

# Análise Fatorial

Análise fatorial exploratória e confirmatória

Victor Ivamoto

16 de dezembro de 2020

# Contents

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Metodologia</b>	<b>3</b>
<b>3</b>	<b>Preparação dos Dados</b>	<b>4</b>
3.1	Valores faltando . . . . .	4
3.2	Testes de normalidade . . . . .	5
3.3	Outliers . . . . .	6
3.4	Correlação . . . . .	10
3.5	Teste de Kaiser-Meyer-Olkin (KMO) . . . . .	11
3.6	Fator de inflação de variância (VIF) . . . . .	13
3.7	Teste de esfericidade de Bartlett . . . . .	13
<b>4</b>	<b>Análise Fatorial Exploratória (EFA)</b>	<b>15</b>
<b>5</b>	<b>Análise Fatorial Confirmatória (CFA)</b>	<b>21</b>
5.1	Modelagem por Equações Estruturais (SEM) . . . . .	24
5.2	Índices de adequação . . . . .	30
<b>6</b>	<b>Confiabilidade</b>	<b>32</b>
6.1	Validade discriminante . . . . .	33
6.2	Validade convergente e discriminante . . . . .	34
	<b>Apêndice</b>	<b>40</b>

# Chapter 1

## Introdução

Este trabalho apresenta um estudo de Análise Fatorial no questionário da Copa do Mundo de 2014 realizada no Brasil. O dataset original contém 3786 observações e 124 atributos. Alguns desses atributos são irrelevantes para a análise estatística dos dados, e outros podem ser agrupados em conjuntos com elementos semelhantes antes da análise estatística. O agrupamento é necessário para redução da dimensionalidade dos dados.

# Chapter 2

## Metodologia

O objetivo da Análise Fatorial Exploratória (EFA) é explicar as correlações entre um conjunto de variáveis observadas e um grupo menor de variáveis fundamentais não-observadas chamadas *fatores*. Supõe-se que cada fator explique a variância compartilhada entre duas ou mais variáveis observadas, portanto, tecnicamente, são chamados de fatores comuns.

O modelo pode ser representado como:

$$X_i = a_1F_1 + a_2F_2 + \dots + a_pF_p + U_i$$

onde  $X_i$  é a  $i$ -ésima variável observada ( $i = 1, \dots, k$ ),  $F_j$  são os fatores comuns ( $j = 1, \dots, p$ ) e  $p < k$ .  $U_i$  é a porção da variável  $X_i$  única a esta variável e que não é explicada pelos fatores comuns.  $a_i$  é o grau que cada fator contribui para a composição da variável observada.

# Chapter 3

## Preparação dos Dados

Os dados são preparados antes de realizar a análise fatorial. Essa etapa consiste em excluir as variáveis categóricas e dicotômicas, tratar os valores faltando e com baixo número de respostas. Todas as variáveis usadas neste estudo são ordinais.

```
# Carrega as bibliotecas
library(readxl)      # importa arquivo Excel
library(tidyverse)   # manipula as tabelas (datasets)
library(MVN)         # teste de normalidade multivariável
library(mctest)      # teste de multicolinearidade (VIF)
library(REdaS)       # teste de Bartlett
library(psych)       # realiza EFA
library(lavaan)      # realiza CFA
library(semPlot)     # para plotar CFA

# Importa as colunas 11 a 118 que possuem variáveis ordinais.
# As demais variáveis são categóricas, dicotômicas ou contínuas
# e não interesam para análise fatorial.
tbl <- readxl::read_excel('Base_Original.xls', sheet = 'CÓDIGO',
                          col_names = TRUE)[-54][11:117]

# Exclui variáveis dicotômicas
dic <- apply(tbl, 2, function(x) length(unique(x)))
tbl <- tbl[dic > 2]
```

### 3.1 Valores faltando

O questionário possui algumas colunas com muitas variáveis faltando, que podem ser excluídas sem prejuízo para análise. Vamos identificar essas colunas e removê-las.

```
# Colunas com valores faltando (NA)
nacols <- colnames(tbl[colMeans(is.na(tbl)) != 0])

# Remove colunas com valores faltando (NA)
tbl <- tbl %>% select(!nacols)
```

Além dos valores em branco removidos acima, as respostas da tabela abaixo não interessam na análise. Vamos substituir esses valores pela média dos valores na coluna correspondente, a fim de minimizar o impacto nos cálculos futuros.

Valor	Significado da resposta
96	Não acompanhou os jogos
97	Não respondeu
98	Outras respostas
99	Não sabe

```
# Substitui 96, 97, 98 e 99 pelo valor médio da coluna
for (x in 96:99){
  while (max(colMeans(tbl == x), na.rm = TRUE) > 0){
    # Substitui 96, 97, 98 e 99 por NA
    col <- which.max(colMeans(tbl == x))
    tbl[tbl[col] == x,col] = NA
    # Substitui NA pelo valor médio da coluna
    tbl <- data.frame(sapply(tbl,
      function(x) ifelse(is.na(x),
        mean(x, na.rm = TRUE), x)))
  }
}
```

O tamanho de amostra recomendado é de pelo menos 300 participantes, e as variáveis que são submetidas à análise fatorial devem ter cada uma pelo menos 5 a 10 observações (Comrey and Lee, 2013). Normalmente dizemos que a proporção de respondentes para variáveis deve ser de pelo menos 10: 1 e que os fatores são considerados estáveis e com validação cruzada com uma proporção de 30: 1. Um tamanho de amostra maior diminuirá o erro em seus dados e, portanto, o EFA geralmente funciona melhor com tamanhos de amostra maiores.

Os dados possuem 3786 observações e 99, ou seja a proporção de 38.2 satisfaz o critério

## 3.2 Testes de normalidade

A análise fatorial pressupõe que os dados possuam distribuição normal (Gaussiana) tanto uni variável quanto multivariável (Child, 2006), e a ausência de *outliers* (Field, 2009).

As distribuições das variáveis não influencia no uso da análise fatorial para descrever as relações entre as variáveis. Se as variáveis são normalmente distribuídas, a solução é aprimorada. Na medida em que a normalidade falha, a solução é degradada, mas ainda pode valer a pena (Tabachnick and Fidell, 2006).

