



SOFTWARE R APLICADO À PSICOMETRIA

Ministrantes:

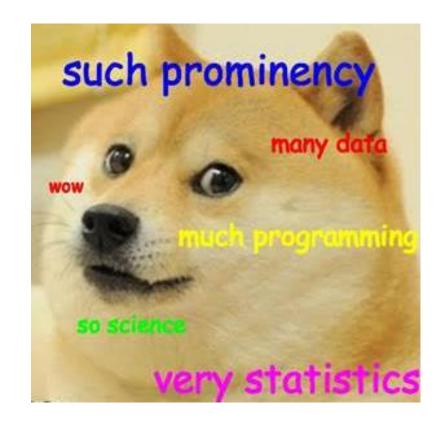
Wagner de Lara Machado (PUCRS) Euclides Mendonça (UFRGS)

MATERIAL DO CURSO

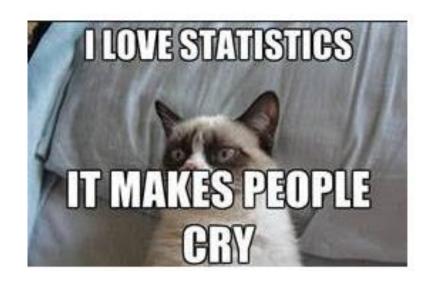
https://github.com/wagnerLM

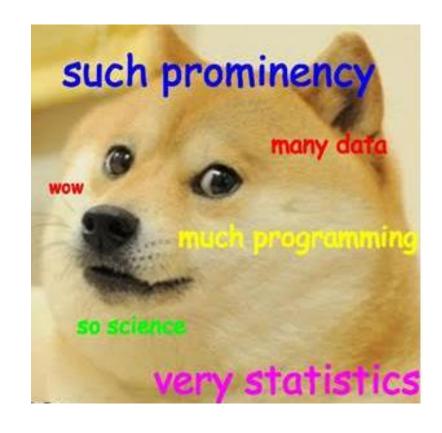
Link: SBP

https://github.com/wagnerLM/SBP











O QUÊ É A LINGUAGEM R?

R é um "dialeto" da linguagem S

1993 R é lançado ao público

1995 General Public License

1997 o R Core Group é formado – Controla a fonte de códigos do R

2000 R version 1.0.0 é lançado

2015 version 3.2.1



Capacidades gráficas muito sofisticadas e melhores que muitos softwares

Linguagem de programação que possibilita o desenvolvimento de novas ferramentas

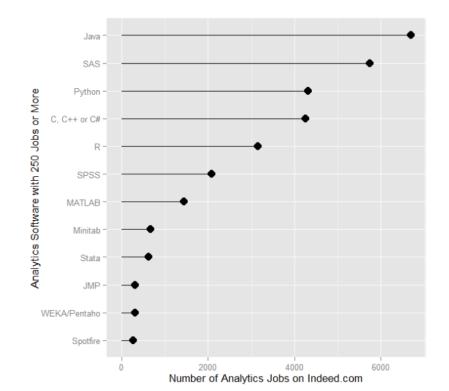
Comunidade de usuários muito ativa e participativa

