

Modelos de Variáveis Latentes

Dr. Jorge Luis Bazán

`jlbazan@icmc.usp.br`

Universidade de São Paulo

`https://jorgeluisbazan.weebly.com`

CARACTERIZAÇÃO DOS MODELOS DE VARIÁVEIS LATENTES

- 1 Modelos de variáveis latentes
 - Introdução
 - Classificação

CARACTERIZAÇÃO DOS MODELOS DE VARIÁVEIS LATENTES

1 Modelos de variáveis latentes

- Introdução
- Classificação

2 Definições

- Definições
- Modelos sob estudo

CARACTERIZAÇÃO DOS MODELOS DE VARIÁVEIS LATENTES

- 1 Modelos de variáveis latentes
 - Introdução
 - Classificação
- 2 Definições
 - Definições
 - Modelos sob estudo
- 3 Observações

CARACTERIZAÇÃO DOS MODELOS DE VARIÁVEIS LATENTES

- 1 Modelos de variáveis latentes
 - Introdução
 - Classificação
- 2 Definições
 - Definições
 - Modelos sob estudo
- 3 Observações
- 4 Modelos
 - Análise Fatorial
 - Modelos de traços latentes ou TRI
 - Análise de Classe Latente

1. Modelos de variáveis latentes

- Muitos constructos ou variáveis de interesse em diferentes áreas não podem ser observados diretamente
- Existem muitos exemplos deste tipo de variáveis, por exemplo no campo das ciências sociais, as humanidades, a medicina quando são propostos constructos por trás das respostas dos questionários aplicados.
- Exemplos nas psicologia são habilidades cognitivas como atenção, memória, lógica, processamento, emoções, afecções, traços de personalidade; atitudes, intenções, percepções, etc
- Exemplos na educação são habilidade matemática, conhecimento do idioma estrangeiro, competências como pensamento crítico, pensamento criativo, comunicação, argumentação, empatia, autoconhecimento, etc.

- Tais constructos ou variáveis podem ser somente medidos indiretamente por meio de indicadores observáveis, como por exemplo, respostas a itens de um teste ou a perguntas de um questionário, respostas numa escala, a escolha de opções numa lista ou inventário, etc.
- Os questionários, testes, escalas, inventários e outros instrumentos são usados principalmente como uma ferramenta simples, que contém perguntas que servem para coletar o máximo de aspectos de diferentes pessoas relacionadas com o constructo de interesse.
- Estes instrumentos, particularmente os chamados instrumentos psicométricos seguem regras para torná-los apropriados como instrumentos de medida

- Vários tipos de técnicas de escalamento tem sido desenvolvidas para derivar a informação de constructos não observáveis de interesse usando os indicadores.
- Uma família importante dessas variáveis não observáveis de interesse é chamada de variáveis latentes.
- Este tipo de variáveis não está restrita a psicologia ou educação, ela está presente em diversas áreas onde o constructo de interesse deseja ser medido porém considerando perguntas ou items de um instrumento.

Uma variável latente pode ser vista como

- Uma variável que não é observável ou não é diretamente medível
- Uma variável que é medida com erro ou somente pode ser medida com erro.
- Uma variável latente pode ser usada para representar a variável “verdadeira” a qual é medida com erro, ou uma variável conceitual, ou um construto que é o resumo de um conceito complexo.

Exemplos de variáveis “verdadeiras” que são medidas com erro: ingestão de cálcio medida na frequência alimentar, atividade física medida por auto-reporte, peso auto-reportado, latência do sono obtida de diários de sono, concentração de bebida alcoólica em uma pessoa

Exemplos de variáveis conceituais e constructos que são desejáveis de serem medidas: liberalismo, qualidade de vida, estresse, autoestima, nível socioeconômico, dieta saudável, habilidade matemática, satisfação do cliente, suporte social, dificuldades da fala, severidade asmática, inteligência, fluência de leitura, etc.

Os modelos de variáveis latentes são

- Um modo de analisar as relações entre variáveis observáveis (manifestas)
- Os modelos de variáveis latentes postulam a existência de variáveis não observáveis ou fatores subjacentes as variáveis manifestas mas que não podem ser observados o medidos diretamente.
- A hipótese fundamental é que a covariação das variáveis manifestas são explicadas pelas covariação entre as variáveis manifestas e as latentes.
- É assumido que as respostas nos indicadores são resultado de uma posição individual na(s) variáveis latente(s).
- O princípio fundamental é a independência condicional: as variáveis manifestas são independentes, condicional as variáveis latentes.