A normalidade multivariada é assumida quando a inferência estatística é usada para determinar o número de fatores. A normalidade multivariada é a suposição de que todas as variáveis e todas as combinações lineares de variáveis são normalmente distribuídas. Embora os testes de normalidade multivariada sejam excessivamente sensíveis, a normalidade entre variáveis únicas é avaliada por assimetria e curtose. Se uma variável tem assimetria e curtose substanciais, a transformação da variável é considerada (Tabachnick and Fidell, 2006).

Em uma grande amostra, o nível de significância da assimetria não é tão importante quanto seu tamanho real (pior quanto mais distante de zero) e a aparência visual da distribuição. Em uma grande amostra, o impacto do afastamento da curtose zero também diminui. Por exemplo, as subestimativas da variância associadas à curtose positiva (distribuições com caudas curtas e grossas) desaparecem com amostras de 100 ou mais casos; com curtose negativa, a subestimação da variância desaparece com amostras de 200 ou mais (Waternaux, 1976).

Isso posto, a análise de normalidade não se faz necessário no presente estudo, dado que o número de amostras é substancialmente elevado.

### 3.3 Outliers

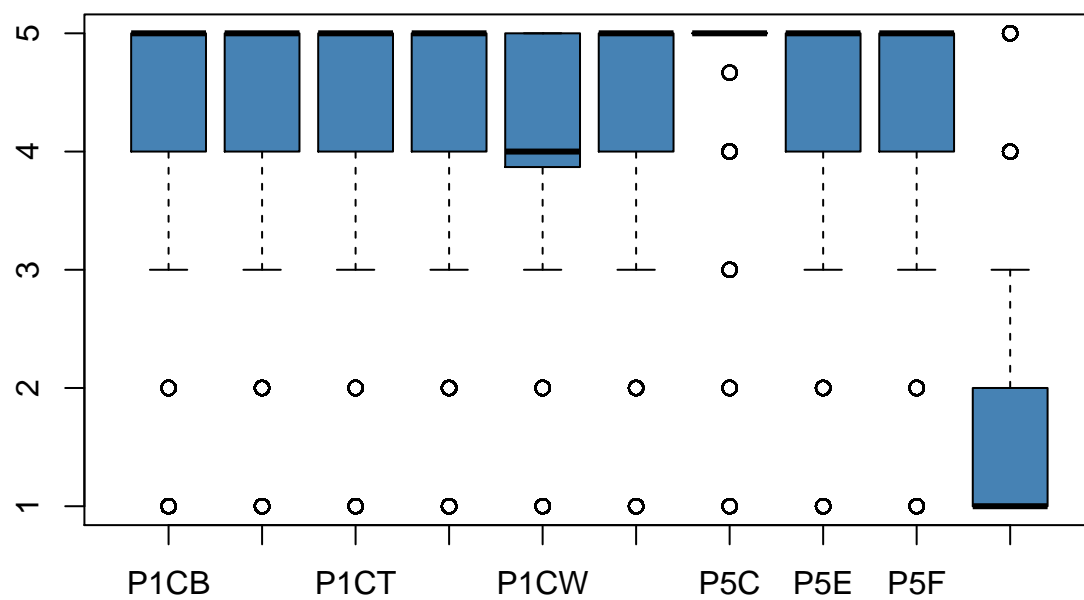
Existem diversas técnicas para identificação de outliers. Este estudo considerou os dados dentro do intervalo do box plot como observações típicas, e as amostras fora desse intervalo como observações atípicas ou *outliers*. Formalmente, o intervalo é dado pela seguinte equação:

$$I = [Q_{25} - 1,5IQR, Q_{75} + 1,5IQR]$$

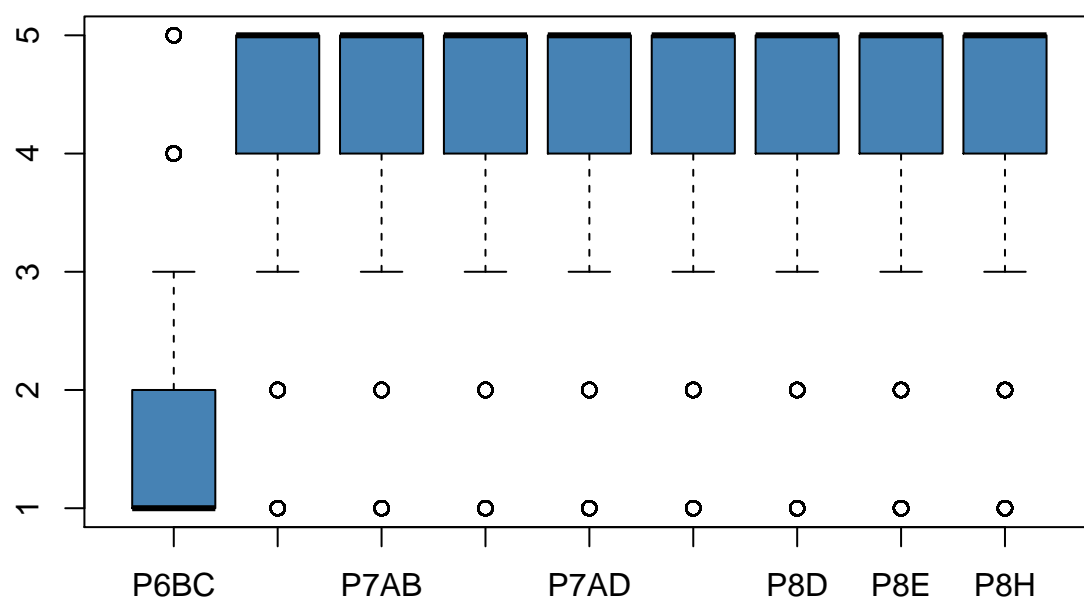
Onde  $Q_{25}$  e  $Q_{75}$  são os quartis de 25% e 75% respectivamente, e  $IQR$  é o intervalo interquartil dado por:

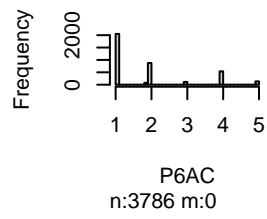
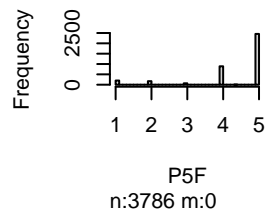
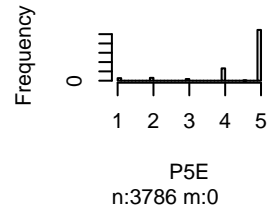
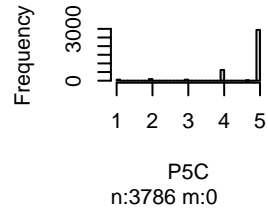
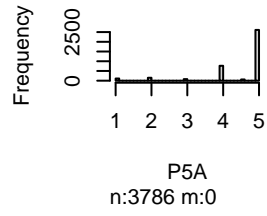
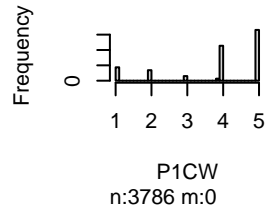
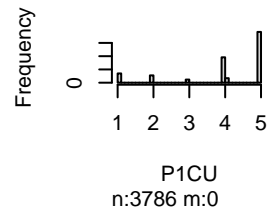
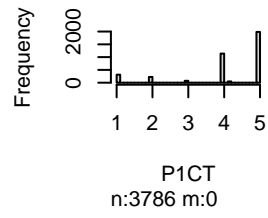
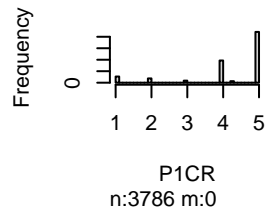
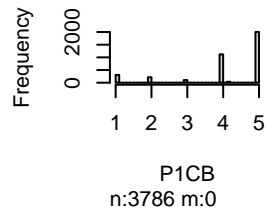
$$IQR = Q_{75} - Q_{25}$$

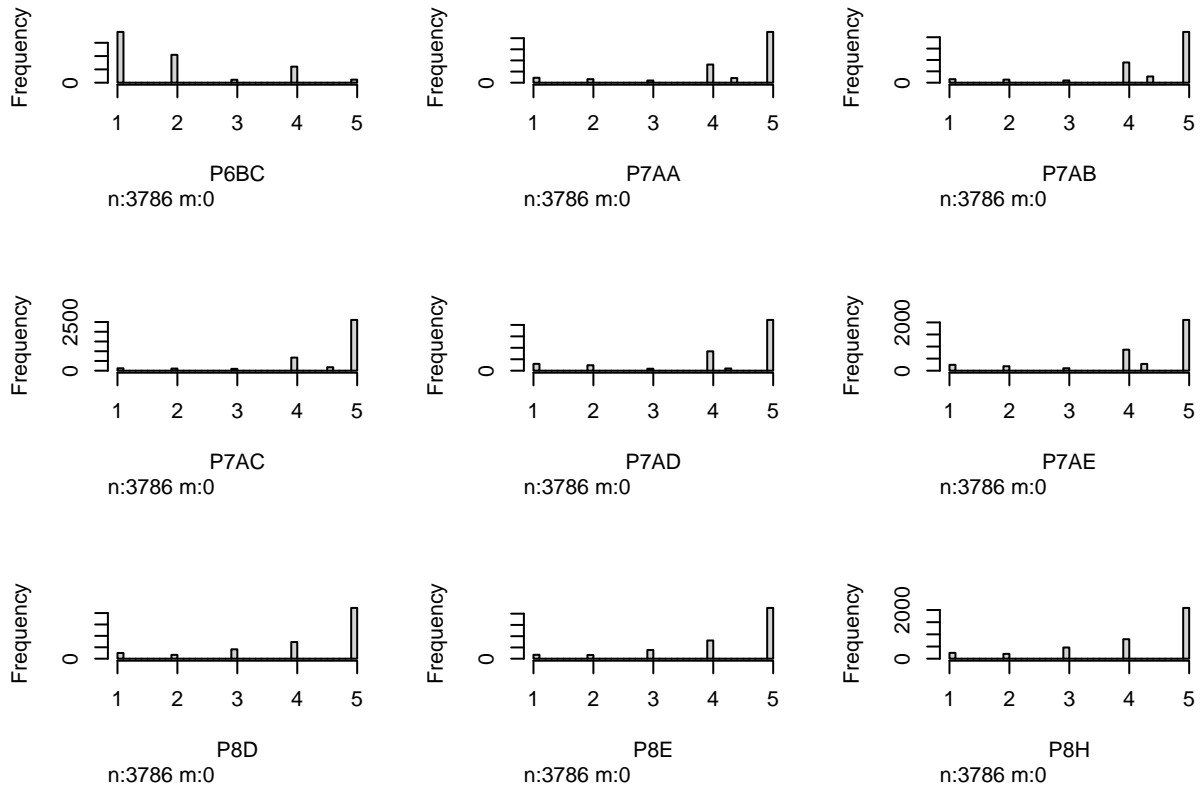
Algumas perguntas do questionário possuem concentração em uma ou duas respostas, causando uma aparente anomalia na distribuição dos dados. Neste caso, fica a critério do pesquisador remover essas variáveis atípicas. Os box plots abaixo mostram as variáveis com *outliers*, e os histogramas mostram as distriuições destas variáveis.







