

# Análise fatorial exploratória

Prof. Dr. Nelson Hauck Filho

Salvador, 27 de junho de 2019.

# Overview.

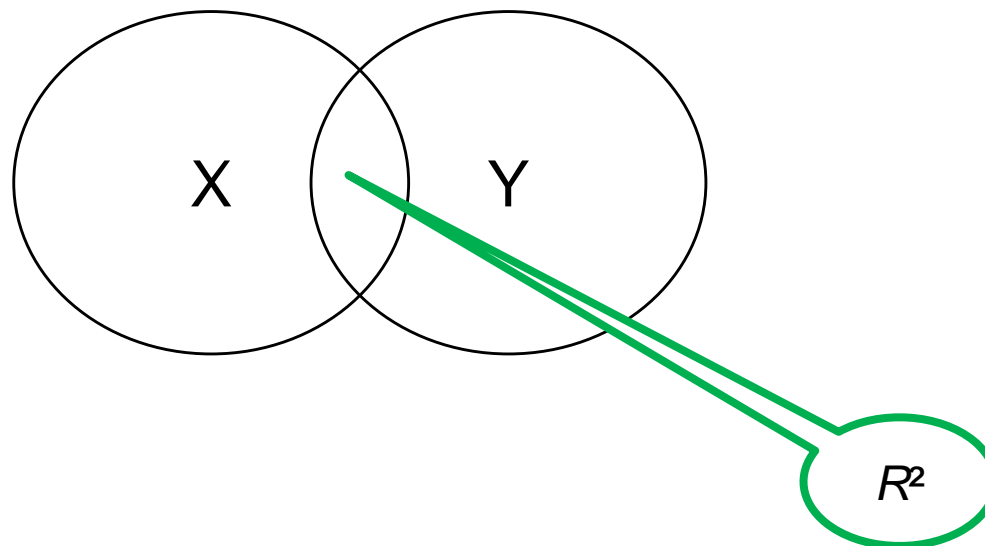
- Apresentação breve de conceitos relacionados à análise fatorial
- Introdução à análise fatorial exploratória com o pacote psych e com o programa jamovi
- Exercícios

# Conceitos básicos

- Regressão linear
- Variáveis latentes
- Modelo reflexivo
- Dimensionalidade
- Independência local
- Análise exploratória versus confirmatória

# Regressão linear

- Utilizada para estimar a força de associação linear entre duas variáveis
- $X$  = variável independente, preditor
- $Y$  = variável dependente, desfecho



