Análise Fatorial Exploratória: Conceitos Básicos

Prof. Dr. Wagner de Lara Machado

PPG Psicologia PUCRS

Grupo de Pesquisa Avaliação em Bem-estar e Saúde Mental

Conteúdo

- Aspectos gerais
- Métodos de extração de fatores
- Métodos de retenção de fatores
- Métodos de rotação de fatores
- Avaliação do modelo fatorial

Análise de itens e pressupostos da análise fatorial exploratória (AFE)

A AFE é uma das técnicas mais utilizadas na psicologia, especialmente na **psicometria**, com a finalidade de testar o desempenho de instrumentos psicométricos ou **gerar teorias**

(Costello & Osborne, 2005; Haig, 2005)

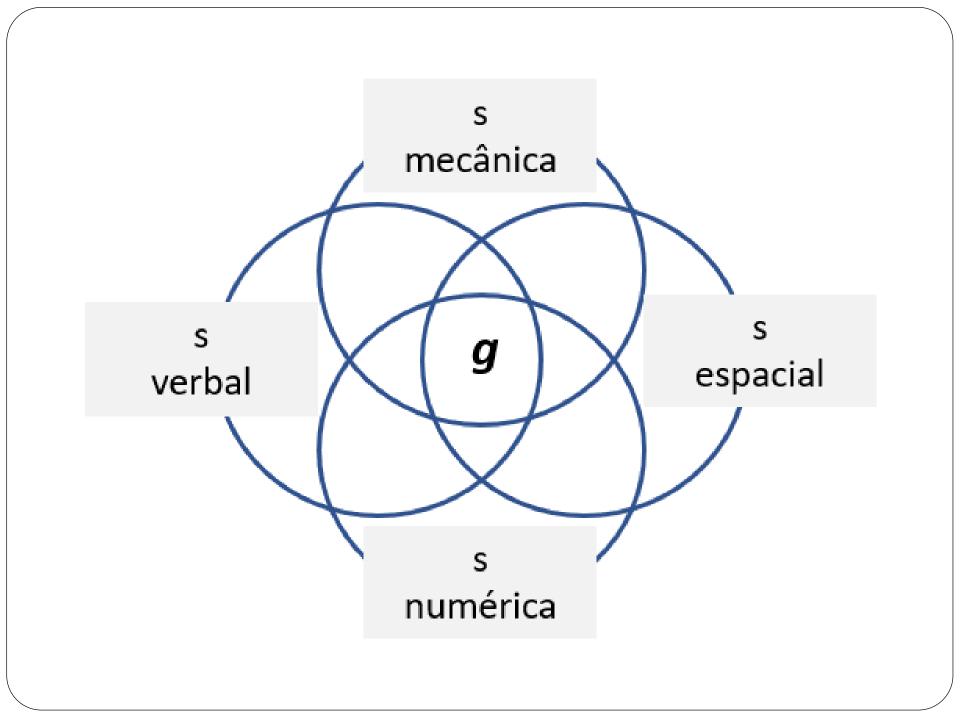
MULTIVARIATE BEHAVIORAL RESEARCH, 40(3), 303–329 Copyright © 2005, Lawrence Erlbaum Associates, Inc.

Exploratory Factor Analysis, Theory Generation, and Scientific Method

Brian D. Haig University of Canterbury

Charles Spearman

- Foi desenvolvida por Spearman e utilizada na geração da teoria do fator geral de inteligência (g)
- Sua hipótese era de que havia uma variável (geral) que explicava as relações entre habilidades específicas



Cinco grandes fatores da personalidade (Big Five)

Por Que Cinco Fatores?

A descoberta dos cinco fatores foi acidental e se constitui em uma generalização empírica, replicada independentemente inúmeras vezes. Como o modelo não foi desenvolvido a partir de uma teoria, não há, consequentemente, uma explicação teórica *a priori* (e satisfatória) dos motivos que levariam a organização da personalidade em cinco (e não quatro, ou sete) dimensões básicas.

1998 Claudio Simon Hutz / Carlos H. Nunes / Alice D. Silveira / Jovana Serra / Márcia Anton / Luciane S. Wieczorek O DESENVOLVIMENTO DE MARCADORES PARA A AVALIAÇÃO DA PERSONALIDADE NO MODELO DOS CINCO GRANDES FATORES



Psicometria

• Na Psicometria, o modelo de AFE é empregado para descobrir (!) o número de variáveis (latentes) necessárias e suficientes para explicar um conjunto de itens ou tarefas de um instrumento psicométrico. Um uso bem comum é na investigação das propriedades psicométricas de escalas, questionários e inventários

Formulação matemática

- $X = b + a_1F_1 + a_2F_1 + a_3F_1 \dots + e$
- Em que:

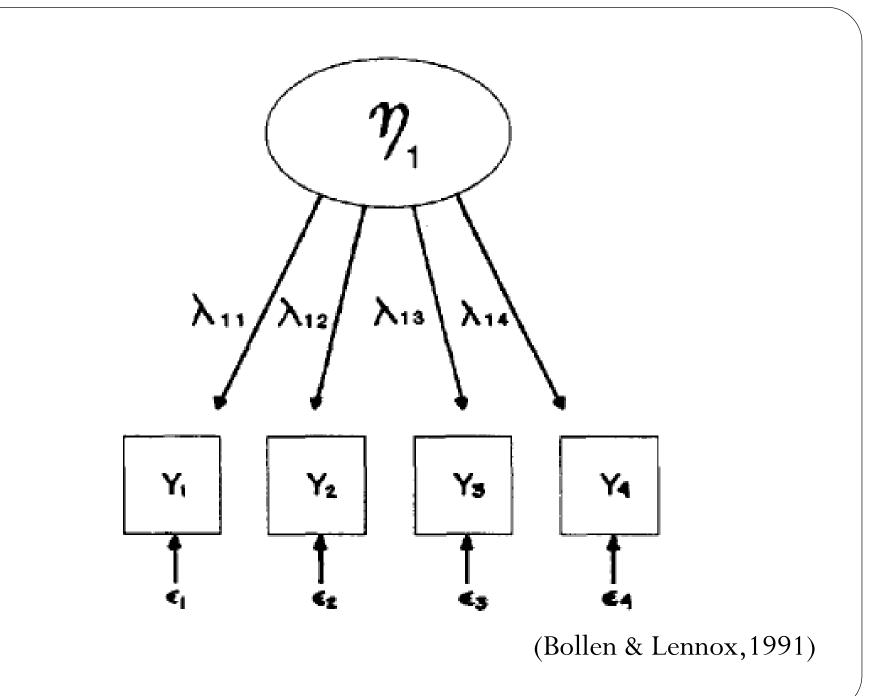
X é o escore observado no teste;

b é um intercepto, em geral fixado em zero;

a é o coeficiente angular, indicando a relação linear entre item e fator;

 $\emph{\textbf{F}}$ é um escore fatorial, valor latente do traço em questão;

e é um erro aleatório com média zero e distribuição normal.