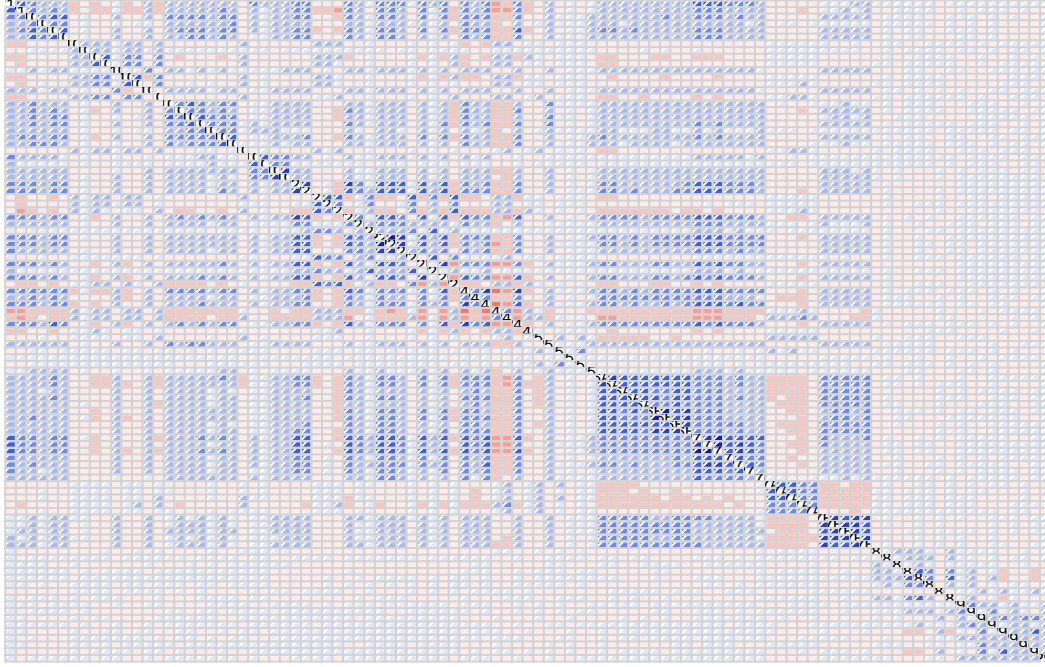




### 3.4 Correlação

A correlação  $r$  deve ser 0,30 ou maior, pois qualquer coisa menor sugeriria uma relação realmente fraca entre as variáveis (Tabachnick and Fidell, 2006). Os pontos claros no correlograma indicam baixa correlação, e os escuros indicam alta correlação. Os pontos azuis são correlação positiva, e os vermelhos são negativa.

```
## Registered S3 method overwritten by 'seriation':
##   method      from
##   reorder.hclust gclus
```



### 3.5 Teste de Kaiser-Meyer-Olkin (KMO)

O índice de KMO, também conhecido como índice de adequação da amostra, é um teste estatístico que sugere a proporção de variância dos itens que pode estar sendo explicada por uma variável latente (Lorenzo-Seva et al., 2011). Tal índice indica o quão adequada é a aplicação da EFA para o conjunto de dados (Hair et al., 2000). O KMO é calculado por meio do quadrado das correlações totais dividido pelo quadrado das correlações parciais, das variáveis analisadas (Field, 2009). Seu valor pode variar de zero a um. Valores iguais ou próximos a zero indicam que a soma das correlações parciais dos itens avaliados é bastante alta em relação à soma das correlações totais. Nesses casos, possivelmente a análise fatorial será inapropriada (Pasquali and Primi, 2003)<sup>1</sup>.

O índice KMO é dado pela seguinte equação:

$$KMO = \frac{\sum_{i=1}^m \sum_{j=1}^m r_{ij}^2}{\sum_{i=1}^m \sum_{j=1}^m r_{ij}^2 + \sum_{i=1}^m \sum_{j=1}^m a_{ij}^2}, \quad j \neq i$$

onde  $R$  é a matriz de correlação baseado nos dados  $X$ ,  $r_{ij}$  é a correlação entre as variáveis  $i$  e  $j$ , e  $a_{ij}$  é a correlação parcial.

<sup>1</sup>[http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712012000200007](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712012000200007)

A medida de adequação da amostra (MSA) é dada pela seguinte equação:

$$MSA_i = \frac{\sum_{j=1}^k r_{ij}^2}{\sum_{j=1}^k r_{ij}^2 + a_{ij}^2}, \quad j \neq i$$

Como referência, [Kaiser \(1970\)](#) sugere a seguinte regra de interpretação do resultado:

KMO	Referência
0,9 - 1,0	Maravilhoso
0,8 - 0,9	Meritório
0,7 - 0,8	Moderado
0,6 - 0,7	Medíocre
0,5 - 0,6	Miserável
menos de 0,5	Inaceitável

Para os testes KMO e MSA, fez-se uso da metodologia de Nakazawa proposta para o R como sugerida por Bates (2011).

```
# Cálculo do KMO e MSA
x <- subset(tbl, complete.cases(tbl))
r <- cor(x)      # correlação de x
r2 <- r^2
i <- solve(r)    # resolve equação a %*% x = b para x
d <- diag(i)     # diagonal da matriz i
p2 <- (-i/sqrt(outer(d,d)))^2
diag(r2) <- diag(p2)<-0
KMO <- sum(r2)/(sum(r2)+sum(p2))
MSA <- colSums(r2)/(colSums(r2)+colSums(p2))
```

O índice KMO em torno de 0.95 está maravilhoso para realizar a análise fatorial.

A medida de adequação da amostra pode ser interpretada com as seguintes orientações ([Hair et al., 2000](#)):

MSA	Interpretação
0,80 ou acima	admirável
0,70 ou acima	mediano
0,60 ou acima	medíocre
0,50 ou acima	ruim
abaixo de 0,50	inaceitável

A tabela seguinte mostra a distribuição da MSA de acordo com os critérios estabelecidos

anteriormente. Todas as variáveis estão com MSA acima de 0,50, considerado aceitável.

Table 3.4

Admirável	Mediano	Medíocre	Ruim	Inaceitável
87.9	12.1	0	0	0

### 3.6 Fator de inflação de variância (VIF)

O fator de inflação de variância (VIF) indica o efeito que as outras variáveis independentes têm sobre o erro padrão de um coeficiente de regressão. O fator de inflação de variância está diretamente relacionado ao valor de tolerância ( $VIF_i = 1/TOL_i$ ). Valores VIF altos também indicam um alto grau de colinearidade ou multicolinearidade entre as variáveis independentes (Hair et al., 2000).

A multicolinearidade cria variância “compartilhada” entre variáveis, diminuindo assim a capacidade de prever a medida dependente, bem como averiguar os papéis relativos de cada variável independente.