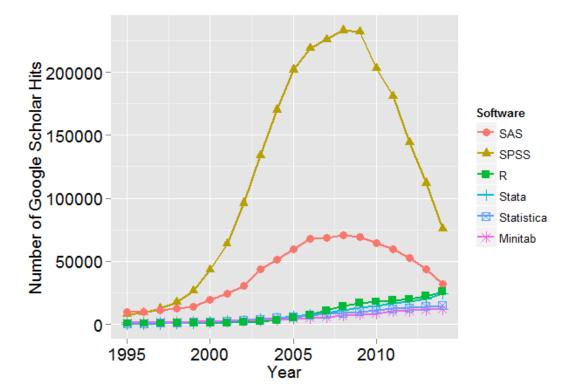


Capacidades gráficas muito sofisticadas e melhores que muitos softwares

Linguagem de programação que possibilita o desenvolvimento de novas ferramentas

Comunidade de usuários muito ativa e participativa

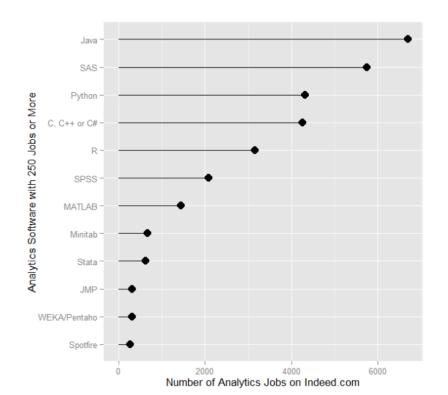


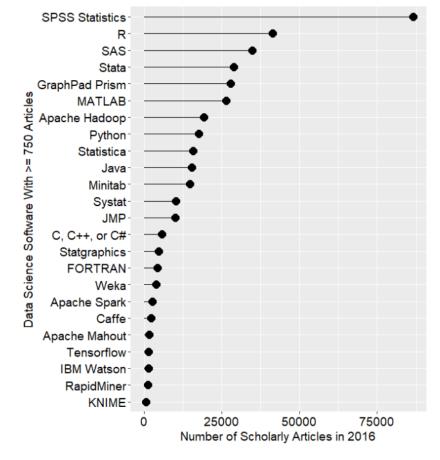


Capacidades gráficas muito sofisticadas e melhores que muitos softwares

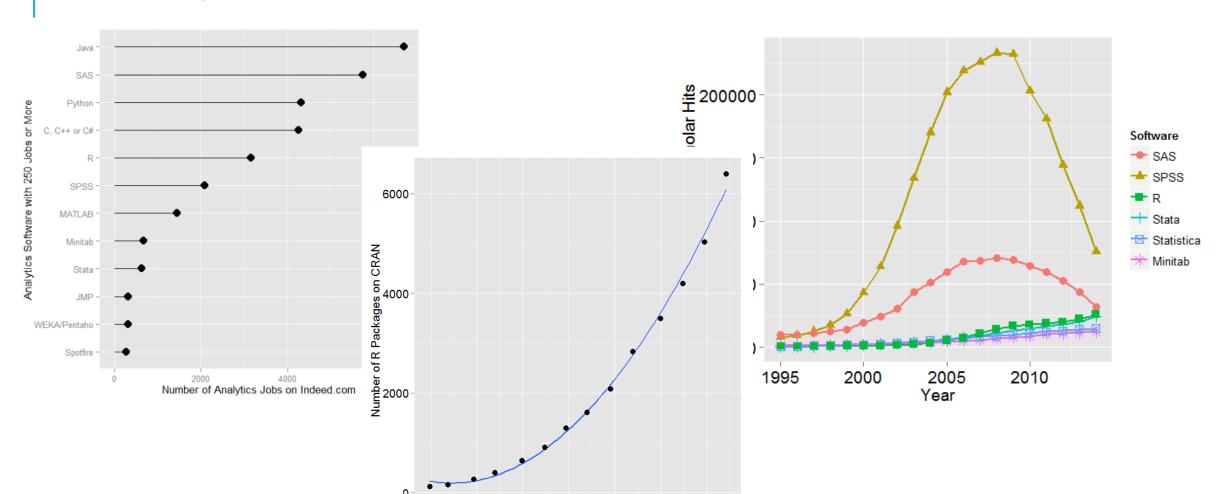
Linguagem de programação que possibilita o desenvolvimento de novas ferramentas

Comunidade de usuários muito ativa e participativa





Year



Capacidades gráficas muito sofisticadas e melhores que muitos softwares

Linguagem de programação que possibilita o desenvolvimento de novas ferramentas

Comunidade de usuários muito ativa e participativa





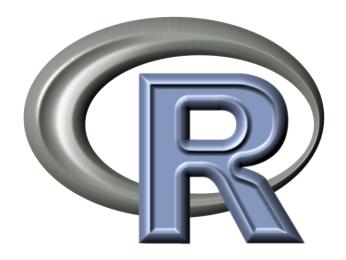


Psicometristas Psicometristas (Brasil) ♦ Compartilhar ✓ Notificações ••• Procurar neste grupo

Personality Project







GETTING STARTED



- É o sistema básico feito de forma colaborativa – Possui funções básicas que podem ser combinadas em pacotes mais avançados
- R Studio Uma interface mais amigável para o uso do R. Permite análise, escrita (tem gente que escreve a tese por aqui), e publicações.