• Fatores comuns Vs. Componentes principais

Componentes principais Var. expecífica + Var. comum + Var. erro

1	r	r
r	1	r
r	r	1

Fatores comuns
Var. comum + Var. erro

h ²	r	r
r	h^2	r
r	r	h ²

• h^2 = Comunalidade, isto é, a variância compartilhada entre a variável e todos os fatores retidos

Uso da análise fatorial exploratória em psicologia

Bruno Figueiredo Damásio¹ – Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

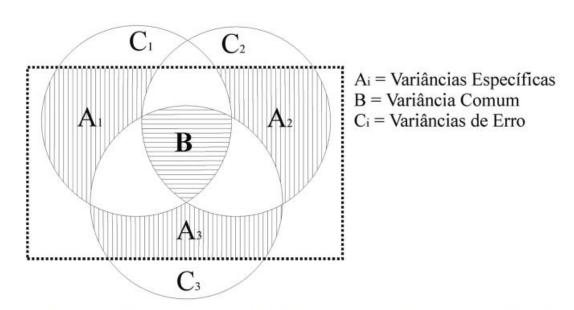


Figura 1 - Ilustração das variâncias de três itens e suas relações com um fator hipotético.

- A AFE é uma técnica que:
- 1) parte da hipótese de que a correlação entre variáveis tem como <u>causa</u> uma ou mais variáveis, também chamada de "princípio da causa comum";
- 2) introduz a modelagem de **variáveis latentes**, isto é, variáveis que não são diretamente observáveis, e, por isto, não estão no banco de dados;
- 3) diferente de outras técnicas estatísticas, a AF é exploratória e interpretativa, alinhada ao raciocínio adbutivo (geração de hipóteses e teorias após a observação empírica)

- Avaliação das variáveis
 - Nível de mensuração
 - A análise de fatores é uma técnica para variáveis métricas, ou seja, intervalares (Hair, Anderson, Tatham, & Black, 2005)
 - Contudo, existem métodos robustos para acomodar variáveis que não cumprem esse pressuposto
 - Variáveis intervalares: matrizes de variância e covariância, matrizes de correlações de *Pearson*
 - Variáveis ordinais: matrizes de correlações policóricas
 - Variáveis dicotômicas: matrizes de correlações tetracóricas

- Avaliação das variáveis
 - Normalidade
 - Univariada: assimetria e curtose ± 1 (restritivo) ou 2 e 7, respectivamente
 - Testes univariados (Shapiro-Wilk e Kolmogorov-Smirnov)
 - Multivariada: Coeficiente de Mardia (p associado)
 - Gráfico de dispersão e matriz de correlações
 - Associação linear
 - Correlações acima de 0,30 (aprox. 10% de var.)

- Tamanho amostral
 - A AFE é uma técnica para grandes amostras
 - São sugeridas as razões 5:1 quando a dimensionalidade é conhecida e de 10:1 quando não conhecida (Pasquali, 2010)
 - O poder estatístico da AFE é uma função da comunalidade (h²)

Table 3: The Effects of Subject to Item Ratio on Exploratory Factor Analysis

Variable:	2:1	5:1	10:1	20:1	F (3,76)=	$\eta^2 =$
% samples with correct factor structure	10%	40%	60%	70%	13.64***	.21
Average number of items misclassified on wrong factor	1.93	1.20	0.70	0.60	9.25***	.16
Average error in eigenvalues	0.41	0.33	.20	.16	25.36***	.33
Average error in factor loadings	.15	.12	.09	.07	36.38***	.43
% analyses failing to converge after 250 iterations	30%	0%	0%	0%	8.14***	.24
% with Heywood cases	15%	20%	0%	0%	2.81	.10
*** p < .0001	* p < .0001 (Costello & Osbori					ne, 2005)

- Razão itens/fatores
 - É sugerida a razão mínima de 3:1, sendo ótima ≥ 5:1
- Medidas de adequação dos dados
 - Kaiser-Meyer-Olkin (KMO): medida do grau de intercorrelações entre as variáveis
 - Interpretação: ≥ 0.80 (ótimo), ≥ 0.70 (bom), ≥ 0.60 (fraco), ≥ 0.50 (ruim) e < 0.50 (inaceitável)
 - Teste de esfericidade de Bartlett
 - Testa a hipótese nula que a matriz de correlações é igual a uma matriz identidade. Nível de significância associado.

Qualidade dos dados

- KMO = proporção entre correlações bivariadas e correlações parciais
 - Quanto mais próximo à unidade, melhor
 - Aceitável acima de 0,6
- Teste de esfericidade de Batlett (p<0,05)

Γ	1	0	0	0
	0	1	0	0
	0	0	1	0
L	0	0	0	1

Métodos de extração de fatores

TABLE 13.7 Summary of Extraction Procedures

Extraction Technique	Program	Goal of Analysis	Special Features
Principal components	SPSS SAS	Maximize variance extracted by orthogonal components	Mathematically determined, empirical solution with common, unique, and error variance mixed into components
Principal factors	SPSS SAS	Maximize variance extracted by orthogonal factors	Estimates communalities to attempt to eliminate unique and error variance from variables
Image factoring	SPSS SAS (Image and Harris)	Provides an empirical factor analysis	Uses variances based on multiple regression of a variable with all other variables as communalities to generate a mathe- matically determined solution with error variance and unique variance eliminated
Maximum likelihood factoring	SAS SPSS	Estimate factor loadings for population that maximize the likelihood of sampling the observed correlation matrix	Has significance test for factors; especially useful for confirmatory factor analysis
Alpha factoring	SPSS SAS	Maximize the generalizability of orthogonal factors	
Unweighted least squares	SPSS SAS	Minimize squared residual correlations	
Generalized least squares	SPSS SAS	Weights variables by shared variance before minimizing squared residual correlations	

(Tabachnick & Fidell, 2007)

Métodos de extração de fatores

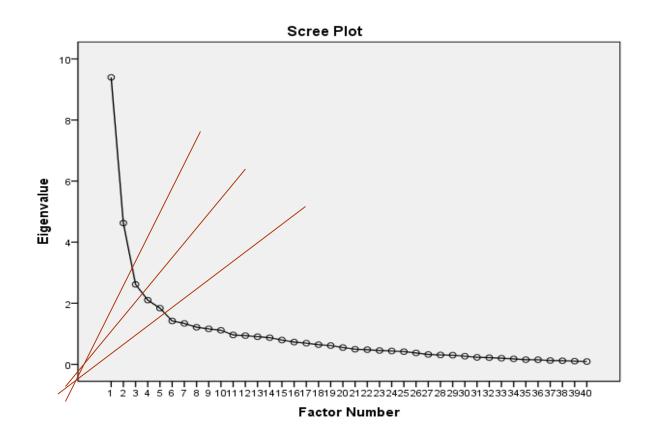
- Métodos de extração de fatores comuns
 - Maximum Likelihood: distribuição normal
 - Principal Factors ou Principal Axis: violações da normalidade

Métodos de retenção dos fatores

- Critério de Kaiser
 - Autovalores (eigenvalues) ≥1
 - Autovalor = corresponde a soma do quadrado das cargas fatoriais. Representa a quantidade de variância explicada pelo fator.