Verificamos a inexistência de multicolinearidade nos dados. Caso houvesse, removeríamos as variáveis com IMC maiores que 10.

```
# Verifica multicolinearidade
model <- lm('P1A~.', data = tbl)
imc <- imcdiag(model, method = "VIF", vif = 10)[[1]][,1]
# Nenhuma variável > 10 encontrada
mean(imc > 10)
```

```
## [1] 0
```

### 3.7 Teste de esfericidade de Bartlett

O teste de esfericidade de Bartlett avalia em que medida a matriz de covariância é similar a uma matriz identidade, ou seja, não apresentam correlações entre si. Segundo Hair et al. (2000), esse teste avalia, também, a significância geral de todas as correlações em uma matriz de dados. Valores do teste de esfericidade de Bartlett com níveis de significância p-valor < 0,05 indicam que a matriz é fatorável (Tabachnick and Fidell, 2006), rejeitando a hipótese nula de que a matriz de dados é similar a uma matriz-identidade (Damásio, 2012).

Se as variáveis estão perfeitamente correlacionadas, apenas um fator é suficiente. Se são ortogonais, então o número de fatores iguala o número de variáveis. Aumentar o tamanho da

amostra faz com que o teste Bartlett se torne mais sensível na detecção de correlações entre as variáveis ([Hair et al., 2000](#)).

O teste apresenta valor estatisticamente significativo, com p-valor próximo a zero, indicando que existem correlações suficientes entre as variáveis para se continuar a análise.

```
REdaS::bart_spher(tbl)
```

```
## Bartlett's Test of Sphericity
##
## Call: REdaS::bart_spher(x = tbl)
##
##      X2 = 138199.065
##      df = 4851
## p-value < 2.22e-16
```

## Chapter 4

# Análise Fatorial Exploratória (EFA)

No modelo matemático de “análise fatorial clássica”,  $p$  denota o número de variáveis ( $X_1, X_2, \dots, X_p$ ) e  $m$  denota o número de fatores subjacentes ( $F_1, F_2, \dots, F_m$ ).  $X_j$  é a variável representada em fatores latentes. Portanto, este modelo assume que existem  $m$  fatores subjacentes em que cada variável observada é uma função linear desses fatores juntamente com uma variável residual. Este modelo pretende reproduzir as correlações máximas.

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + \epsilon_j$$

Onde  $j = 1, 2, \dots, p$ . As cargas fatoriais são  $a_{j1}, a_{j2}, \dots, a_{jm}$ , o que denota que  $a_{j1}$  é a carga fatorial da  $j$ -ésima variável no primeiro fator (Yong and Pearce, 2013).

Uma decisão crítica em análise fatorial é o número de fatores a serem mantidos para interpretação e uso posterior. Uma base quantitativa exata para decidir o número de fatores a serem extraídos ainda não foi desenvolvida. Critérios de parada para o número de fatores a serem extraídos incluem a raiz latente ou autovalor, definição a priori, percentual de variância e teste *scree* (Hair et al., 2000).

Uma maneira de determinar o número de fatores ou componentes em uma matriz de dados ou uma matriz de correlação é examinar o gráfico *scree* dos autovalores sucessivos. Quebras bruscas no gráfico sugerem o número apropriado de componentes ou fatores a serem extraídos. A análise paralela é uma técnica alternativa que compara a sequência de fatores dos dados observados com a de uma matriz de dados aleatórios do mesmo tamanho da original. Isso pode ser feito para dados contínuos, dicotômicos ou politômicos usando correlações de Pearson, tetracóricas ou policóricas (Revelle, 2017).

Além de determinar o número de fatores, as variáveis com baixa comunalidade são excluídas do modelo. O tamanho da comunalidade é um índice útil para avaliar o quanto de variância em uma variável particular é explicado pela solução fatorial. Valores mais altos de comunalidade indicam que uma grande quantia da variância em uma variável foi extraída pela solução fatorial. Comunalidades pequenas mostram que uma porção substancial da variância da variável não é explicada pelos fatores. Apesar de nenhuma diretriz estatística indicar exatamente o que é “pequeno” ou “grande”, considerações práticas sugerem um nível mínimo



de 0,50 para comunalidades nesta análise (Hair et al., 2000). Para efeitos deste estudo, o valor mínimo de 0,40 foi adotado.

A comunalidade é a variância nas variáveis observadas que são contabilizadas por um fator comum ou variância comum (Child, 2006). A comunalidade é denotada por  $h^2$  e é a soma das correlações quadradas da variável com os fatores. A fórmula para derivar as comunalidades é  $h_j^2 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2$  onde  $a$  são as cargas para variáveis  $j$  (Yong and Pearce, 2013).

A exclusão começa com a variável com menor comunalidade, então a comunalidade de todas as variáveis é recalculado e a próxima variável é removida. O processo prossegue até que todas as variáveis possuam valor acima de 0,40. Em seguida, é realizada a análise fatorial.

```
nfactors <- psych::fa.parallel(tbl, plot = FALSE)
nfact <- nfactors$nfact      # Número de fatores
corMat <- cor(tbl)           # Matriz de correlação
N <- nrow(tbl)               # Número de observações
faPC <- fa(r = corMat, nfactors = nfact, n.obs = N, rotate = "varimax")

# Remove as variáveis com baixa comunalidades (coluna h2 < 0,40).
h2 <- faPC$communality      # Comunalidade (coluna h2)
exclui <- mean(h2 < 0.40) > 0 # TRUE se houver h2 < 0,40
while(exclui){
  tbl <- tbl[!(h2 == min(h2))] # Exclui variável com menor h2
  nfactors <- fa.parallel(tbl, plot = FALSE) # Recalcula nfactors
  nfact <- nfactors$nfact      # Número de fatores
  N <- nrow(tbl)              # Número de linhas na nova tabela
  corMat <- cor(tbl)           # Recalcula a correlação
  faPC <- fa(r = corMat, nfactors = nfact, n.obs = N, rotate = "varimax")
  h2 <- faPC$communality      # Coluna h2
  exclui <- mean(h2 < 0.40) > 0 # Continua enquanto h2 < 0,40
}
```

Após remover as variáveis com baixa comunalidade ( $h^2 < 0,40$ ), prosseguimos para a análise fatorial.

```
# Análise fatorial
fc <- factanal(tbl, factors=nfact, scores="regression")
fa(r = corMat, nfactors = nfact, n.obs = N, rotate = "varimax")

## Factor Analysis using method = minres
## Call: fa(r = corMat, nfactors = nfact, n.obs = N, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR6  MR11  MR2  MR1  MR3  MR12  MR5  MR7  MR4  MR9  MR8  MR10
## P1CG -0.05 -0.07  0.02 -0.08  0.11 -0.12  0.02 -0.07 -0.05  0.78  0.00 -0.06
## P1CH -0.05 -0.09 -0.06 -0.03  0.11 -0.06  0.09  0.00 -0.02  0.62  0.02 -0.01
```