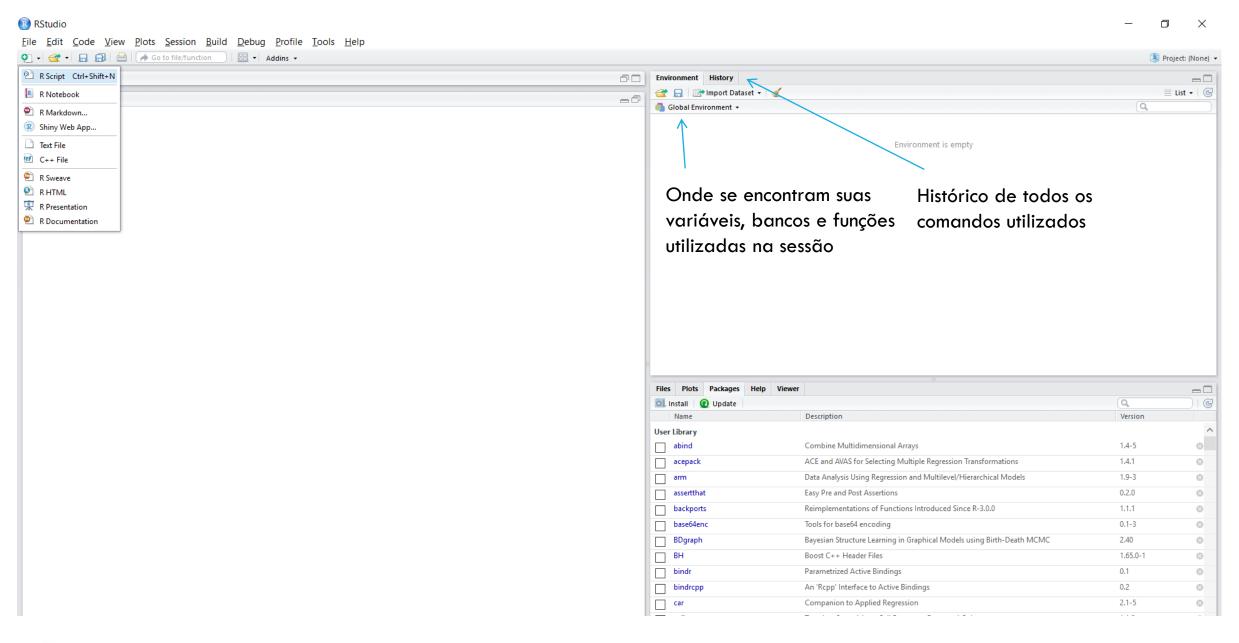


R DESING

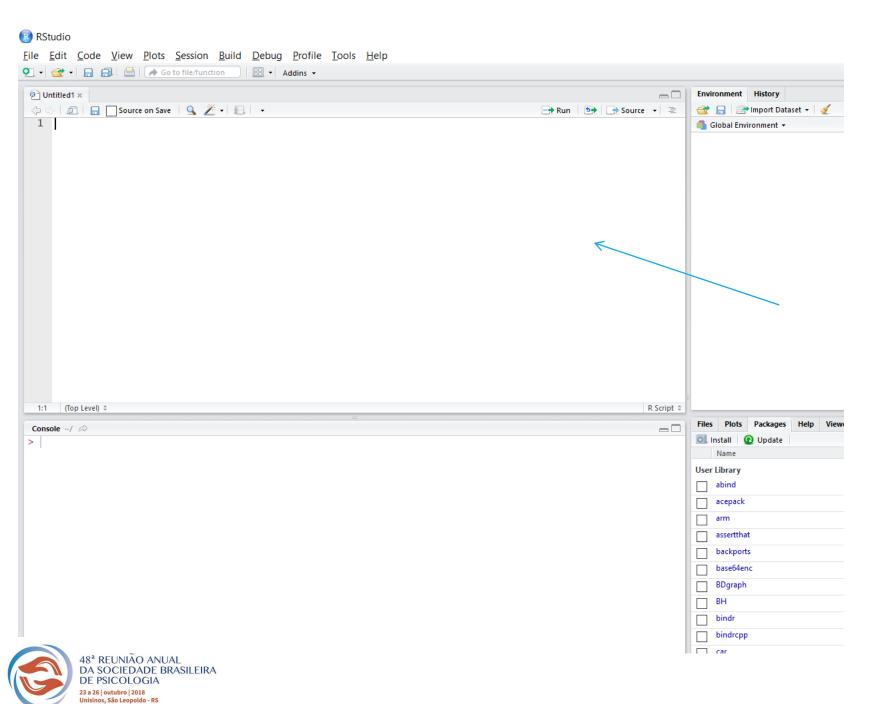
O sistema R é divido em duas partes:

- A "base" do sistema R que pode-se fazer o download no CRAN
- Todo o resto Funcionalidades chamadas de packages
- Atualmente existem mais de 6000 packages

