83	88	90	75	96
3	95	83	81	71	96
.	.	.	.	.	.
.	.	.	.	.	.
87	25	36	25	25	35
88	20	50	31	14	29

- La matriz de correlación es

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

Ejemplo 3

- 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

- ¿Cómo se puede interpretar el factor común?

- El factor común se puede interpretar como un factor de habilidad general en matemáticas

Ejemplo 3

- 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

- El primer factor se puede interpretar como un factor de habilidad general en matemáticas
- El segundo factor se puede interpretar como un factor de habilidad en exámenes cerrados y abiertos

Ejemplo 3

- Para $k = 1$

Estimación por factores principales

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

Estimación por 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

Ejemplo 3

- Para $k = 2$

Estimación por factores principales

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

Estimación por 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

Ejemplo 3

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

$$U = n' \min(F) = 7.749803 < 11.0705 = \chi^2_{5,.95}$$

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

- Hacemos la prueba de hipótesis para $k = 2$, obteniendo

$$U = n' \min(F) = 0.02823699 < 3.841459 = \chi^2_{1,.95}$$

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

Rotaciones ortogonales

- Para tener unicidad en el modelo se consideró la restricción $\Lambda^T\Psi^{-1}\Lambda$ sea diagonal

- Para la interpretación de los factores es preferible que:
  1. Cada variable esté asociada fuertemente a lo más a un factor
  2. Las cargas sean muy grandes y positivas o cercanas al cero con algunos valores intermedios

- Ejemplo de rotaciones ortogonales: varimax, quartimax, orthomax, etc.

Rotación varimax

- Rotación ortogonal propuesta por Kaiser (1958)

- Sea Λ la matriz de cargas y $\mathbf{G}$ matriz ortogonal entonces la matriz de cargas rotadas es:

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

- Objetivo: maximizar ϕ

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

Ejemplo 3 (continuación)