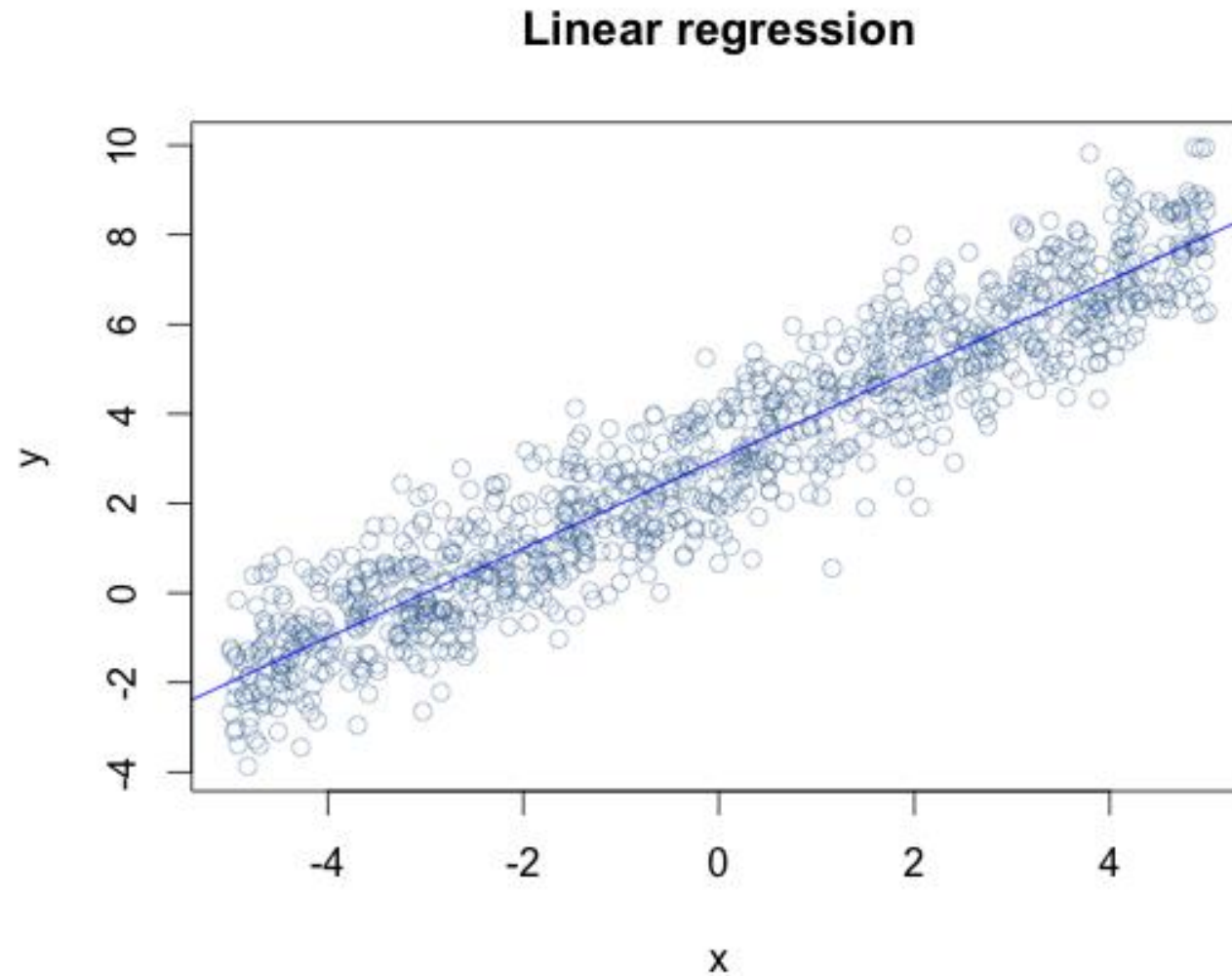
# Regressão linear simples.

$$Y = B_0 + B_x X + E$$

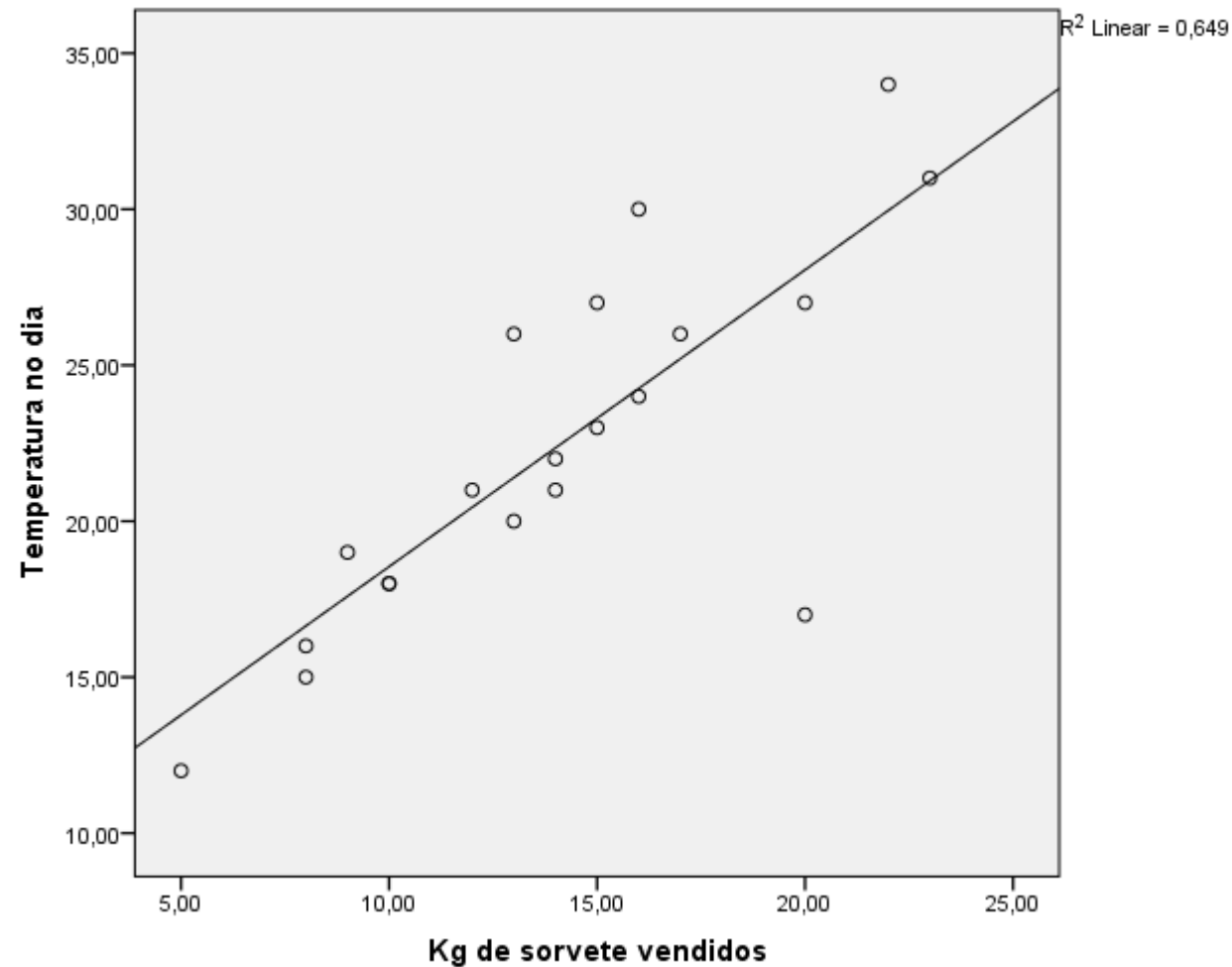
Em que:

- $Y$  = variável dependente.
- $B_0$  = intercepto.
- $B_x$  = coeficiente angular associado à variável  $X$ .
- $X$  = variável independente.
- $E$  = erro aleatório.

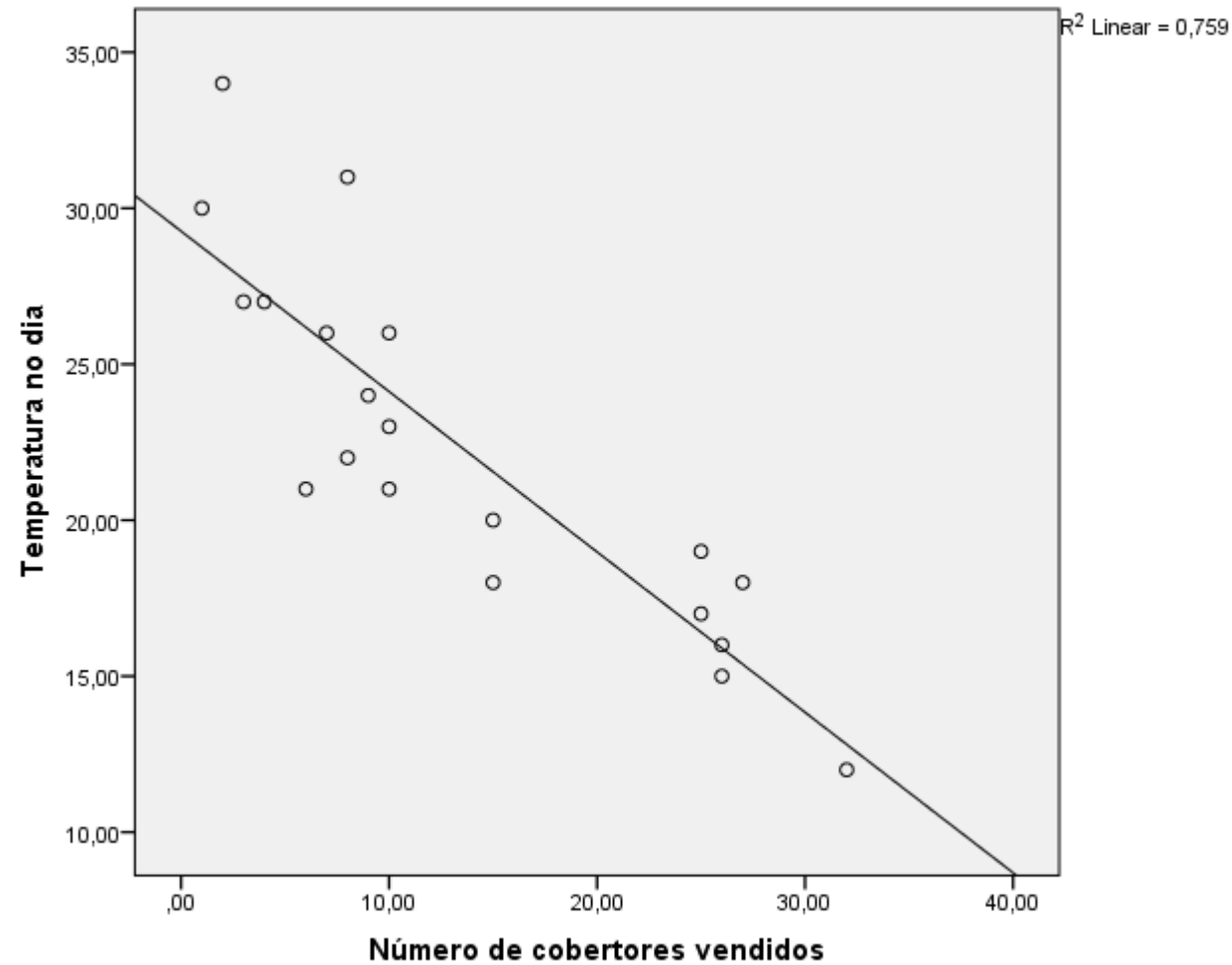
# Regressão linear simples.