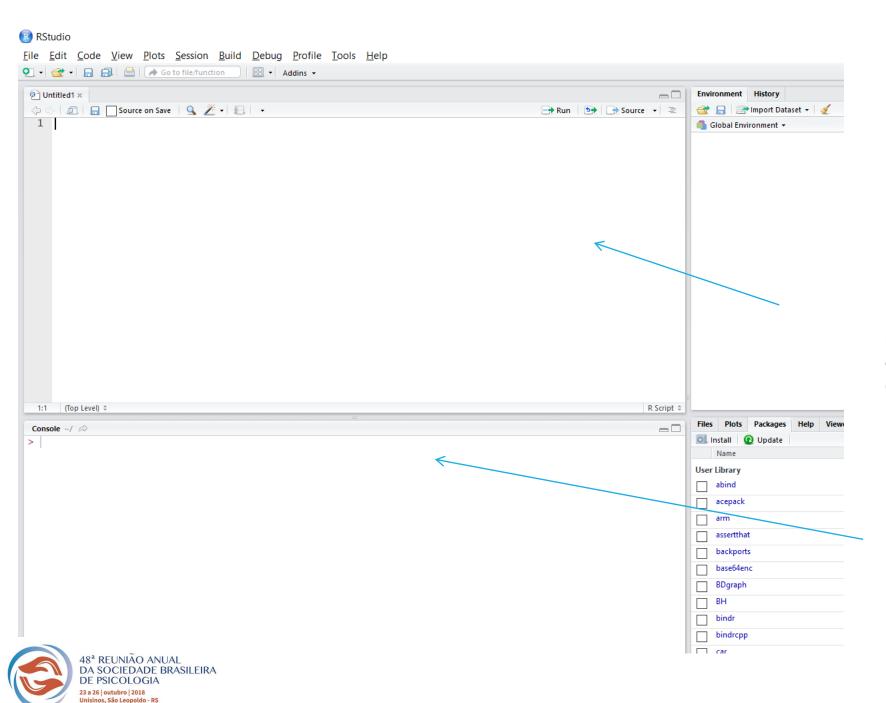






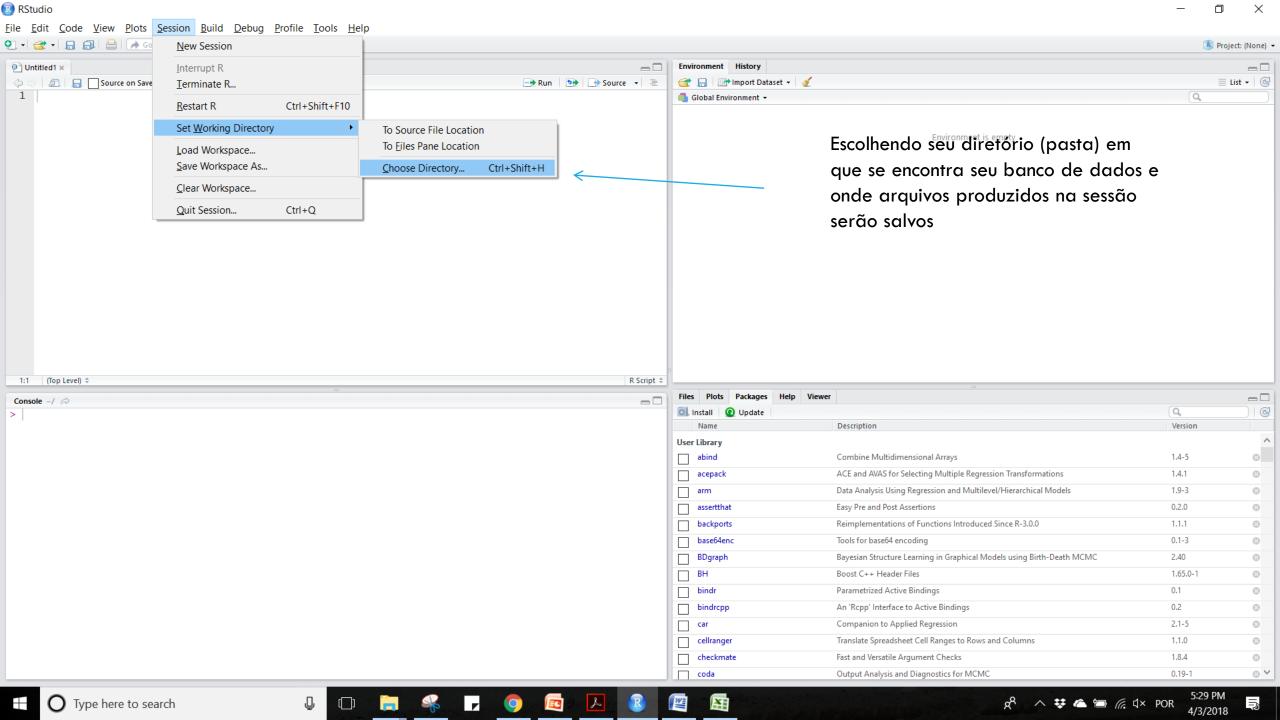


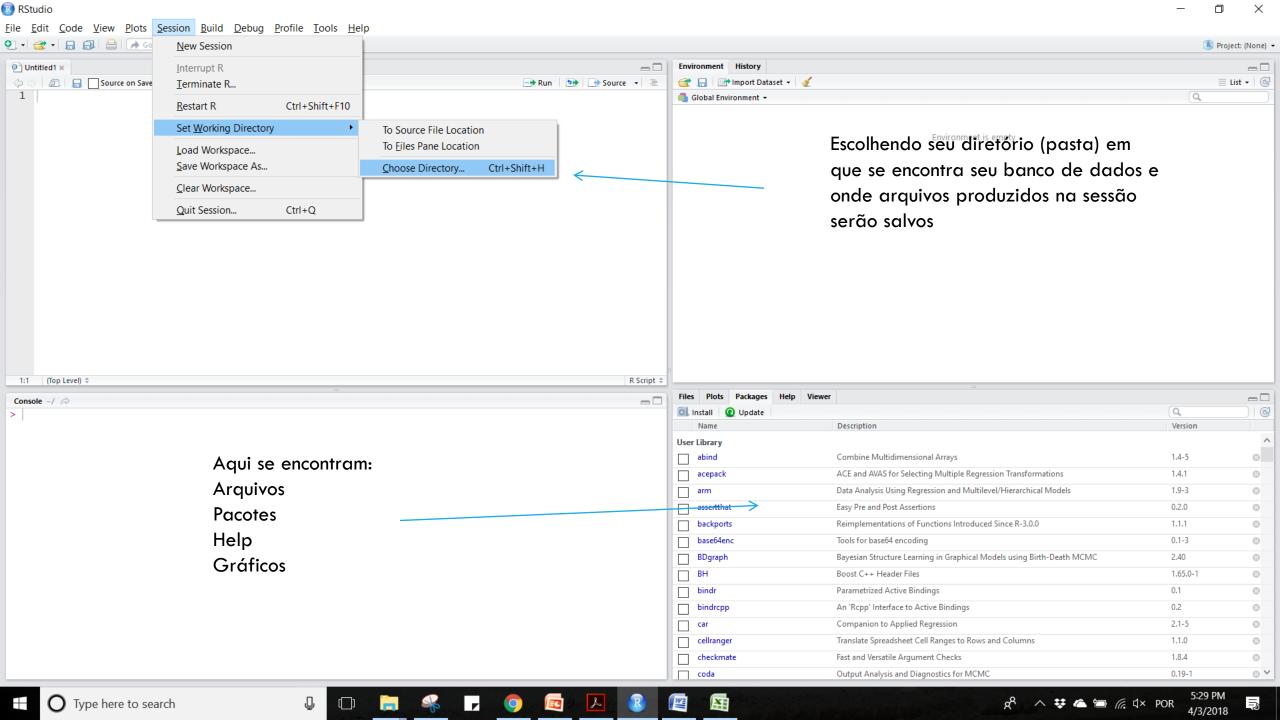
Editor de sintaxes.
Os comandos podem ser escritos e "rodados" nessa caixa de script



Editor de sintaxes. Os comandos podem ser escritos e "rodados" nessa caixa de script

Interface do R básico. Aqui saem os outputs das análises. Comandos também podem ser escritos aqui, mas não é aconselhável





FUNCIONALIDADES BÁSICAS

- Assinalando valores
- Vetor X Matriz
- Função
- Tipos de variáveis
- Carregando Banco de Dados
- Instalar Packages



O Banco de dados no R é definido por linhas e colunas em que:

Linhas – Representam casos

Colunas – Variáveis

Para selecionar essas dimensões do banco utiliza-se os [,]

Ex: Banco [Linhas, Colunas]



Ex: Banco [Linhas, Colunas]

Para selecionar os 10 primeiros casos e todas as variáveis

- Banco[1:10,]

Para selecionar os 10 primeiros casos e 2 primeiras variáveis

- Banco[1:10, 1:2]



Ex: Banco [Linhas, Colunas]

Para selecionar os 10 primeiros casos e 2 primeiras variáveis

- ESV [1:10, 1:2]

```
> ESV[1:10,1:2]
ESV1 ESV2
1 5 5
2 4 4
3 5 5
4 4 3
5 2 7
6 4 5
7 7 7
8 4 4
9 6 5
10 4 7
```



Uso do operador "\$"

ESV\$ESV1

Exibe todos os valores do Item1 da ESV



Selecionando somente os itens do Banco ESV ESV[,- c (6,7)] - O sinal de "-" exclui as colunas 6 e 7

> 1	ESVL ,	-0(0	,/)]		
	ESV1	ESV2	ESV3	ESV4	ESV5
1	5	5 4	5	6	2
2	5 4 5	4	5	5	2
3	5	5	5	6	5
4	4	3	3	4	2
5	2	7	4	6	2 2 5 2 4
4 5 6 7	4 2 4 7	5 7	4 7	3 7	3 7
	7		7		
8	4	4 5 7	5	4	6
9	6	5	5	6	6
10	4	7	2	6	4
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26	4 6 4 6 4 5 3	6	5 2 6 6 5	6	4
12	6	7 4 6	6	6	4
13	4	4	5	6	6
14	5	6	5	6	5 6
15	3	4 2	3	7	6
16	3	2	4	4	6
17	1 4 7	3 4 7	3	3	6
18	4	4	4 7	4 7	6
19	7				7
20	4	4 6	5 6	5	6 7
21	7	6	6	5 7 7	7
22	7	7	7		7
23	5 7	5 6 5	5	5	1
24	7	6	4 5	4 6	3
25	5 6 5	5	5	6	6
26	6	5 6	7	6	5
27 28		6	5	6	2 6
28	5	4	5	6	
29	6	6	6	6	5
30	7	6	5	6	7



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.

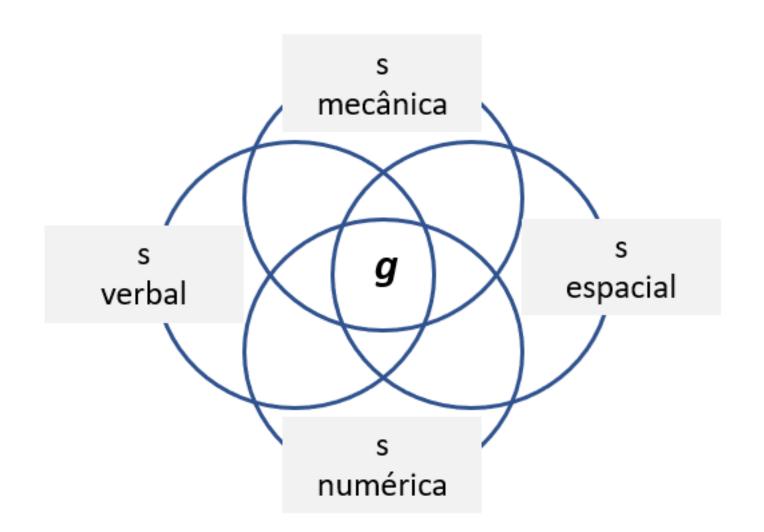
Exploratory Factor Analysis, Theory Generation, and Scientific Method

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

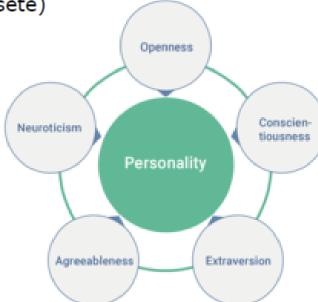


CINCO GRANDES FATORES DA PERSONALIDADE (BIG FIVE)

Por Que Cinco Fatores?

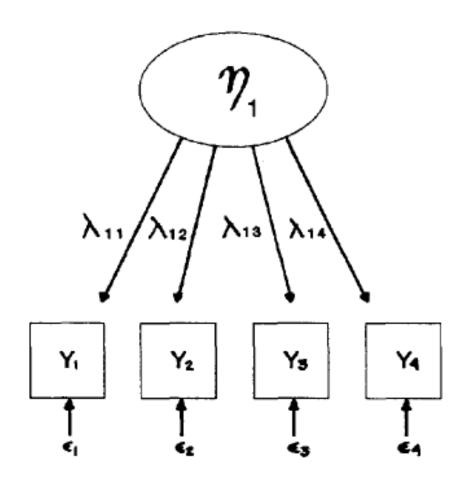
A descoberta dos cinco fatores foi acidental e se constitui em uma generalização empírica, replicada independentemente inúmeras vezes. Como o modelo não foi desenvolvido a partir de uma teoria, não há, consequentemente, uma explicação teórica *a priori* (e satisfatória) dos motivos que levariam a organização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

Psicología Reflexão e Crítica Universidad Federal do Rio Grande do Sul prcrev@ufrgs.br ISSN: 0102-7972 BRASIL