- Para $k = 2$

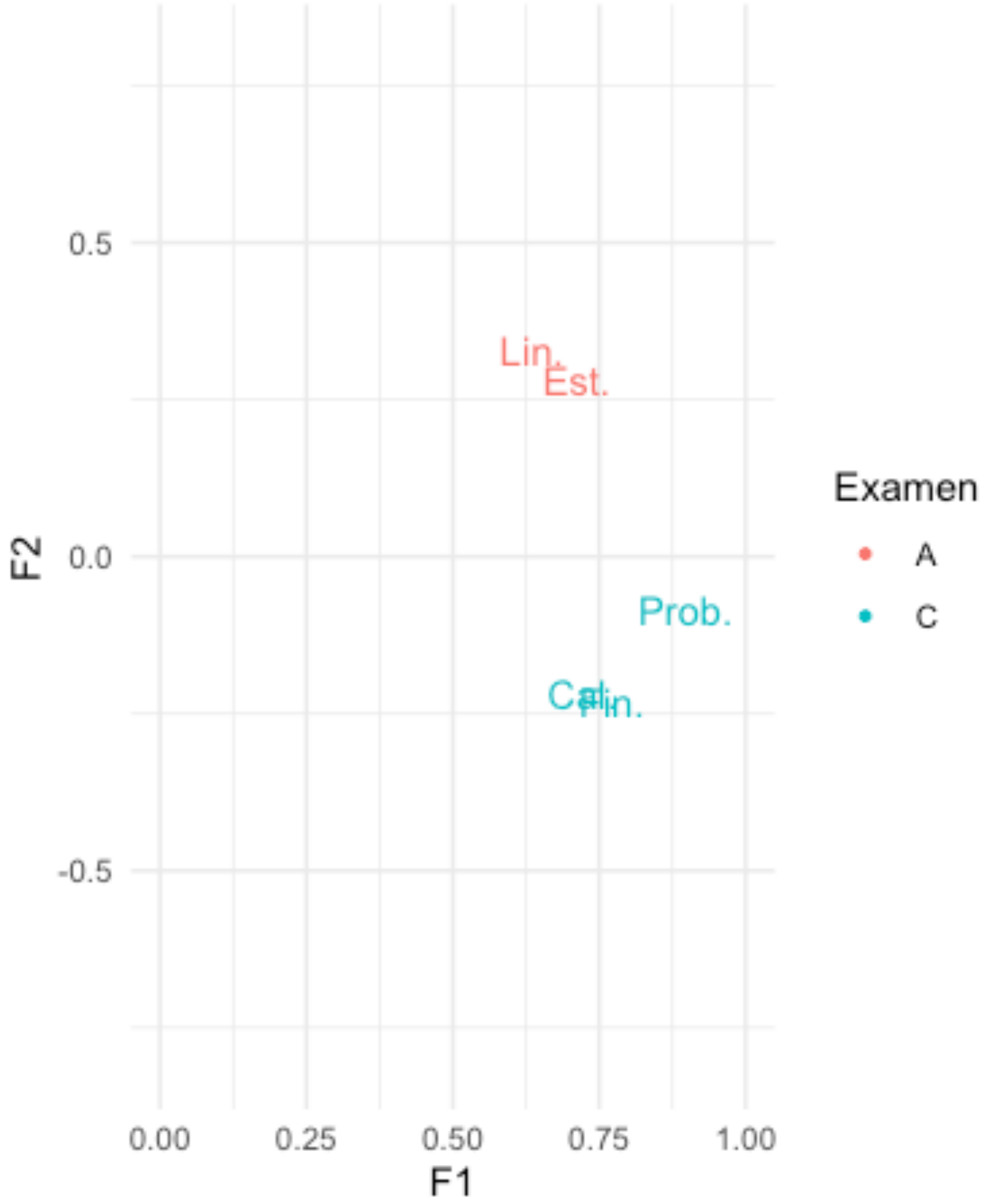
Factores principales sin rotación

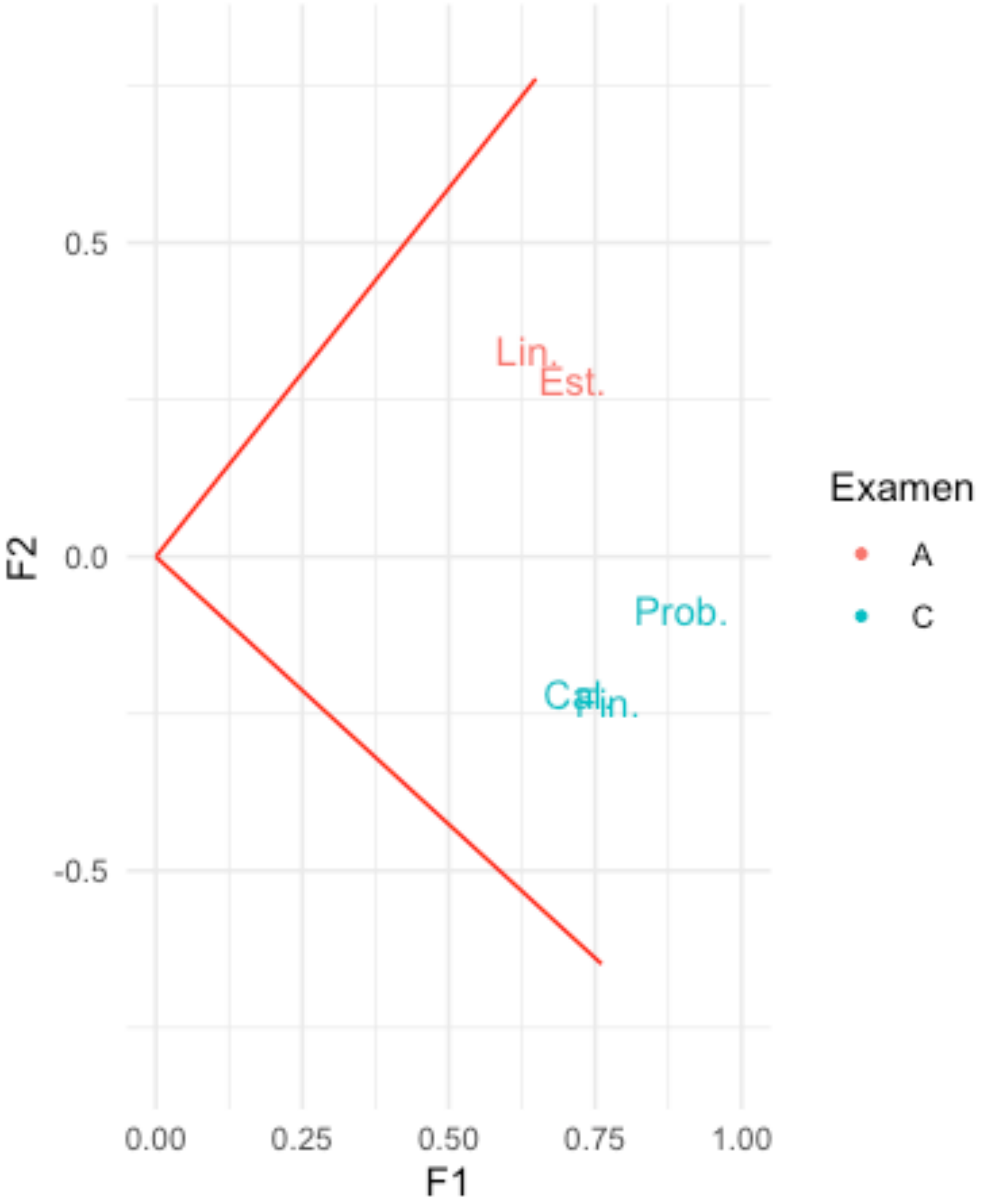
Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.64	0.33	0.51	0.49
Estadística	0.71	0.28	0.59	0.41
Probabilidad	0.90	-0.08	0.81	0.19
Finanzas	0.77	-0.23	0.65	0.35
Cálculo	0.72	-0.22	0.57	0.43