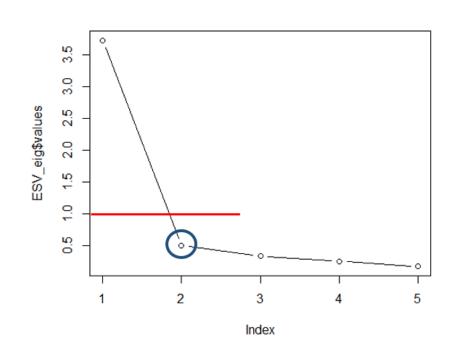
Métodos de retenção dos fatores

• Critério gráfico de Cattell (scree-test)



Técnicas de retenção de fatores

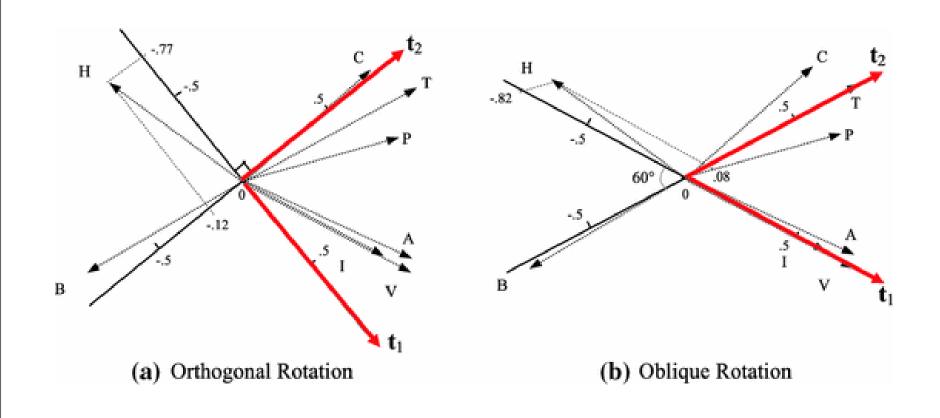
- Critério de Kaiser
- Critério de Cattell



Métodos de rotação de fatores

- O objetivo das técnicas de rotação de é simplificar a estrutura de relações entre fatores e variáveis
- Existem dois tipos de rotação de fatores
 - Ortogonais
 - Impõe que os fatores sejam não-correlaciados
 - Oblíquas
 - Permite a correlação entre os fatores

• A rotação fatorial é uma técnica que visa ajustar a solução fatorial (cargas fatoriais) de modo a deixá-la mais interpretável, ou clara. A solução não rotacionada pode levar aos itens apresentarem correlações com vários fatores. Para diminuir este efeito, os eixos dos fatores são rotacionados no espaço, de modo a permitir (rotação oblíqua) ou não (rotação ortogonal) a correlação entre os fatores.



Métodos de rotação de fatores

TABLE 13.9 Summary of Rotational Techniques

(Tabachnick & Fidell, 2007)

Rotational Technique	Program	Туре	Goals of Analysis	Comments
Varimax	SAS SPSS	Orthogonal	Minimize complexity of factors (simplify columns of loading matrix) by maximizing variance of loadings on each factor.	Most commonly used rotation; recommended as default option
Quartimax	SAS SPSS	Orthogonal	Minimize complexity of variables (simplify rows of loading matrix) by maximizing variance of loadings on each variable.	First factor tends to be general, with others subclusters of variables.
Equamax	SAS SPSS	Orthogonal	Simplify both variables and factors (rows and columns); compromise between quartimax and varimax.	May behave erratically
Orthogonal with gamma (orthomax)	SAS	Orthogonal	Simplify either factors or variables, depending on the value of gamma (Γ) .	Gamma (Γ) continuously variable
Parsimax	SAS	Orthogonal	Simplifies both variables and factors: $\Gamma = (p^*(m-1))/p + m - 2.$	

Métodos de rotação de fatores

TABLE 13.9 Summary of Rotational Techniques

(Tabachnick & Fidell, 2007)

Rotational Technique	Program	Туре	Goals of Analysis	Comments
Direct oblimin	SPSS	Oblique	Simplify factors by minimizing cross-products of loadings.	Continuous values of gamma, or delta, δ (SPSS), available; allows wide range of factor intercorrelations
(Direct) quartimin	SPSS	Oblique	Simplify factors by minimizing sum of cross-products of squared loadings in pattern matrix.	Permits fairly high correlations among factors. Achieved in SPSS by setting $\delta = 0$ with direct oblimin.
Orthoblique	SAS (HK) SPSS	Both orthogonal and oblique	Rescale factor loadings to yield orthogonal solution; non-rescaled loadings may be correlated.	
Promax	SAS	Oblique	Orthogonal factors rotated to oblique positions.	Fast and inexpensive
Procrustes	SAS	Oblique	Rotate to target matrix.	Useful in confirmatory FA

Avaliação do modelo fatorial

- Comunalidades (h²):
 - Valores ≥ 0.40
- Cargas fatoriais
 - > de 0,30 em um dos fatores retidos
 - <0,30 nos demais fatores, cargas cruzadas*
- Correlação entre os fatores
 - Não deve exceder 0,80
- Variância explicada
 - Var. total 50 a 60%
- Consistência interna Alpha de Cronbach

Métodos robustos em AFE

- Métodos robustos são preferíveis quando há violação dos pressupostos analíticos dos testes paramétricos
- Esses pressupostos são:
 - Normalidade univariada
 - Normalidade multivariada
 - Homoscedasticidade (igualdade de variâncias)
 - Nível de mensuração intervalar ou escalar
- A utilização de métodos paramétricos sob violação de seus pressupostos pode ocasionar a super ou subestimação de intervalos de confiança, tamanhos de efeito e níveis de significância estatística
- Métodos robustos mantém fidedignos os valores de probabilidade (α), associação, diferenças, etc...