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



Uso da análise fatorial exploratória **EM PSICOLOGIA**

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

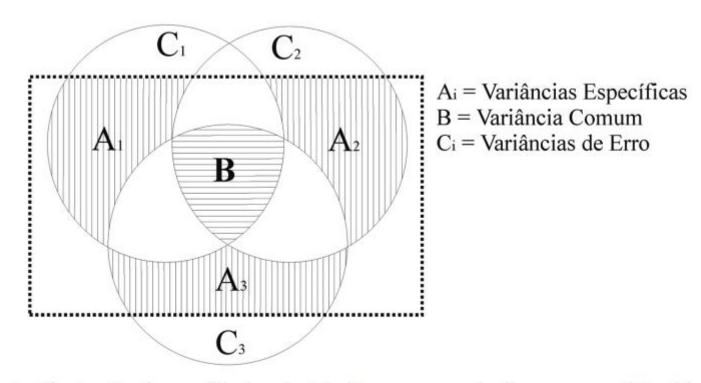


Figura 1 - Ilustração das variâncias de três itens e suas relações com um fator hipotético.

EM SUMA

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 isso, 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)

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;

F é um escore fatorial, valor latente do traço em questão;

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

PRESSUPOSTOS

Distribucionais – método de extração

Correlações – qualidade dos dados

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)

DISTRIBUIÇÕES

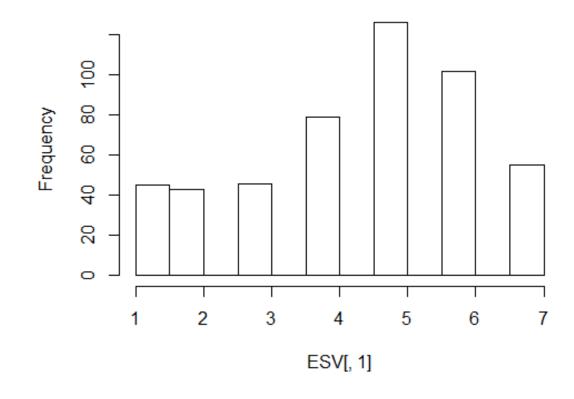
hist(ESV[,1]) shapiro.test(ESV[,1])

Normalidade: Pearson

Violações da normalidade:

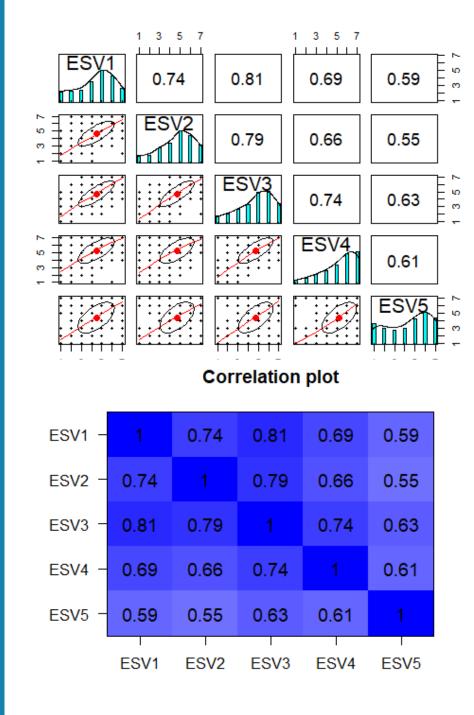
- Spearman
- Tetracóricas (dicotômicos)
- Policóricas (categorias ordenadas)

Histogram of ESV[, 1]



DISTRIBUIÇÕES E CORRELAÇÕES

cor.plot(ESV[,-c(6,7)],numbers = TRUE,cex = 0.8) pairs.panels(ESV[,-c(6,7)])



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)

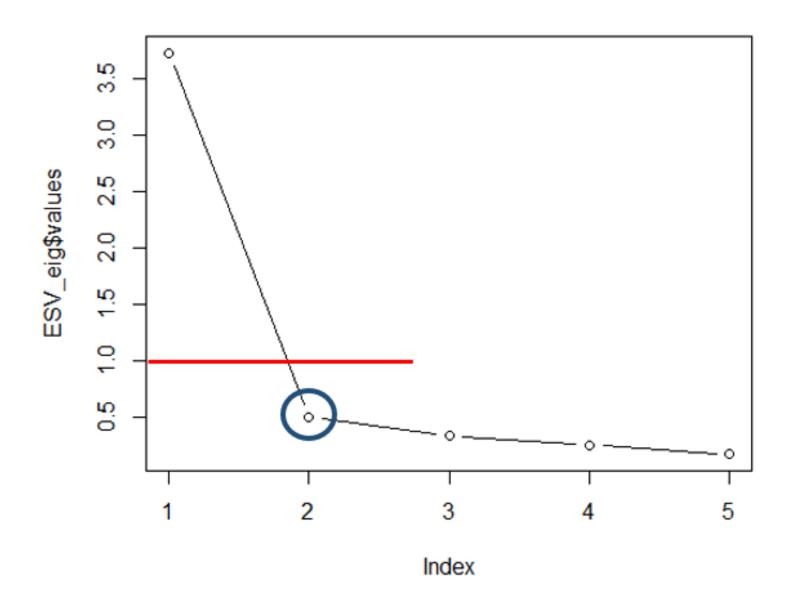
$$KMO = rac{\displaystyle\sum_{j
eq k} r_{jk}^2}{\displaystyle\sum_{j
eq k} r_{jk}^2 + \displaystyle\sum_{j
eq k} p_{jk}^2}$$