# Temperatura e venda de sorvetes



# Temperatura e venda de cobertores





# Regressão linear multivariada.

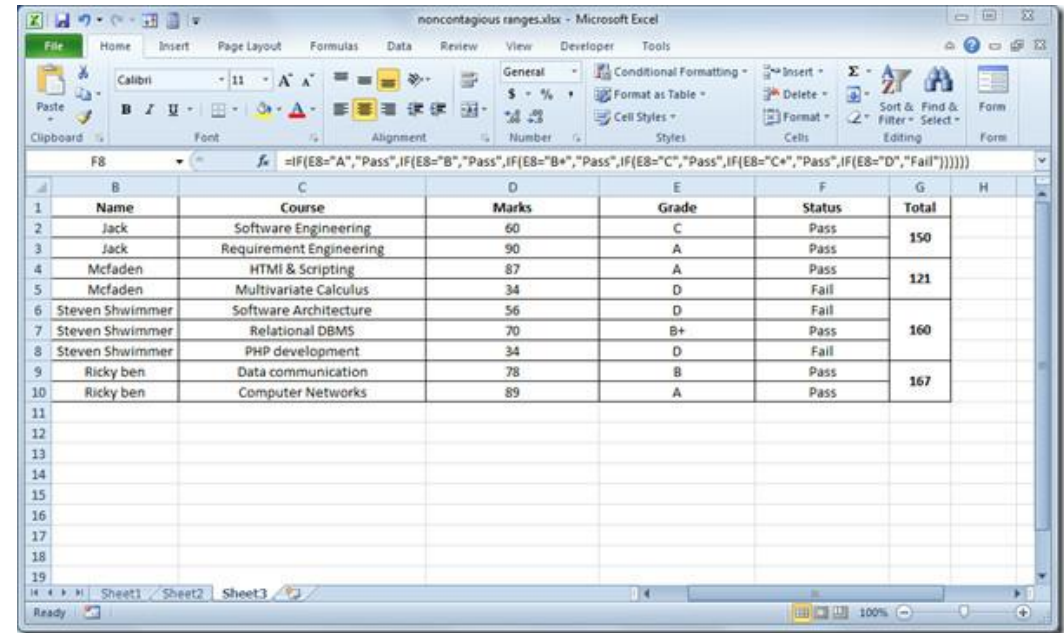
$$Y = B_0 + B_{x1}X_1 + B_{x2}X_2 + \dots + B_nX_n + E$$

- $Y$  = variável dependente.
- $B_0$  = intercepto.
- $B_{x1}$  = coeficiente angular associado à variável  $X_1$ .
- $B_{x2}$  = coeficiente angular associado à variável  $X_2$ .
- $B_n$  = coeficiente angular associado à variável  $X_n$ .
- $E$  = erro aleatório.