Considerando adicionalmente a classificação das variáveis manifestas e latentes nos podemos ter os seguintes modelos

- Análise Fatorial: variáveis manifestas e latentes contínuas
- Análise de traço latente ou Modelos de Resposta ao item: variáveis manifestas categóricas e variáveis latentes contínuas
- Análise de perfil latente: variáveis manifestas contínuas e variáveis latentes categóricas
- Análise de Classe latente: variáveis manifestas categóricas e variáveis latentes categóricas

Modelos de variáveis latentes

	Variáveis latentes	
Variáveis manifestas	Contínuas	Catóricas
Contínuas	Análise Fatorial (AF)	Análises de Perfis latentes (APF)
Catóricas	Análise de Traço Latente (TRI)	Análise de Classe Latente (ACL)

- Na AF e na TRI, as variáveis latentes são usualmente assumidas como variáveis contínuas normalmente distribuídas.
- Na APL e ACL as variáveis latentes discreta é geralmente assumida como uma distribuição multinomial.
- As variáveis manifestas no AF e APL são contínuas. Em muitos casos, a distribuição condicional dada a variável latente é assumida normal.
- Na TRI e na ACL, os indicadores são variáveis dicotômicas ou policotômicas: ordinais ou nominais, e as distribuições condicionais são assumidas a ser binomiais ou multinomial.

Atualmente, a distinção entre os modelos de indicadores contínuos e discretos acaba por não ser fundamental em tudo. A especificação das distribuições condicionais dos indicadores segue-se naturalmente a de seus tipos de escala.

O mais recente desenvolvimento na modelagem de variáveis latentes é para permitir distribuições diferentes para cada indicador. Estes podem, por exemplo, seguir as distribuições normal, t-student, log-normal, gama, ou exponencial para as variáveis contínuas, binomial para variáveis dicotômicas, multinomial para variáveis ordinais e nominais, e Poisson, binomial, ou negativa-binomial para contagens.

Dependendo se a variável latente é tratada como contínua ou discreta, obtém-se uma forma generalizada de TRI ou ACL.

2. Definição de variável latente

Seja (\mathbf{X}, \mathbf{U}) um modelo latente com $\mathbf{X} = (X_1, \dots, X_I)$ e \mathbf{U} vetores aleatórios de valor real não necessariamente da mesma dimensão.

\mathbf{X} é um vetor de variáveis manifestas e \mathbf{U} é um vetor de variáveis latentes.

Assuma-se que as distribuições de \mathbf{X} , \mathbf{U} e (\mathbf{X}, \mathbf{U}) existem.

(\mathbf{X}, \mathbf{U}) modelo latente satisfaz:

- *Independência condicional latente sse*

$$F(\mathbf{x}|\mathbf{u}) = \prod_1^I F_i(x_i|\mathbf{u}) \quad \forall i \in \{1, \dots, I\}, \forall \mathbf{u}$$

Nesse caso dizemos que o modelo é um modelo latente condicionalmente independente.

- *A monotonicidade latente sse*

$$1 - F_i(x|\mathbf{u}) = P(X_i > x|\mathbf{u})$$

são funções não decrescentes de \mathbf{u} , $\forall x$ e $\forall i \in \{1, \dots, I\}$

- *A unidimensionalidade latente sse \mathbf{u} é unidimensional (é variável aleatória)*

(\mathbf{X}, \mathbf{U}) é um *modelo de variável latente monótona* sse satisfaz a independência condicional e a monotonicidade latente.

Se além disso o modelo é unidimensional se diz que é um modelo de *variável latente monótono unidimensional*.

- *Teoría da Resposta ao Item.*

$\{1, \dots, I\}$ são ítems. X_i são respostas binárias (1=correto, 0=incorreto). $P(X_i = 1|U = u)$ representa a probabilidade de responder corretamente o item dada U , habilidade latente unidimensional do sujeito.

- *Modelos de Classe Latente.*

$\{1, \dots, I\}$ são sintomas. X_i são respostas binárias (1=têm sintoma, 0=não têm sintoma). U classe latente (variável nominal discreta) que toma valores num conjunto ordenado finito S .

- *Análise Fatorial.*

$\{1, \dots, I\}$ são variáveis manifestas contínuas. X_i são normais i.i.d. \mathbf{U} é um vetor normal multivariado de fatores latentes. X_i e \mathbf{U} cumprem

$$X = \Lambda \mathbf{U} + \mathbf{e}$$

com Λ matriz fixa de cargas fatoriais e \mathbf{e} vetor de erros normais i.i.d. que são também independentes de \mathbf{U} .