Métodos robustos em AFE

Tabela 1. Classificação de variáveis de Stevens.

Escala	Operação	Exemplos	Localização	Dispersão	Associação	Teste
Nominal	Igualdade	Números de jogadores	Moda			Qui-quadrado
Ordinal	Maior ou menor	Dureza de minerais	Mediana	Percentil	Correlação de	Testes de
		Números de rua			ordem (rank)	ranking
		Escores brutos				
Intervalar	Distância	Temperatura Celsius	Média aritmética	Desvio Padrão	Correlação	t-teste
		Tempo			produto	F-teste
		Escores padronizados?			momento	
Razão	Razão	Distância	Média geométrica	% de variação		
		Temperatura Kelvin	Média harmônica			
		Temperatura Kelvin	Média harmônica			

(Chachamovich, 2007)

Métodos robustos em AFE

- Exemplos:
- Método de extração Principal Factors ou Principal Axis (Eixos principais)
 - Robusto para desvios de normalidade univariada

(Costelo & Osborne, 2005)

- Matriz de correlações policóricas
 - Robusto para desvios de normalidade univariada e nível de mensuração ordinal

(Holgado-Tello et al., 2010)

Escores fatoriais

- Fiz a AFE, e agora?
- () somar os itens e usar o resultado em outras análises
- () usa a informação da modelagem estatística, ponderando a importância de cada item, em outras análises
- Escores fatoriais são estimados por meio da resolução do modelo linear da análise fatorial e são ideais para conduzir outros testes estatísticos, pois retiram o erro de mensuração dos itens.

Consistência interna

- Alpha de Cronbach
- Mais utilizado porém polêmico
- Representa o "lower bound" da fidedignidade

PSYCHOMETRIKA—VOL. 74, NO. 1, 107–120 MARCH 2009 DOI: 10.1007/s11336-008-9101-0

ON THE USE, THE MISUSE, AND THE VERY LIMITED USEFULNESS OF CRONBACH'S ALPHA

KLAAS SIJTSMA
TILBURG UNIVERSITY

• (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2792363/pdf/11336 2008 Article 9101.pdf)

Um exemplo passo a passo

• AFE da Escala de Satisfação com a Vida

Anexo

Escala de Satisfação com a Vida

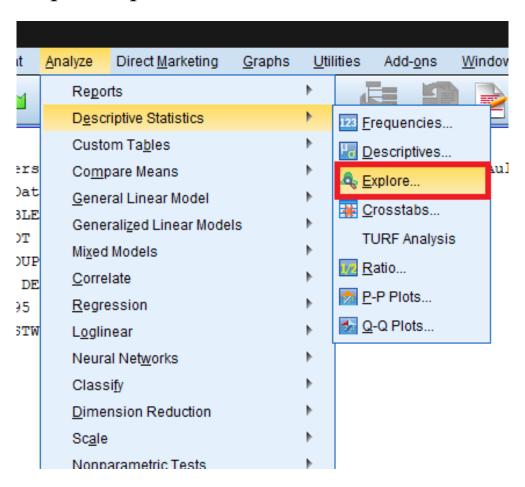
Instruções

Abaixo você encontrará cinco afirmações com as quais pode ou não concordar. Usando a escala de resposta a seguir, que vai de 1 a 7, indique o quanto concorda ou discorda com cada uma; escreva um número no espaço ao lado da afirmação, segundo sua opinião. Por favor, seja o mais sincero possível nas suas respostas.

- 7 = Concordo totalmente
- 6 = Concordo
- 5 = Concordo ligeiramente
- 4 = Nem concordo nem discordo
- 3 = Discordo ligeiramente
- 2 = Discordo
- 1 = Discordo totalmente
- Na maioria dos aspectos, minha vida é próxima ao meu ideal.
- As condições da minha vida são excelentes.
- Estou satisfeito(a) com minha vida.
- Dentro do possível, tenho conseguido as coisas importantes que quero da vida.
- 5.____Se pudesse viver uma segunda vez, não mudaria quase nada na minha vida.

(http://www.vvgouveia.net/sp/images/Gouveia 2005 medindo a satisfao com a vida dos mdicos no brasil.pdf)

Avaliando pressupostos



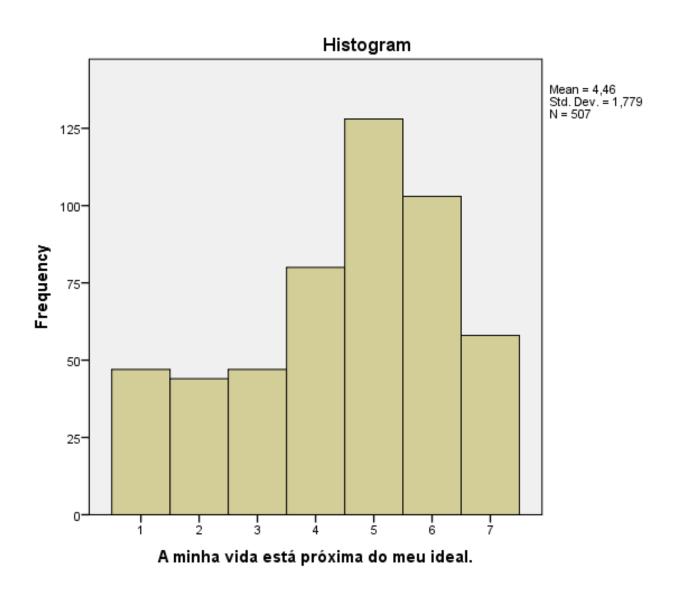
Descriptives

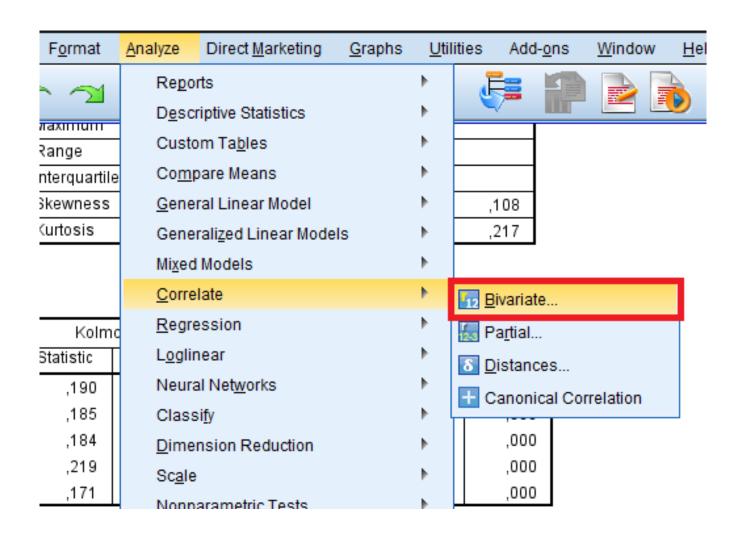
lean 5% Confidence Interval or Mean	Lower Bound	4,46	,079
	Lower Bound		1
or Mean		4,30	
	Upper Bound	4,61	
% Trimmed Mean		4,51	
ledian		5,00	
Variance		3,166	
Std. Deviation		1,779	
Minimum		1	
Maximum		7	
Range		6	
nterquartile Range		3	
kewness		-,492	,108
(urtosis		-,712	,217
lean		4,68	,074
/a	ariance d. Deviation inimum aximum ange terquartile Range kewness	ariance d. Deviation inimum aximum ange terquartile Range kewness	ariance 3,166 id. Deviation 1,779 inimum 1 aximum 7 ange 6 terquartile Range 3 xewness -,492 urtosis -,712