# Fatores em uma análise fatorial

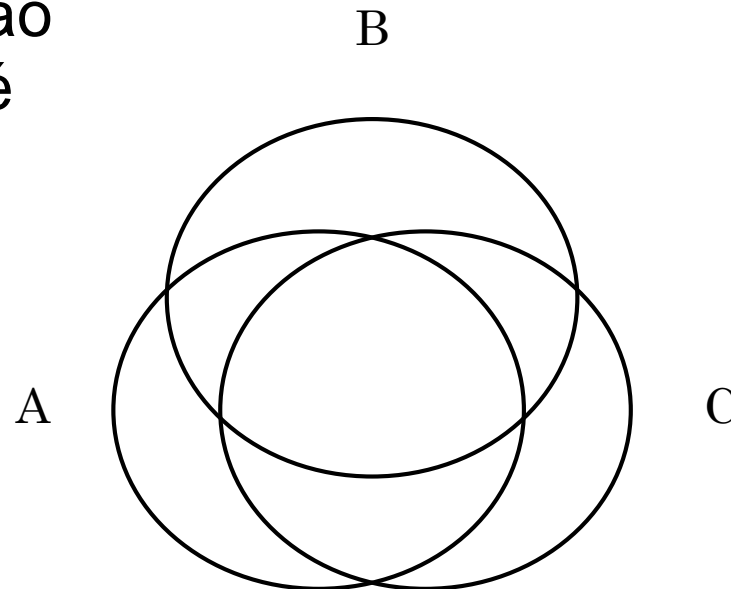
- Variáveis preditoras, independentes,  $X$
- Não observadas, não estão disponíveis no banco de dados
- Explicam variáveis observadas (dependentes,  $Y$ )

- Vamos pensar em um banco de dados no Excel. Linhas são casos (pessoas), colunas são variáveis. Vamos imaginar que são 200 pessoas e 10 variáveis. Uma variável latente é uma variável que explica os escores em uma ou mais das colunas do banco de dados, mas que não está nesse banco, que não é nenhuma das 10 colunas.



The screenshot shows a Microsoft Excel spreadsheet titled 'noncontagious ranges.xlsx'. The spreadsheet has columns labeled B through H. Column B is 'Name', C is 'Course', D is 'Marks', E is 'Grade', F is 'Status', and G is 'Total'. The data is as follows:

B	C	D	E	F	G	H
Name	Course	Marks	Grade	Status	Total	
Jack	Software Engineering	60	C	Pass	150	
Jack	Requirement Engineering	90	A	Pass		
Mcfaden	HTML & Scripting	87	A	Pass	121	
Mcfaden	Multivariate Calculus	34	D	Fail		
Steven Shwimmer	Software Architecture	56	D	Fail	160	
Steven Shwimmer	Relational DBMS	70	B+	Pass		
Steven Shwimmer	PHP development	34	D	Fail		
Ricky ben	Data communication	78	B	Pass	167	
Ricky ben	Computer Networks	89	A	Pass		

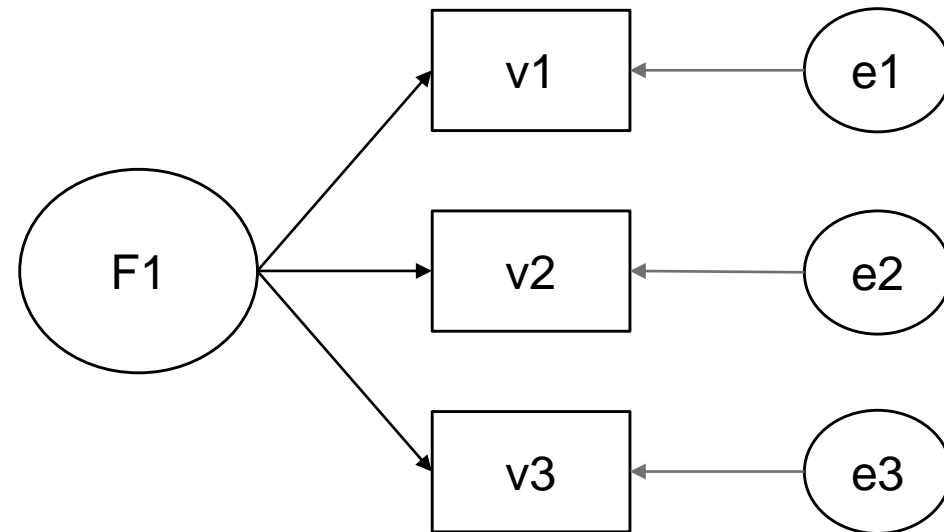


	Item A	Item B	Item C	Total
Joãozinho	1	1	1	3
Mariazinha	0	1	1	2
Zezinho	0	0	1	1
Total	1	2	3	-

