Análise Fatorial

EFT

28 de outubro de 2014

índice

Introdução

A Análise Fatorial (AF) tem por objetivo explicar um conjunto de variáveis observadas por meio de um pequeno conjunto de variáveis latentes ou não observadas, que são chamados de Fatores.

Suponha que tomamos 20 medidas físicas do corpo de um indivíduo: altura, tamanho das extremidades, largura dos ombros, peso, etc.

Supõe-se que todas as medidas não são independentes entre si. Uma explicação é que as dimensões do corpo humano dependem de fatores que caso sejam conhecidos, poderiamos prever as dimensões com erro pequeno.

A AF está relacionado com ACP, mas existem algumas diferenças: 1) as CP se constroem para explicar as variâncias, no entanto, os Fatores se constroem para explicar as covariâncias ou correlações entre variáveis. 2) a CP é uma ferramenta descritiva, enquanto a AF presupõe um modelo estatístico formal de geração dos dados.

Hipóteses

Supomos que observamos um vetor de variáveis **x**, de dimensões $(p \times 1)$ em elementos de uma população. O modelo de AF estabelece que este vetor de dados é gerado mediante a relação:

$$\mathbf{x} = \mu + \Delta\ \mathbf{f} + \mathbf{u} \tag{1}$$

Hipóteses

onde:

1. **f** $(m \times 1)$ é um vetor de variáveis latentes ou fatores não observados. Suporemos que **f**$\sim N_m(\mathbf{0}, \mathbf{I})$. I.e, os fatores são v. com média zero e independentes entre si e com distribuição normal.
2. Δ $(p \times m)$ é uma matriz de constantes desconhecidas $(m < p)$. Contém os coeficientes que descrevem como os fatores **f** afetam as variáveis observadas **x**. É denotada também de matriz de cargas.
3. **u** $(p \times 1)$ é um vetor de perturbações não observadas. Acumula o efeito de todas as variáveis que não são os fatores que exercem influência sobre **x**. Suporemos que **u**$\sim N_p(\mathbf{0}, \psi)$, onde ψ é diagonal e que as perturbações são não correlacionadas com os fatores **f**.

Hipóteses

deduzimos então que:

- μ é a média de **x**
- **x**$\sim N_p(\mu, \mathbf{V})$ (**V** é a matriz de covariâncias.

Isto implica que dada uma a.a.s. de n elementos gerada pelo modelo fatorial, cada x_{ij} pode ser escrito como :

$$x_{ij} = \mu_j + \lambda_{j1} f_{1i} + \ldots + \lambda_{jm} fmi + u_{ij}, i = 1, ..., n, j = 1, ..., p$$