##	P1CP	0.16	0.16	0.06	0.09	-0.09	0.15	0.02	0.10	0.04	-0.02	0.00	0.56
##	P1CQ	0.16	0.14	0.09	0.13	-0.08	0.13	-0.03	0.08	0.05	-0.06	0.03	0.64
##	P1CX	0.13	0.15	0.09	0.11	-0.05	0.07	-0.01	0.83	0.02	-0.06	-0.01	0.08
##	P1CY	0.13	0.11	0.11	0.13	-0.06	0.05	-0.03	0.72	0.10	-0.02	-0.02	0.11
##	P2A	0.60	0.17	0.10	0.29	-0.04	0.08	-0.06	0.09	0.03	-0.04	0.00	0.07
##	P2B	0.61	0.17	0.12	0.28	-0.03	0.11	-0.05	0.08	0.02	-0.03	-0.01	0.07
##	P2C	-0.06	-0.07	-0.01	0.01	0.74	-0.08	0.08	0.00	-0.03	0.09	-0.04	-0.03
##	P2E	-0.10	-0.11	-0.05	-0.06	0.66	-0.10	0.07	-0.05	-0.01	0.06	-0.02	-0.03
##	P2F	0.58	0.17	0.14	0.26	-0.15	0.14	-0.06	0.07	0.05	-0.05	0.01	0.11
##	P2I	0.79	0.15	0.10	0.18	-0.08	0.14	-0.04	0.06	0.07	-0.04	0.00	0.06
##	P2J	0.82	0.15	0.09	0.16	-0.09	0.13	-0.02	0.05	0.06	-0.05	0.01	0.05
##	P2K	0.63	0.12	0.08	0.17	0.05	0.08	-0.05	0.06	0.02	0.01	-0.01	0.07
##	P2M	0.57	0.12	0.09	0.20	-0.28	0.16	-0.01	0.06	0.11	-0.06	-0.02	0.12
##	P2O	0.56	0.13	0.08	0.20	-0.29	0.18	-0.01	0.07	0.10	-0.07	-0.03	0.16
##	P2P	-0.13	-0.08	-0.03	-0.04	0.66	-0.05	0.05	-0.05	-0.04	0.08	0.00	-0.08
##	P4A	0.28	0.16	0.13	0.28	-0.12	0.63	-0.04	0.07	0.08	-0.07	0.04	0.16
##	P4C	0.26	0.17	0.13	0.35	-0.04	0.57	-0.04	0.06	0.04	-0.06	0.02	0.12
##	P4D	-0.16	-0.11	-0.04	-0.15	0.19	-0.53	0.01	-0.03	-0.04	0.15	-0.04	-0.09
##	P4F	0.30	0.21	0.14	0.23	-0.13	0.49	-0.05	0.07	0.07	-0.06	0.03	0.15
##	P6AA	0.14	0.70	0.20	0.11	-0.12	0.15	-0.07	0.08	0.07	-0.08	0.02	0.04
##	P6AB	0.15	0.69	0.17	0.10	-0.11	0.16	-0.07	0.10	0.10	-0.11	0.01	0.03
##	P6AC	0.14	0.66	0.21	0.12	-0.06	0.04	-0.14	0.03	0.03	-0.02	-0.05	0.10
##	P6AD	0.16	0.63	0.19	0.18	-0.05	0.10	-0.08	0.07	0.06	-0.04	-0.02	0.08
##	P6AE	0.13	0.63	0.20	0.13	-0.07	0.05	-0.10	0.05	0.03	-0.04	-0.02	0.11
##	P6BA	0.14	0.48	0.15	0.13	-0.07	0.10	-0.03	0.10	0.66	-0.06	0.02	0.07
##	P6BB	0.14	0.47	0.16	0.13	-0.09	0.10	-0.03	0.10	0.66	-0.07	0.02	0.06
##	P6BC	0.14	0.55	0.20	0.19	-0.03	0.03	-0.11	0.04	0.31	-0.01	-0.02	0.08
##	P6BD	0.16	0.51	0.18	0.21	-0.05	0.11	-0.06	0.08	0.39	-0.01	-0.03	0.08
##	P7A	0.33	0.18	0.17	0.64	-0.07	0.25	-0.01	0.08	0.07	-0.06	0.01	0.09
##	P7B	0.31	0.19	0.12	0.75	-0.03	0.17	-0.05	0.06	0.05	-0.05	0.00	0.04
##	P7C	0.32	0.19	0.12	0.75	-0.07	0.21	-0.02	0.08	0.06	-0.05	0.01	0.07
##	P7D	0.22	0.17	0.11	0.55	-0.03	0.04	-0.07	0.09	0.06	-0.03	-0.04	0.09
##	P7E	0.25	0.17	0.16	0.58	0.00	0.17	-0.02	0.06	0.06	-0.03	-0.02	0.06
##	P7AA	-0.03	-0.08	-0.11	-0.02	0.07	-0.05	0.62	-0.02	0.00	0.02	0.00	0.01
##	P7AB	-0.04	-0.09	-0.10	-0.02	0.05	-0.01	0.68	0.00	-0.05	0.05	0.03	0.00
##	P7AC	-0.06	-0.16	-0.14	-0.06	0.06	-0.01	0.64	-0.01	0.00	0.05	0.03	-0.02
##	P7BA	0.11	0.21	0.75	0.13	-0.03	0.04	-0.11	0.02	0.02	0.00	0.01	0.02
##	P7BB	0.07	0.19	0.72	0.08	-0.03	0.07	-0.08	0.07	0.08	-0.02	-0.01	0.04
##	P7BC	0.12	0.18	0.66	0.09	-0.02	0.08	-0.08	0.07	0.10	-0.02	0.02	0.05
##	P7BD	0.09	0.19	0.77	0.10	-0.03	0.06	-0.14	0.02	0.02	-0.01	0.00	0.04
##	P7BE	0.11	0.21	0.70	0.13	-0.02	0.05	-0.09	0.07	0.05	-0.02	0.00	0.06
##	P8D	-0.02	-0.03	-0.01	-0.03	-0.06	0.04	0.00	-0.02	0.02	-0.01	0.71	0.02
##	P8E	0.01	-0.01	0.02	0.00	0.01	0.02	0.05	0.00	-0.01	0.02	0.70	-0.01
##	MR13	h2	u2	com									
##	P1CG	-0.01	0.65	0.35	1.2								

```

## P1CH  0.00 0.43 0.57 1.2
## P1CP  0.00 0.42 0.58 1.7
## P1CQ  0.02 0.52 0.48 1.6
## P1CX  0.01 0.77 0.23 1.2
## P1CY  0.01 0.61 0.39 1.4
## P2A   0.30 0.60 0.40 2.5
## P2B   0.23 0.58 0.42 2.3
## P2C   0.04 0.58 0.42 1.1
## P2E   0.01 0.49 0.51 1.2
## P2F   0.19 0.56 0.44 2.5
## P2I  -0.22 0.77 0.23 1.5
## P2J  -0.30 0.85 0.15 1.6
## P2K  -0.12 0.48 0.52 1.5
## P2M   0.19 0.57 0.43 2.7
## P2O   0.20 0.59 0.41 3.0
## P2P  -0.05 0.48 0.52 1.2
## P4A   0.07 0.66 0.34 2.5
## P4C  -0.03 0.59 0.41 2.8
## P4D   0.04 0.41 0.59 2.1
## P4F   0.07 0.51 0.49 3.6
## P6AA  0.05 0.62 0.38 1.6
## P6AB  0.07 0.61 0.39 1.7
## P6AC  0.02 0.55 0.45 1.6
## P6AD -0.03 0.53 0.47 1.7
## P6AE -0.03 0.51 0.49 1.6
## P6BA  0.01 0.76 0.24 2.4
## P6BB  0.02 0.75 0.25 2.4
## P6BC -0.02 0.52 0.48 2.6
## P6BD -0.03 0.55 0.45 3.1
## P7A  -0.01 0.67 0.33 2.4
## P7B  -0.01 0.75 0.25 1.7
## P7C   0.01 0.78 0.22 1.9
## P7D   0.02 0.42 0.58 1.9
## P7E   0.02 0.49 0.51 2.0
## P7AA -0.01 0.42 0.58 1.1
## P7AB  0.01 0.49 0.51 1.1
## P7AC -0.01 0.47 0.53 1.3
## P7BA  0.05 0.65 0.35 1.3
## P7BB -0.01 0.58 0.42 1.3
## P7BC -0.01 0.52 0.48 1.4
## P7BD  0.01 0.67 0.33 1.3
## P7BE -0.01 0.58 0.42 1.4
## P8D   0.00 0.52 0.48 1.0
## P8E  -0.01 0.49 0.51 1.0
##