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ESV01	,190	507	,000	,917	507	,000
ESV02	,185	507	,000	,923	507	,000
ESV03	,184	507	,000	,915	507	,000
ESV04	,219	507	,000	,870	507	,000
ESV05	,171	507	,000	,893	507	,000

a. Lilliefors Significance Correction



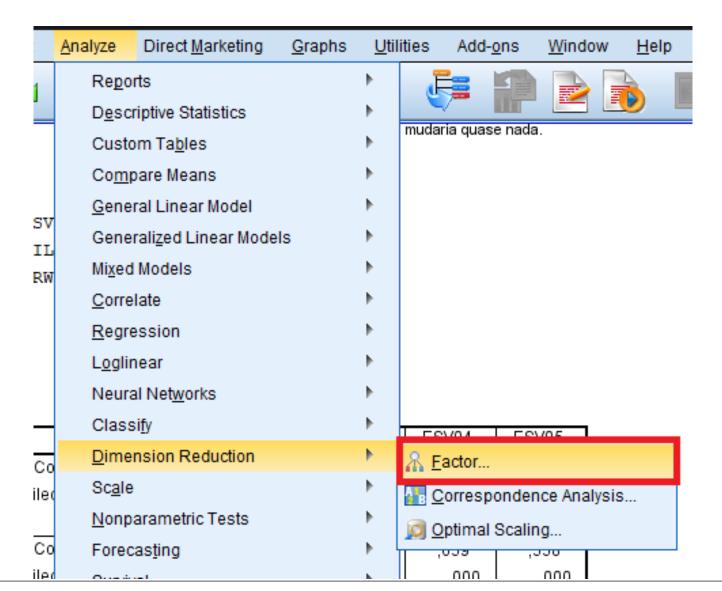


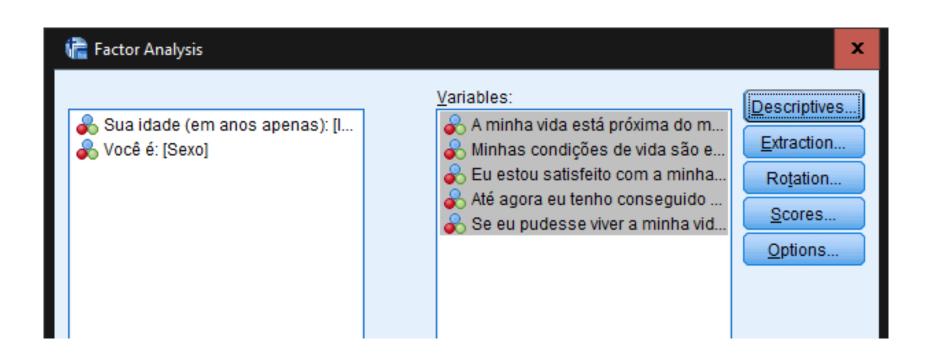
Correlations

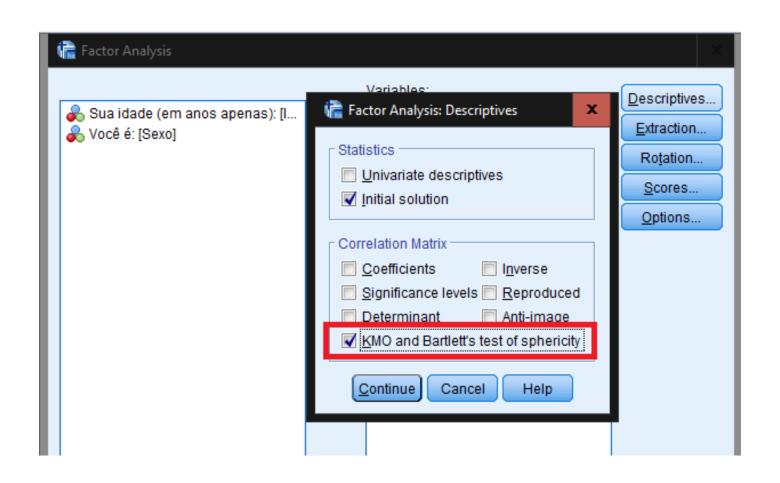
		ESV01	ESV02	ESV03	ESV04	ESV05
ESV01	Pearson Correlation	1	,743**	,797**	,694**	,596**
	Sig. (2-tailed)		,000	,000	,000	,000
	N	510	510	509	510	508
ESV02	Pearson Correlation	,743**	1	,782**	,659**	,558**
	Sig. (2-tailed)	,000		,000	,000	,000
	N	510	510	509	510	508
ESV03	Pearson Correlation	,797**	,782**	1	,733**	,634**
	Sig. (2-tailed)	,000	,000		,000	,000
	N	509	509	509	509	507
ESV04	Pearson Correlation	,694**	,659**	,733**	1	,615**
	Sig. (2-tailed)	,000	,000	,000		,000
	N	510	510	509	510	508
ESV05	Pearson Correlation	,596**	,558**	,634**	,615**	1
	Sig. (2-tailed)	,000	,000	,000	,000	
	N	508	508	507	508	509

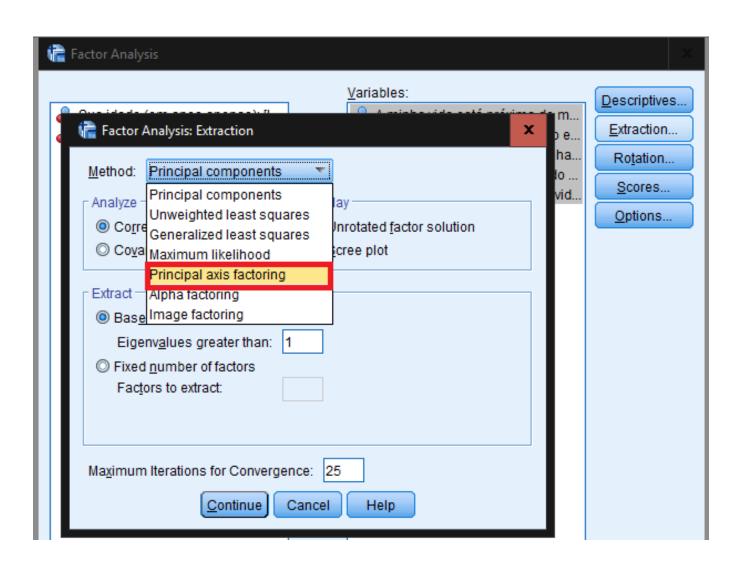
^{**.} Correlation is significant at the 0.01 level (2-tailed).