Factores principales con varimax

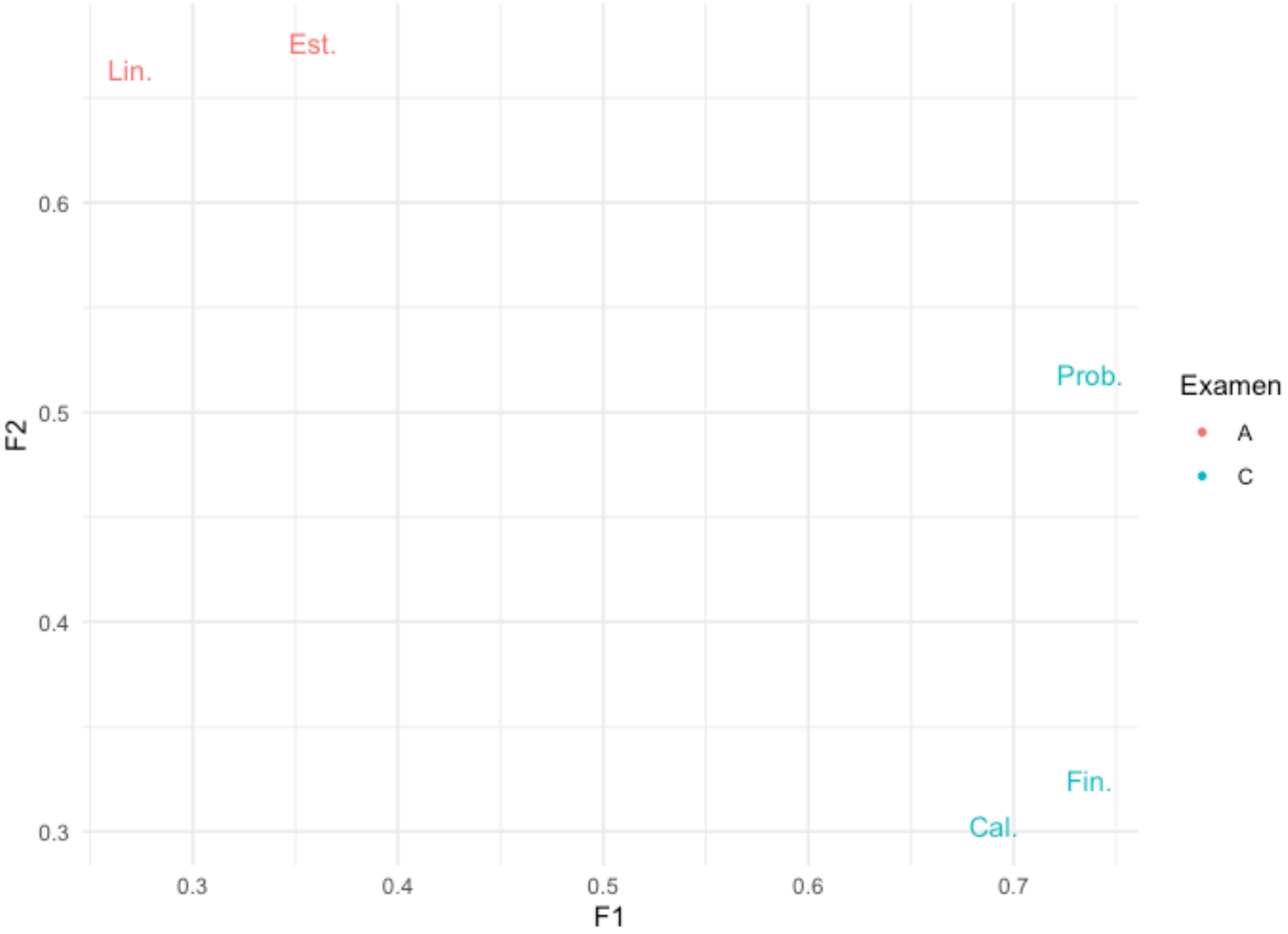
Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	0.27	0.66	0.51	0.49
Estadística	0.36	0.68	0.59	0.41
Probabilidad	0.74	0.52	0.81	0.19
Finanzas	0.74	0.32	0.65	0.35
Cálculo	0.69	0.30	0.57	0.43

Ejemplo 3 (continuación)





Ejemplo 3 (continuación)



Rotaciones oblicuas

- Permiten que los factores estén correlacionados
- Pueden proporcionar soluciones más simples
- Para factores latentes puede ser complicado tener que explicar la correlación entre ellos
- Ejemplos: quartimin, covarimin, biquartimin, oblimin (por default en R), etc.