```

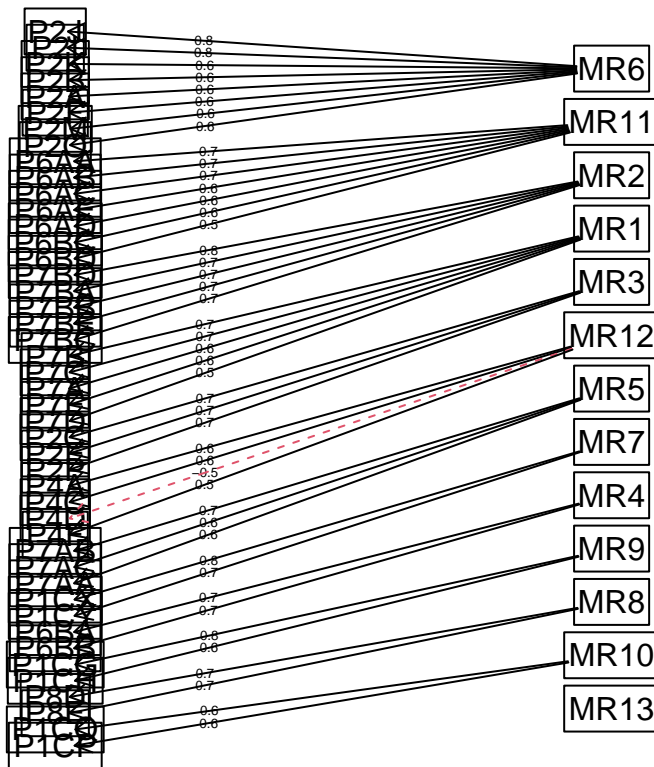
```

##          MR6 MR11  MR2  MR1  MR3 MR12  MR5  MR7  MR4  MR9  MR8
## SS loadings          4.42 4.02 3.21 3.17 1.82 1.74 1.43 1.40 1.25 1.11 1.02
## Proportion Var       0.10 0.09 0.07 0.07 0.04 0.04 0.03 0.03 0.03 0.02 0.02
## Cumulative Var       0.10 0.19 0.26 0.33 0.37 0.41 0.44 0.47 0.50 0.52 0.55
## Proportion Explained 0.17 0.15 0.12 0.12 0.07 0.07 0.06 0.05 0.05 0.04 0.04
## Cumulative Proportion 0.17 0.32 0.45 0.57 0.64 0.71 0.76 0.81 0.86 0.91 0.94
##          MR10 MR13
## SS loadings          0.99 0.45
## Proportion Var       0.02 0.01
## Cumulative Var       0.57 0.58
## Proportion Explained 0.04 0.02
## Cumulative Proportion 0.98 1.00
##
## Mean item complexity = 1.8
## Test of the hypothesis that 13 factors are sufficient.
##
## The degrees of freedom for the null model are 990 and the objective function was 2
## The degrees of freedom for the model are 483 and the objective function was 0.66
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.02
##
## The harmonic number of observations is 3786 with the empirical chi square 963.47 w
## The total number of observations was 3786 with Likelihood Chi Square = 2489.86 wi
##
## Tucker Lewis Index of factoring reliability = 0.95
## RMSEA index = 0.033 and the 90 % confidence intervals are 0.032 0.034
## BIC = -1489.61
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
##          MR6 MR11  MR2  MR1  MR3 MR12
## Correlation of (regression) scores with factors 0.93 0.89 0.91 0.90 0.86 0.81
## Multiple R square of scores with factors        0.87 0.79 0.84 0.81 0.75 0.65
## Minimum correlation of possible factor scores    0.75 0.59 0.67 0.61 0.49 0.31
##
##          MR5  MR7  MR4  MR9  MR8 MR10
## Correlation of (regression) scores with factors 0.83 0.89 0.85 0.83 0.82 0.75
## Multiple R square of scores with factors        0.69 0.80 0.72 0.69 0.67 0.56
## Minimum correlation of possible factor scores    0.37 0.60 0.44 0.39 0.34 0.12
##
##          MR13
## Correlation of (regression) scores with factors 0.76
## Multiple R square of scores with factors        0.58
## Minimum correlation of possible factor scores    0.16

```

Diagrama da carga fatorial. Identifica os fatores com as variáveis correspondentes.