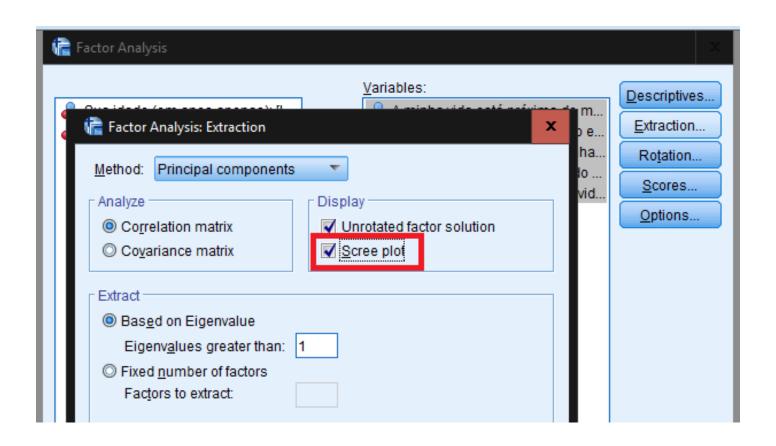
Análise Fatorial

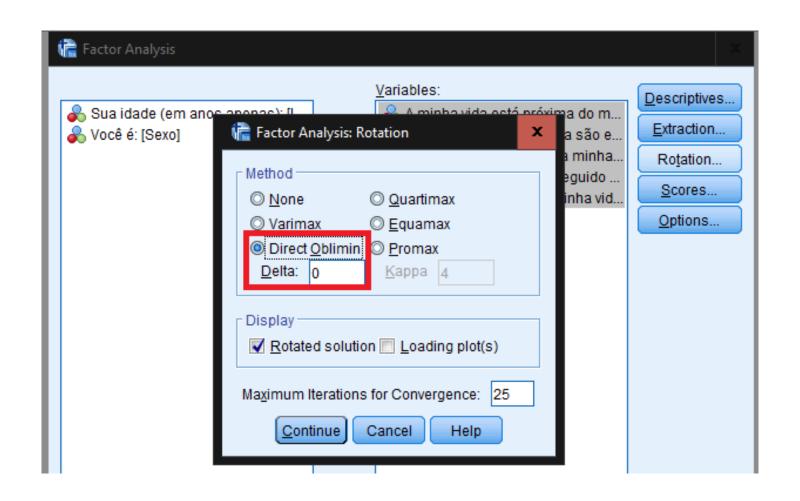


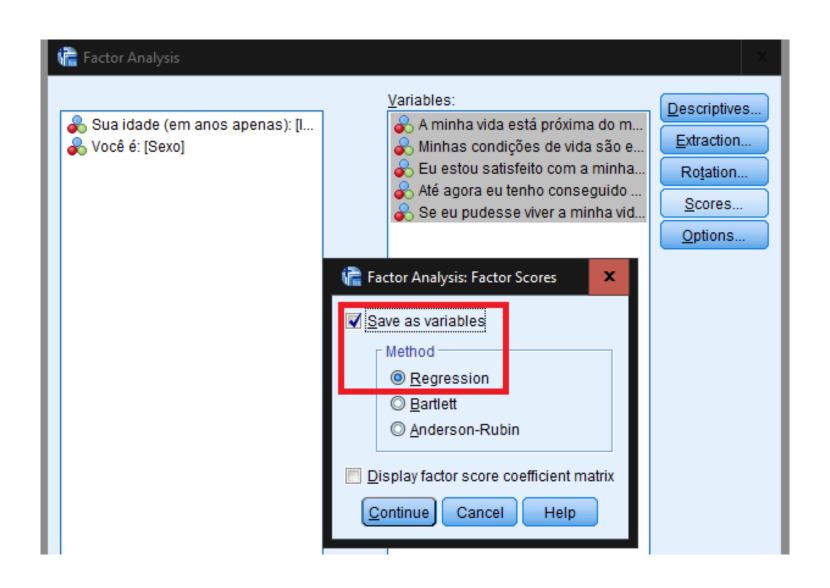


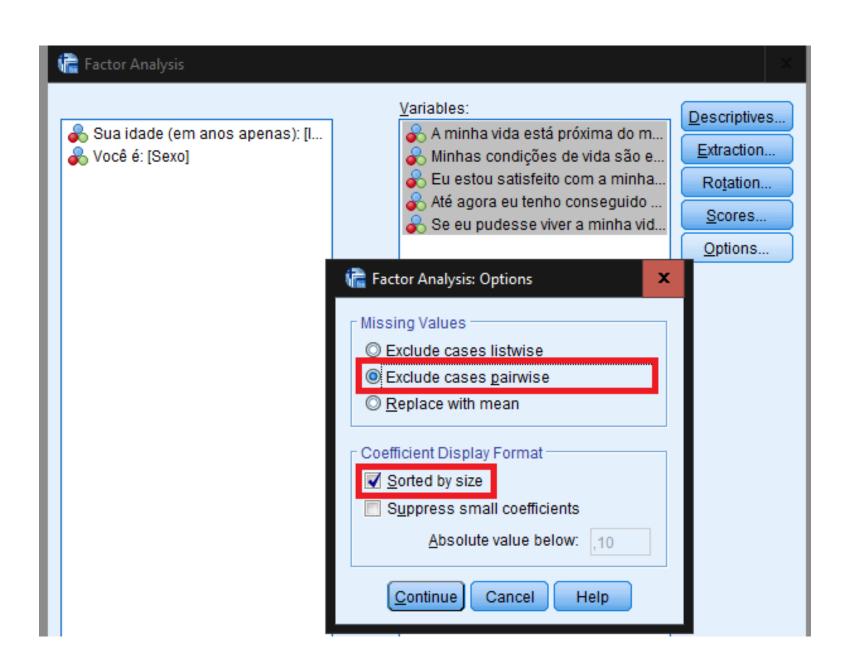












KMO and Bartlett's Test

Kaiser-Meyer-Olkin Mea	,889	
Bartlett's Test of	Approx. Chi-Square	1783,229
Sphericity	df	10
	Sig.	,000