Ejemplo 3 (continuación)

- Para $k = 2$

Factores principales con 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

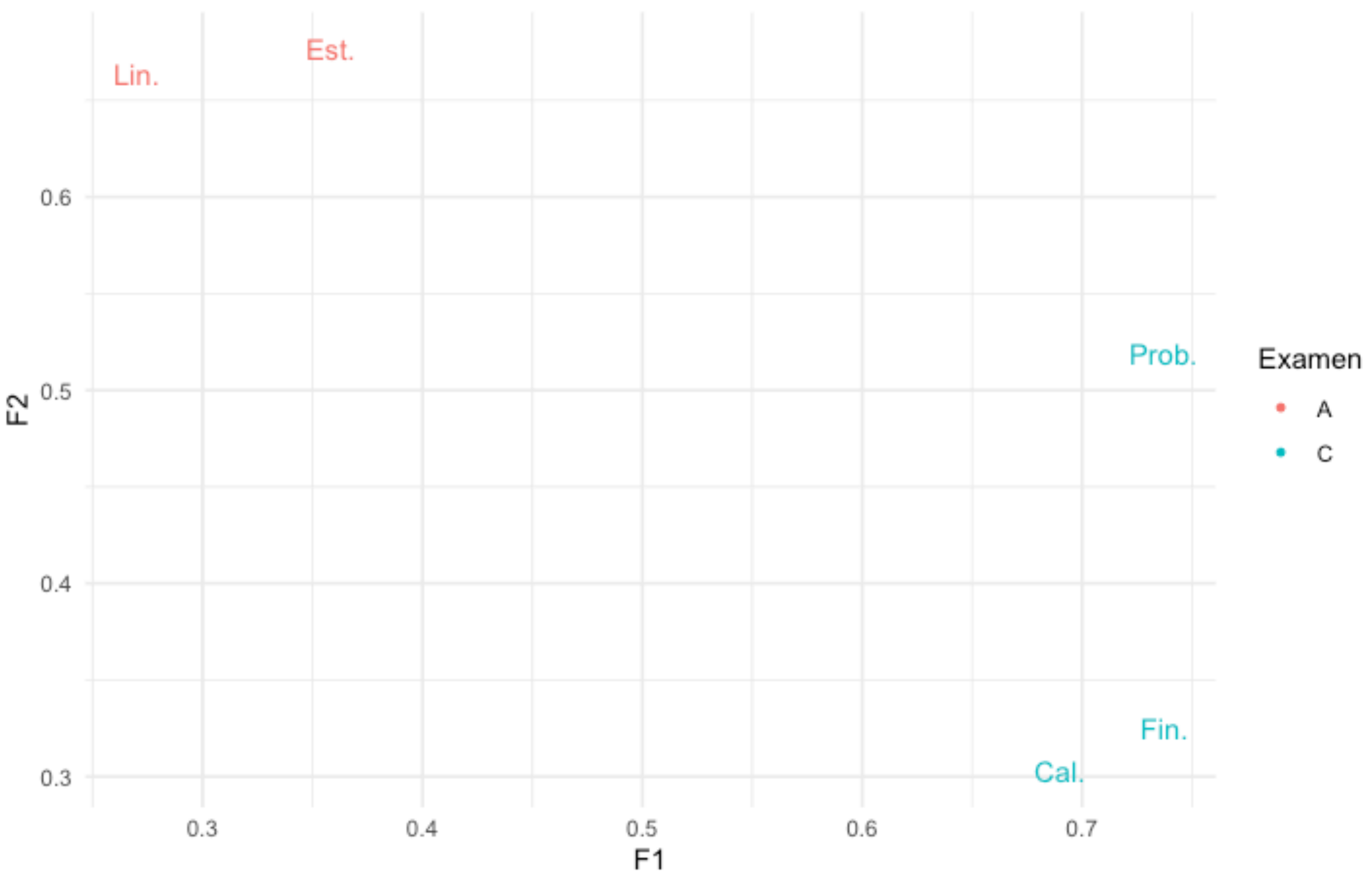
Factores principales con oblimin

Variable	λ_{1i}	λ_{2i}	h_i^2	Ψ_{ii}
Lineal	-0.03	0.74	0.51	0.49
Estadística	0.09	0.70	0.59	0.41
Probabilidad	0.72	0.22	0.81	0.19
Finanzas	0.85	-0.05	0.65	0.35
Cálculo	0.79	-0.05	0.57	0.43

- La correlación de los factores es 0.76

Ejemplo 3 (continuación)

Factores principales con varimax



Factores principales con oblimin