$$\left[\begin{array}{ccccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

TECNICAS DE RETENÇÃO DE FATORES

Critério de Kaiser

Critério de Cattell

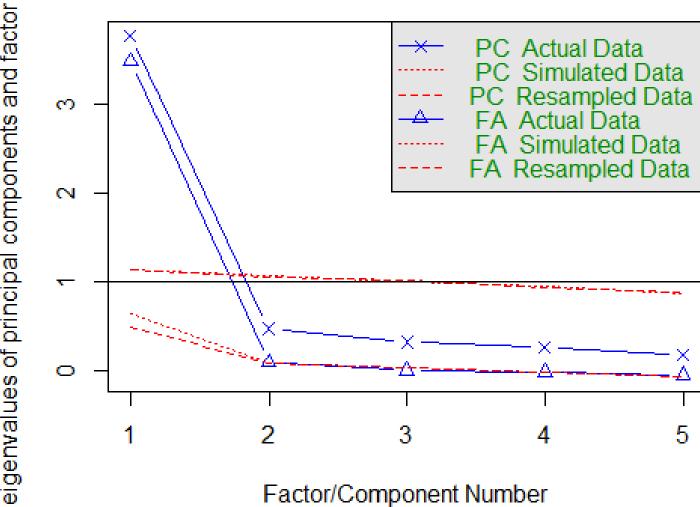


ANÁLISE PARALELA

Simular dados e comparar o poder explicativo dos fatores

Simulações: Monte Carlo (paramétrica) e por permutação dos valores

Parallel Analysis Scree Plots



fa(ESV[,-c(6,7)],cor="poly",fm="minrank")

Standardized loadings (pattern matrix) based upon correlation matrix

```
MRFA1 h2 u2 com

ESV1 0.87 0.75 0.25 1

ESV2 0.84 0.70 0.30 1

ESV3 0.92 0.85 0.15 1

ESV4 0.82 0.68 0.32 1

ESV5 0.74 0.55 0.45 1
```

MRFA1

SS loadings	3.52	
Proportion Var	0.70	

CONSISTÊNCIA INTERNA

0.9 0.91 0.92

Alpha de Cronbach alpha(ESV[,-c(6,7)])

```
Call: alpha(x = ESV[, -c(6, 7)])

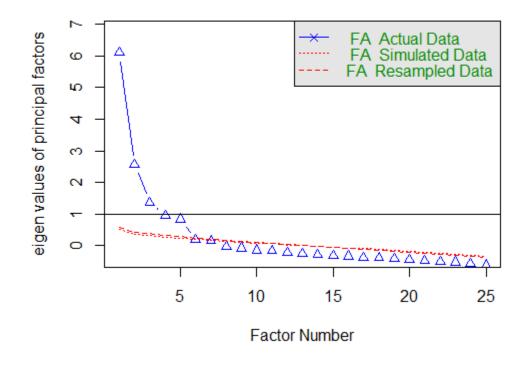
raw_alpha std.alpha G6(smc) average_r S/N ase mean sd
0.91 0.91 0.9 0.68 11 0.0065 4.7 1.5

lower alpha upper 95% confidence boundaries
```

MODELOS MULTIDIMENSIONAIS

$$X = b + a_1F_1 + a_1F_2 + a_2F_1 \dots + e$$

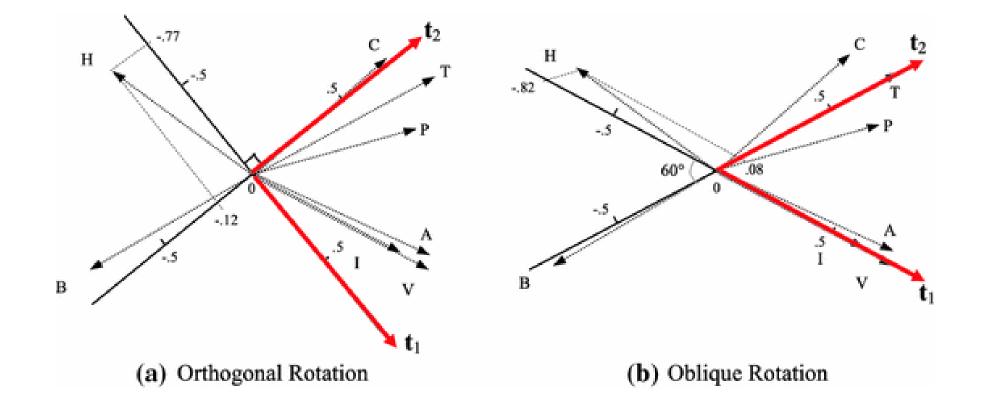
Parallel Analysis Scree Plots



ROTAÇÃO FATORIAL

Rotação fatorial

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.



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

 $Big5_fa < -fa(Big5[,-c(26,27)],5,cor="poly",fm="minrank",rotate = "oblimin",scores = "regression")$

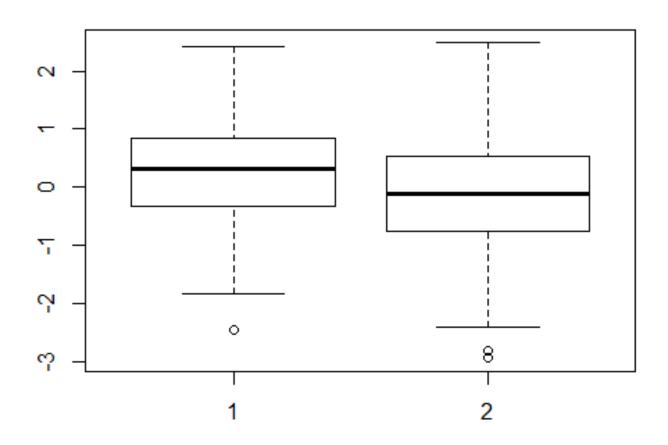
View(Big5_fa\$scores)