Quando todas as cargas fatoriais são não negativas se têm um modelo de variável latente monótona. Se em adição U é escalar e $\Lambda \geq 0$ é um vetor se têm um modelo de variável latente monótona unidimensional. Se U é escalar e Λ um vetor de $1s$, então o modelo se reduz à *teoría do puntaje verdadeiro*. De maneira geral, o *modelo unifatorial não linear*

$$X = f(U) + \mathbf{e}$$

com U escalar, \mathbf{e} erros independentes e f uma função não decrescente monótona de valor vetorial, é um modelo de variável latente monótona unidimensional.

3. Observações

- Um de nossos propósitos é usar os modelos estatísticos para medir variáveis latentes associando elas com coisas que são observáveis (isto é, questionários de itens, resultados de testes, ou qualquer ferramenta observável).
- O debate filosófico ao respeito da existência fundamental ou não existência de variáveis latentes conceituais é evitada e uma posição pragmática é considerada. Ao respeito pode ser revisado:
Borsboom, D., Mellenbergh, G. and van Heerden, J. (2003). The Theoretical Status of Latent Variables. *Psychological Review*. Vol. 110, No. 2, 203-219.
- O intuito é obter informação e explicações para as relações dos dados considerando a modelagem estatístico.

- As variáveis latentes são também usadas em diferentes técnicas de modelagem estatístico como uma conveniência matemática em que elas não usualmente o foco do interesse, isto é, o propósito não é “medir”. Isto acontece por exemplo, em dados perdidos, heterogeneidade não observável (por exemplo em dados de sobrevivência, efeitos aleatórios, dados longitudinais, ou dados em cluster).
- Quando o propósito é medir, a motivação pode ser para pesquisa, para avaliação e para decisões de política.

Análise Fatorial Exploratória (AFE) foi inicialmente proposta pelo Spearman em 1904. O propósito geral é

- Determinar quantos fatores subjacentes são necessários para explicar a maior correlação e variância dos dados.
- Determinar a relação via rotação entre cada um dos fatores subjacentes com cada uma das variáveis observadas de maneira significativa de modo que possam ser interpretadas e denominadas.
- Para eliminar variáveis observadas que não tendem a medir bem os fatores subjacentes comuns dadas por outras variáveis.
- Propor blocos de variáveis que podem ser posteriormente usados para criar um simples escala baseado em escores.
- Para propor um modelo de Análise Fatorial Confirmatório AFC

Análise Fatorial Confirmatória tem o propósito de

- Para definir um modelo de medição para a relação entre observações multivariadas e factores subjacentes
- Para testar a significância estatística das cargas fatoriais e correlações ou para testar se as cargas fatoriais rotadas obtidos de um AFE que são assumidas como zero são, de fato, significativamente diferente de zero ou não.
- Para testar se o modelo de medição para um grupo é o mesmo que o modelo de medição para algum outro grupo.
- Como um precursor para um modelo de equações estruturais

Métodos originalmente propostos na avaliação educacional (as variáveis latentes são chamadas de traços latentes), existe uma grande literatura relacionada com a TRI.

Respostas (0, 1) a uma serie de k questões, assim são possíveis 2^k possíveis patroses de resposta (dados dicotômicos). Respostas $(1, 2, \dots, c)$ para uma serie de k questões, assim existem c^k possíveis padrões de resposta (dados policotômico).

Questões a serem respondidas:

- Quantas das diferenças entre as respostas podem ser explicadas supondo que os itens podem ser explicados por uma o mais variáveis latentes continuas?
- Quantas variáveis subjacentes existem?
- Quantas variáveis ajudam a discriminar os indivíduos melhor?
- Qual é a melhor maneira de combinar as variáveis observadas com o propósito de criar uma escala o score para cada individuo?

O crédito é usualmente dado para o Paul Lazarsfeld:

Lazarsfeld, P.F. e Henry, N. W. (1968) Latent Structure Analysis. Houghton Mifflin.

ACL é um método estatístico para encontrar subtipos de casos relacionados considerando dados categóricos multivariantes.

Questões a serem respondidas:

- Quantas classes subjacentes existem?
- Qual é a prevalência em cada uma das classes latentes?
- Qual é a relação entre as respostas observadas e as classes latentes?
- Qual é a probabilidade que um particular indivíduo pode ser uma classe particular?

Uma revisão histórica e uma visão geral dos modelos de variáveis latentes pode ser vista em:

Bartholomew, D. (2002). Old and new approaches to latent variable modeling. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure*. (pag 1-14) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.