- A maioria das variáveis relacionadas ao funcionamento psicológico são latentes ou, ao menos, envolvem processos latentes
  - Cognição
  - Motivação
  - Emoção
  - Personalidade
  - Psicopatologia
  - Aprendizagem

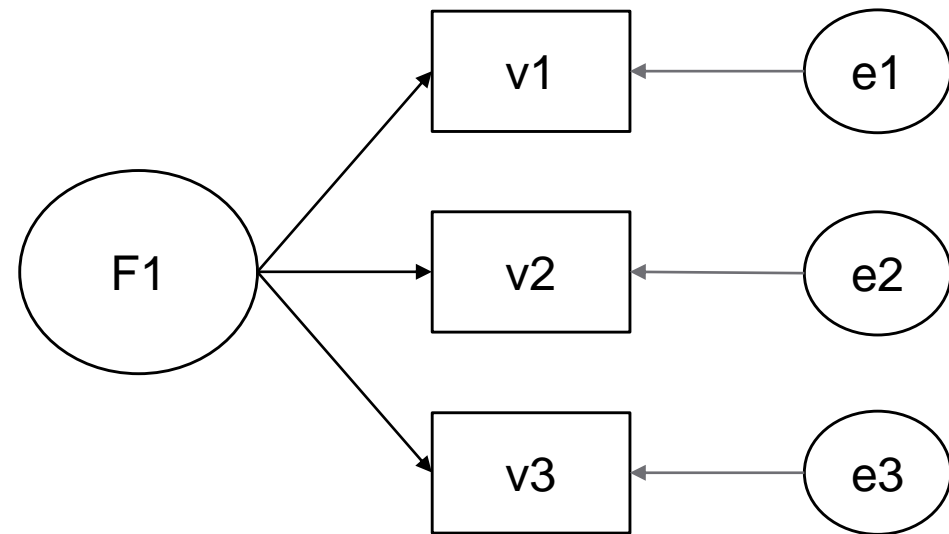
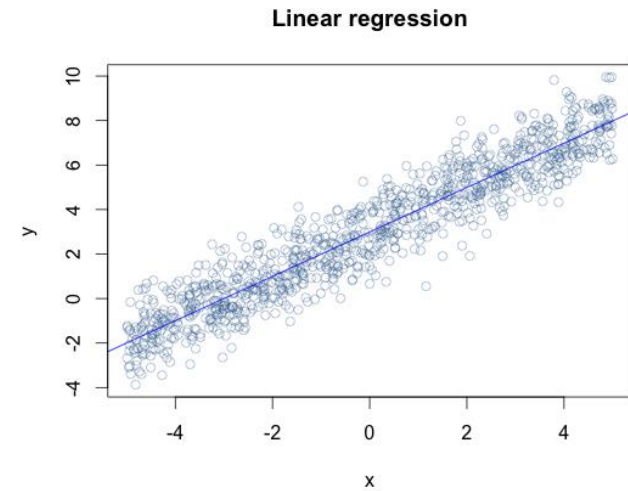
# Modelo reflexivo

- O modelo em que a variável latente é a causa para a variação nos indicadores
- Elipses: variáveis latentes
- Retângulos: variáveis observadas
- A técnica dos componentes principais não segue esse modelo



$$Y1 = v + \lambda_1 F1_1 + e1$$

- $Y1$  = resposta a item 1.
- $v$  = intercepto.
- $\lambda_1$  = carga do item  $Y1$  no fator  $F1$ .
- $F1$  = fator (variável latente).
- $e$  = erro aleatório.