Communalities

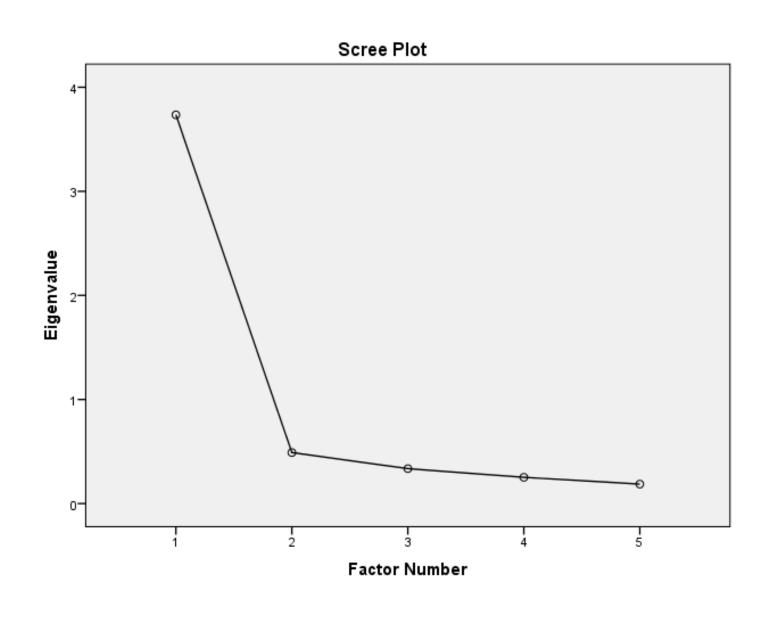
	Initial	Extraction
ESV01	,692	,754
ESV02	,658	,696
ESV03	,754	,841
ESV04	,602	,658
ESV05	,461	,490

Extraction Method: Principal Axis Factoring.

Total Variance Explained

	Initial Eigenvalues			Extraction	n Sums of Square	ed Loadings
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,735	74,704	74,704	3,439	68,783	68,783
2	,490	9,802	84,506			
3	,336	6,710	91,216			
4	,252	5,044	96,260			
5	,187	3,740	100,000			

Extraction Method: Principal Axis Factoring.



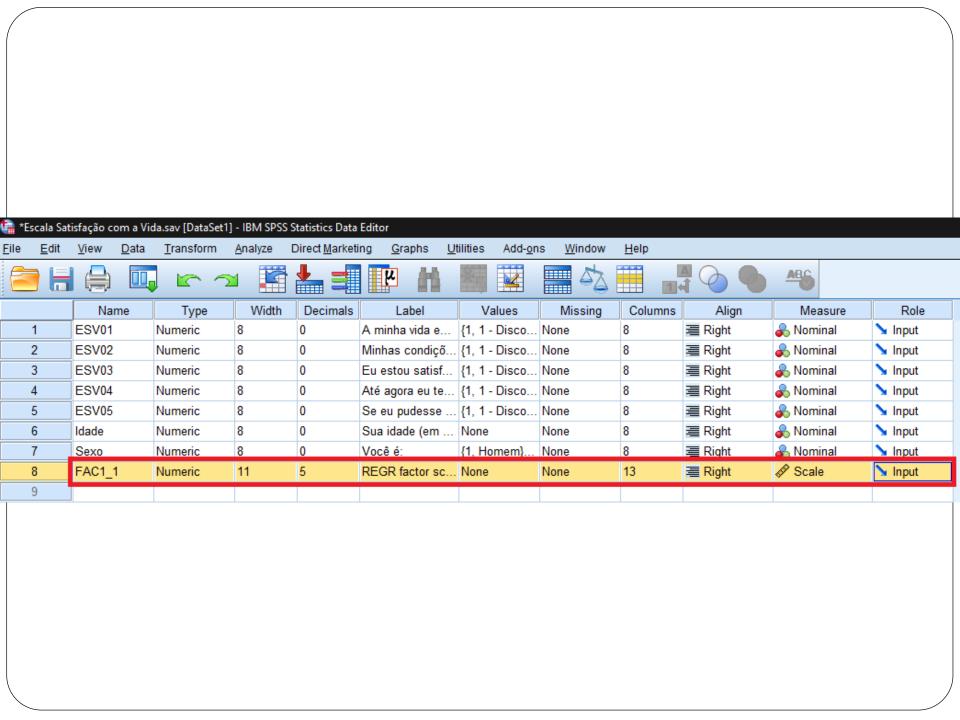
Factor Matrix^a

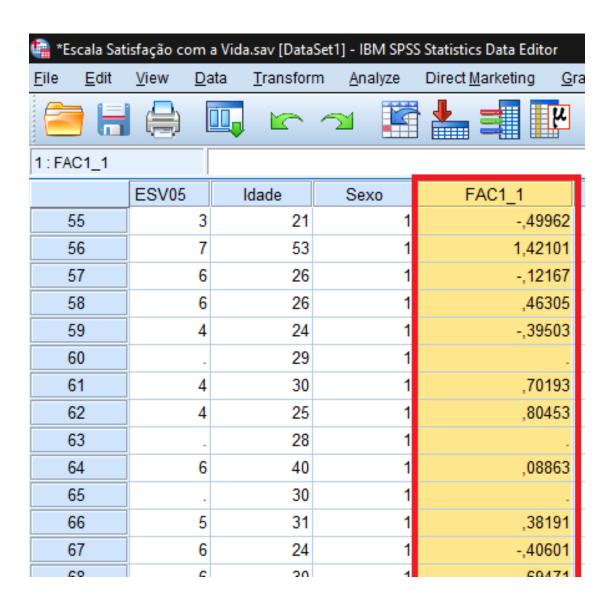
	Factor
	1
ESV03	,917
ESV01	,868,
ESV02	,834
ESV04	,811
ESV05	,700