*	MRFA1 [‡]	MRFA2 [‡]	MRFA4 [‡]	MRFA5 [‡]	MRFA3 [‡]
1	1.743510658	-0.61406410	-1.38189079	0.75822935	0.68997423
2	NA	NA	NA	NA	NA
3	-0.689141688	-0.00883950	-0.97104036	-1.71044114	-0.09944628
4	-0.500946099	1.07885105	-1.02285449	-0.12124280	1.77448972
5	-1.762444583	0.55404260	-0.53821931	-1.43584878	0.30273618
6	-0.130637187	-0.74167365	-1.62212615	-2.90191335	0.33061680
7	NA	NA	NA	NA	NA
8	0.493125538	-1.04918353	1.25134447	-1.81605333	0.14035380
9	-0.229904915	2.75605764	0.59508649	-0.60651766	-0.17919788

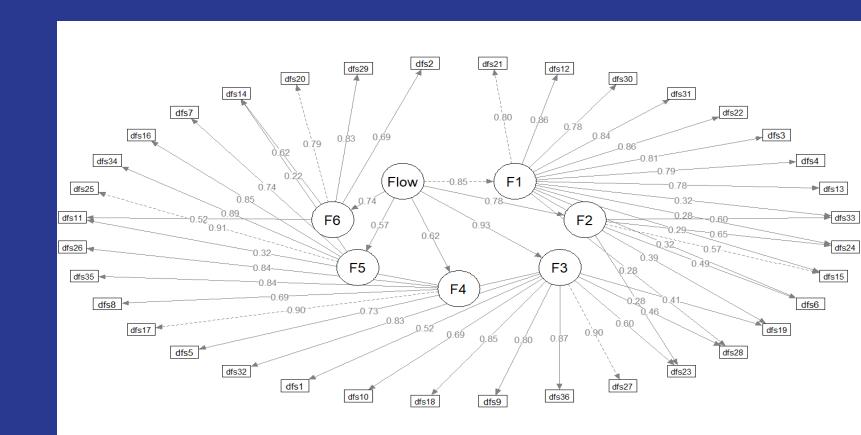
ESCORES FATORIAIS

t.test(Big5_fa\$scores[,5]~Big5\$Sexo)

boxplot(Big5_fa\$scores[,5]~Big5\$Sexo)



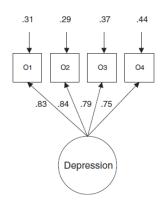
Análise Fatorial Confirmatória em Psicologia





Semelhanças e diferenças AFE & AFC

- •Ambos visam identificar fatores que explicam a variação e co-variação entre um conjunto de indicadores
- •AFE Data driven geralmente um procedimento descritivo ou exploratório
- •AFC Theory Driven pesquisador deve pré-especificar todos os aspectos do modelo latente: o número de fatores, o padrão de cargas entre indicador e fator, e assim por diante



Pacote LAVAAN

- Latent Variable Analysis
- Permite especificação do modelo O pesquisador indica qual item ou variável devem carregar em cada fator
- Permite fazer análises:
- o multi-grupo
- oinvestigação de índices de modificação
- Modelagem por equações regressão de fatores e mediação

Índices de Ajuste do Modelo

Resultado ESV

						•
HEAR	model	Varcue	haca	Ina	mode	
0361	illoue i	versus	vase	HIIC	IIIOuc	

Comparative Fit Index (CFI)
Tucker-Lewis Index (TLI)

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)
Loglikelihood unrestricted model (H1)

Number of free parameters Akaike (AIC) Bayesian (BIC) Sample-size adjusted Bayesian (BIC)

Root Mean Square Error of Approximation:

RMSEA
90 Percent Confidence Interval
P-value RMSEA <= 0.05

Standardized Root Mean Square Residual:

SRMR

0.	995
0.	990

-4054.623 -4047.836

> 10 8129.245 8171.311 8139.571

0.059 0.022 0.097 0.298

0.018

Valores de Referência

AIC (Akaike Information Criterion), o BIC (Bayesian Information Criterion) = Quanto menor melhor!

CFI (*Comparative Fit Index*) & TLI (*Tucker-Lewis Index*) = acima de 0,95 sugerem ajuste excelente, já valores acima de 0,90 indicam que a qualidade de ajuste é satisfatória

RMSEA (Root Mean Square Error of Approximation) e SRMR (Standardized Root Mean Square Residual) = menores que 0,05 indicam bom ajuste, enquanto valores menores que 0,08 indicam ajuste aceitável



Resultado ESV

Estimate 0.756

0.713

0.860

0.637

0.466

R-Square:

ESV1

ESV2

ESV3

ESV4

ESV5

Latent Variables: Std.Err z-value P(>|z|) Estimate Std.lv Std.all SV =~ 1.535 ESV1 1.000 0.869 ESV2 0.916 0.037 24.778 0.000 1.406 0.845 0.035 ESV3 1.029 29.316 0.000 1.579 0.928 ESV4 0.875 0.039 22.449 0.000 1.344 0.798 ESV5 0.911 0.052 17.611 0.000 1.398 0.683 Carga não-Significância da Carga padronizada carga fatorial padronizada

Variância

explicada do item

pelo modelo

Cargas fatoriais padronizadas e não padronizadas

Significância da carga no fator

Comunalidade