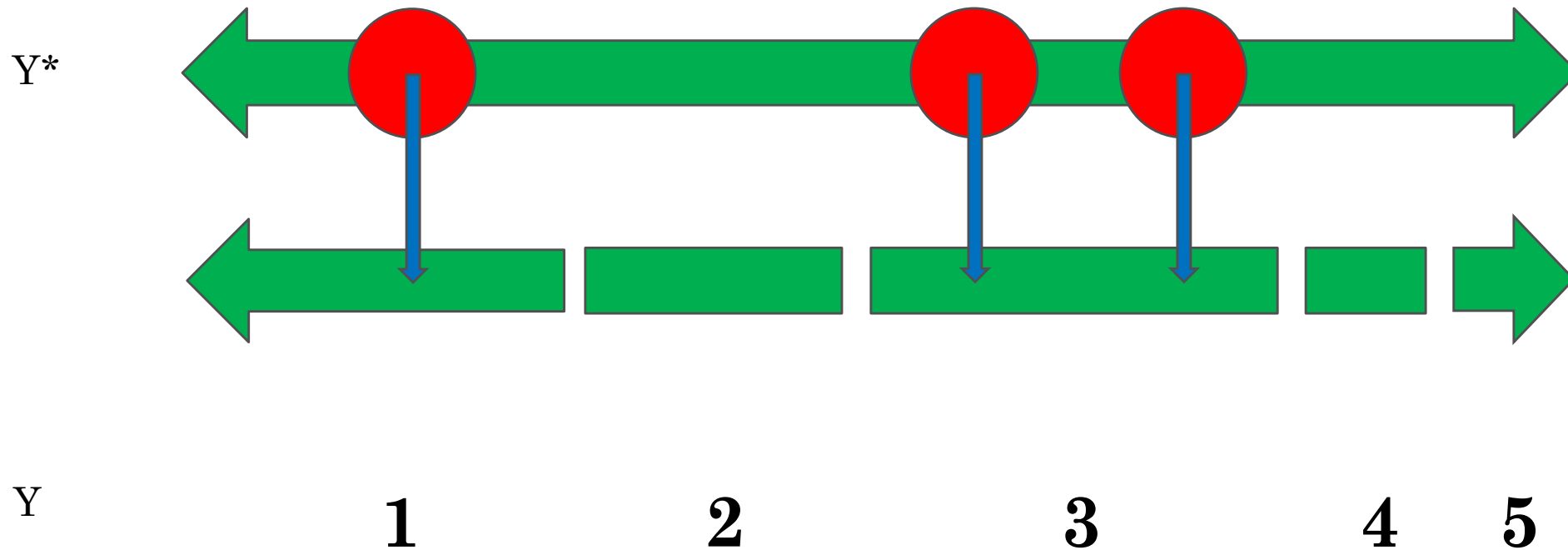


# Tipos de indicadores.

- Categórica=ocorrência ou não de um desfecho, tipos ou categorias.
- Categórica ordenada=escala Likert.
- Contínua=tempo de reação, nível de cortisol, idade, tempo de exposição diário ao sol.
- Contagem=número de doses de álcool consumido a cada episódio de ingesta, número de vezes em que algo aconteceu.
- Variável censurada=uma das variáveis anteriores em que ocorreu de haver um limite numérico no instrumento de medida.