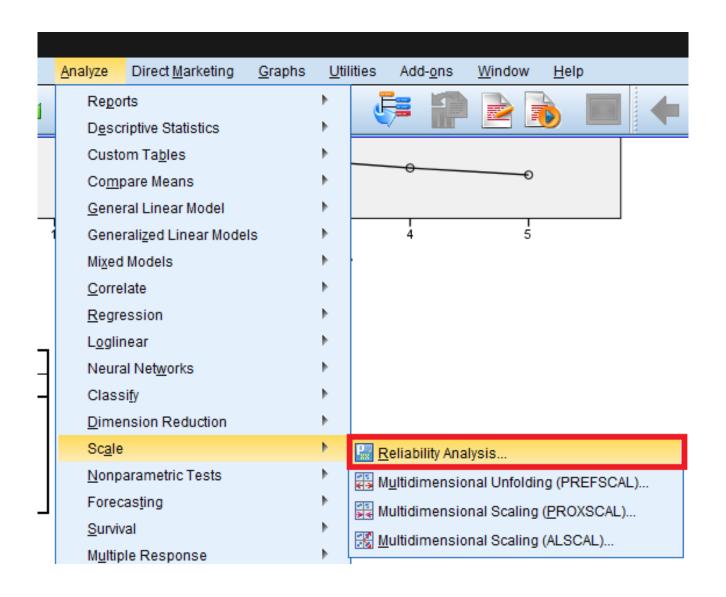
Extraction Method:

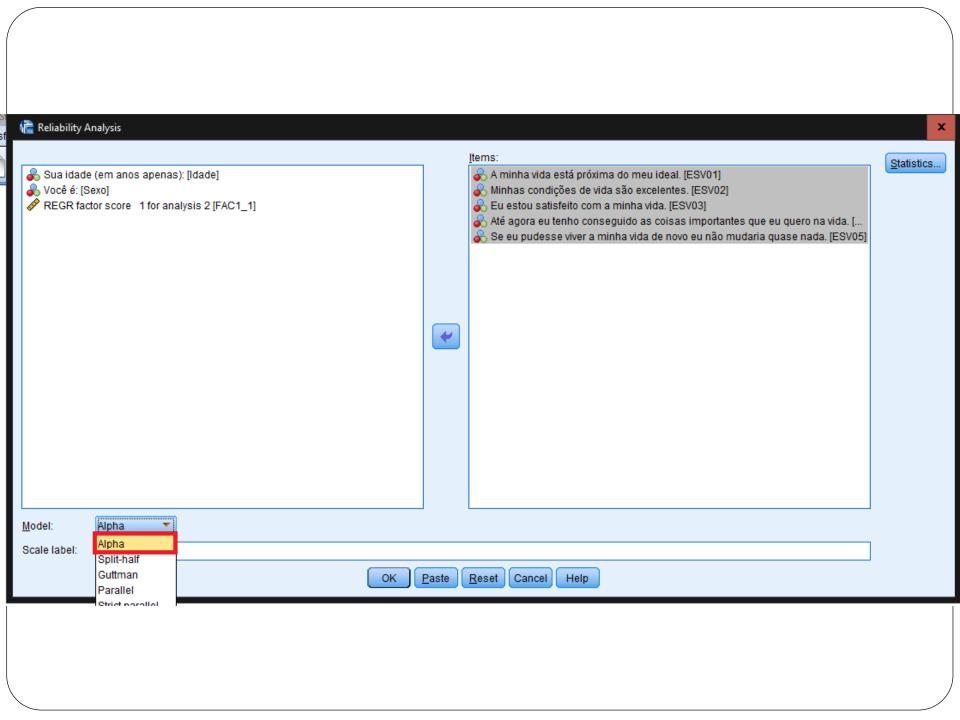
Principal Axis Factoring.

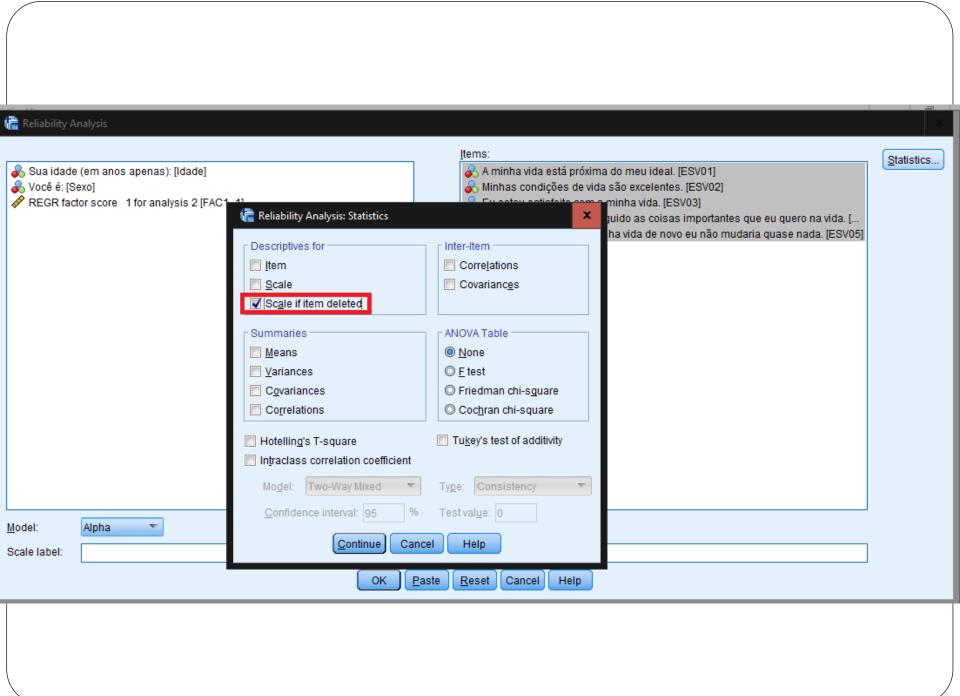
a. 1 factors extracted. 6 iterations required.











Reliability Statistics

Cronbach's		
Alpha		N of Items
	,911	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
ESV01	19,03	37,890	,815	,883
ESV02	18,81	39,644	,783	,890
ESV03	18,76	37,790	,864	,874
ESV04	18,26	39,571	,774	,892
ESV05	19,10	37,797	,670	,919

Outros exemplos

- Escala de Estresse Percebido
- Marcadores Adjetivos dos Cinco Grandes Fatores
- Escala de Afetos Positivos e Negativos PANAS

Como reportar?

• Foi conduzida uma análise fatorial exploratória, com o método de extração XXX e método de rotação XXX. As medidas de adequação dos dados ao modelo fatorial consideradas foram o índice KMO e o teste de esfericidade de Bartlett. Para interpretação do modelo foram considerados os autovalores (Eigenvalues), as comunalidades e as cargas fatoriais dos itens, bem como, a correlação entre os fatores.

Como reportar?

Itens	Fator 1	Fator 2
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Conteúdo total ou parcial	Cargas fatoriais	Cargas fatoriais
Variância/Eigenvalue		
Fidedignidade (alpha)		
Descritivos (M e SD)		