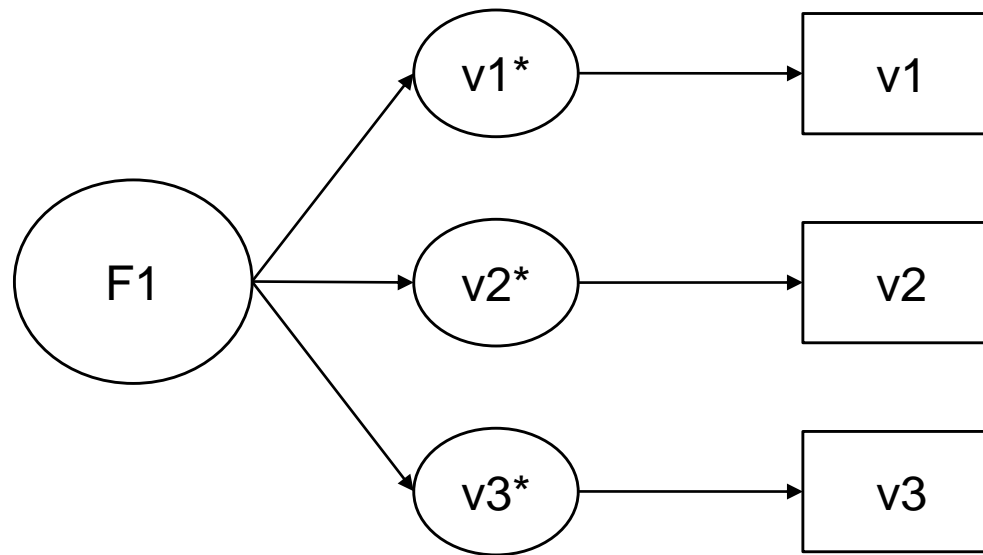


# Resposta contínua e escolha de categorias em escala Likert.



# Resposta contínua e escolha de categorias em escala Likert.

- Apenas um modelo teórico, não deve ser especificado desta forma!



## Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables

Francisco Pablo Holgado –Tello · Salvador Chacón –Moscoso ·  
Isabel Barbero –García · Enrique Vila –Abad

Published online: 7 September 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Given that the use of Likert scales is increasingly common in the field of social research it is necessary to determine which methodology is the most suitable for analysing the data obtained; although, given the categorization of these scales, the results should be treated as ordinal data it is often the case that they are analysed using techniques designed for cardinal measures. One of the most widely used techniques for studying the construct validity of data is factor analysis, whether exploratory or confirmatory, and this method uses correlation matrices (generally Pearson) to obtain factor solutions. In this context, and by means of simulation studies, we aim to illustrate the advantages of using polychoric rather than Pearson correlations, taking into account that the latter require quantitative variables measured in intervals, and that the relationship between these variables has to be monotonic. The results show that the solutions obtained using polychoric correlations provide a more accurate reproduction of the measurement model used to generate the data.

**Keywords** Construct validity · Polychoric correlations · Pearson correlation · Factor analysis

## Developing Multidimensional Likert Scales Using Item Factor Analysis: The Case of Four-point Items

Rodrigo A. Asún<sup>1</sup>, Karina Rdz-Navarro<sup>1</sup>,  
and Jesús M. Alvarado<sup>2</sup>

### Abstract

This study compares the performance of two approaches in analysing four-point Likert rating scales with a factorial model: the classical factor analysis (FA) and the item factor analysis (IFA). For FA, maximum likelihood and weighted least squares estimations using Pearson correlation matrices among items are compared. For IFA, diagonally weighted least squares and unweighted least squares estimations using items polychoric correlation matrices are compared. Two hundred and ten conditions were simulated in a Monte Carlo study considering: one to three factor structures (either, independent and correlated in two levels), medium or low quality of items, three different levels of item asymmetry and five sample sizes. Results showed that IFA procedures achieve equivalent and accurate parameter estimates; in contrast, FA procedures yielded biased parameter estimates. Therefore, we do not recommend classical FA under the conditions considered. Minimum requirements for achieving accurate results using IFA procedures are discussed.

# Dimensionalidade

- Número de variáveis latentes responsáveis pela variabilidade nos dados do pesquisador
- A quantidade de causas ocultas para as respostas aos instrumentos
- Aspecto crucial para a análise factorial exploratória

# Dimensionalidade

- Autovalor  $> 1$
- Análise paralela
- Very Simple Structure
- BIC
- Hull

# Estimador

- O método específico que vai permitir estimar valores para os parâmetros do modelo
- Mais confiável/consistente
  - ML
- Mais indicados para escalas Likert
  - ULS
  - WLSMV

# Interpretação fatorial

- Identificar as maiores cargas em cada fator
- Verificar o sentido do valor da carga (+ ou -)
  - Sugestão: não inverter os itens na base de dados original
- Manter itens com carga  $> 0,30$  ou algum valor superior a esse

# Apresentação

- Desenvolvido por William Revelle
- Autor possui uma página virtual sobre Psicometria, personalidade e R: <https://www.personality-project.org/>
- Pacote psych faz todas as análises psicométricas e estatísticas básicas
- A página do autor contém diversos tutoriais para usar as diferentes funções do programa



# Exercícios

- Baixar os arquivos
  - banco1.txt
  - Labels banco 1.txt
  - Script psych.R