## Factor Analysis



# Chapter 5

## Análise Fatorial Confirmatória (CFA)

Modelagem de equações estruturais (SEM) é um método confirmatório que fornece um meio abrangente para validar o modelo de medição de construtos latentes. O procedimento de validação é denominado Análise Fatorial Confirmatória (CFA). O método CFA tem a capacidade de avaliar a Unidimensionalidade, Validade e Confiabilidade de um construto latente. O pesquisador precisa realizar CFA para todos os construtos latentes envolvidos no estudo antes de modelar sua inter-relação em um modelo estrutural (SEM). No entanto, a avaliação da unidimensionalidade deve ser feita antes de avaliar a validade e confiabilidade.

Com o CFA, qualquer item que não se encaixe no modelo de medição devido à carga fatorial baixa deve ser removido do modelo. A adequação de um modelo de medição é indicada por meio de determinados índices de adequação. No entanto, a exclusão de itens não deve exceder 20% do total de itens em um modelo. Caso contrário, o próprio construto particular é considerado inválido, uma vez que falhou na própria “confirmação”. Os pesquisadores poderiam executar o CFA para cada modelo de medição separadamente ou executar os modelos de medição agrupados de uma vez. No entanto, o CFA para modelos de medição agrupados é mais eficiente e altamente sugerido ([Zainudin, 2012](#)).

Cada variável é associada ao fator com maior carga fatorial, evitando assim carga cruzada e simplificando a análise. Executamos a análise fatorial confirmatória para esse modelo e analisamos os índices.

Para que algo seja rotulado como um fator, ele deve ter pelo menos três variáveis, embora isso dependa do desenho do estudo ([Tabachnick and Fidell, 2006](#)). Como regra geral, os fatores que têm duas ou menos variáveis devem ser interpretados com cautela. Um fator com duas variáveis só é considerado confiável quando as variáveis são altamente correlacionadas entre si ( $r > 0,70$ ), mas razoavelmente não correlacionadas com outras variáveis ([Yong and Pearce, 2013](#)). A tabela abaixo resume a correlação das variáveis dos fatores com 2 variáveis.

Os fatores MR9, MR10, MR7, MR8 possuem correlação abaixo de 0,70 e podem ser retirados do modelo. As variáveis do fator MR4 possuem correlação entre si acima de 0,70 e alta correlação com variáveis fora do fator (máxima correlação de 0.603 e 0.579), portanto esse fator também pode ser removido do modelo. O novo modelo teórico está na tabela 5.3:

Qualquer item com carga fatorial menor que 0,6 e  $R^2$  menor que 0,4 deve ser excluído do

Table 5.1

<b>Modelo inicial.</b>
MR9 =~ P1CG + P1CH
MR10 =~ P1CP + P1CQ
MR7 =~ P1CX + P1CY
MR6 =~ P2A + P2B + P2F + P2I + P2J + P2K + P2M + P2O
MR3 =~ P2C + P2E + P2P
MR12 =~ P4A + P4C + P4D + P4F
MR11 =~ P6AA + P6AB + P6AC + P6AD + P6AE + P6BC + P6BD
MR4 =~ P6BA + P6BB
MR1 =~ P7A + P7B + P7C + P7D + P7E
MR5 =~ P7AA + P7AB + P7AC
MR2 =~ P7BA + P7BB + P7BC + P7BD + P7BE
MR8 =~ P8D + P8E

modelo de medição. No entanto, o pesquisador pode não fazê-lo se os índices de adequação para aquele modelo de medição já atingiram o nível exigido conforme mostrado na Tabela 1 (Categoria do Índice e nível de aceitação). Um item com carga fatorial baixa significa simplesmente que um item específico é considerado inútil para medir aquele construto específico. Manter um item inútil em um modelo afetará o índice de adequação do modelo. Um item pode ter carga fatorial baixa por vários motivos. Entre as razões estão declarações ambíguas, declaração de duplo sentido, declaração sensível, declaração tendenciosa, etc (Zainudin, 2012).

O modelo apenas com variáveis com cargas fatoriais acima de 0,60 está na tabela 5.4

Identificamos seis fatores capazes de agregar as variáveis mais significativas.

Diagrama de trajetória para modelos SEM.

```
# Calcula os índices de fitness do CFA
fit <- lavaan::cfa mdl, data= tbl)

# Diagrama
semPlot::semPaths(fit, 'par', edge.label.cex=1.2, fade=FALSE)
```

Table 5.2

Fator	Variáveis	Correlação
MR9	P1CG e P1CH	0.518
MR10	P1CP e P1CQ	0.469
MR7	P1CX e P1CY	0.68
MR4	P6BA e P6BB	0.771
MR8	P8D e P8E	0.498

Table 5.3

Modelo apenas com fatores com mais de duas variáveis.
MR6 $\sim$ P2A + P2B + P2F + P2I + P2J + P2K + P2M + P2O
MR3 $\sim$ P2C + P2E + P2P
MR12 $\sim$ P4A + P4C + P4D + P4F
MR11 $\sim$ P6AA + P6AB + P6AC + P6AD + P6AE + P6BC + P6BD
MR1 $\sim$ P7A + P7B + P7C + P7D + P7E
MR5 $\sim$ P7AA + P7AB + P7AC
MR2 $\sim$ P7BA + P7BB + P7BC + P7BD + P7BE

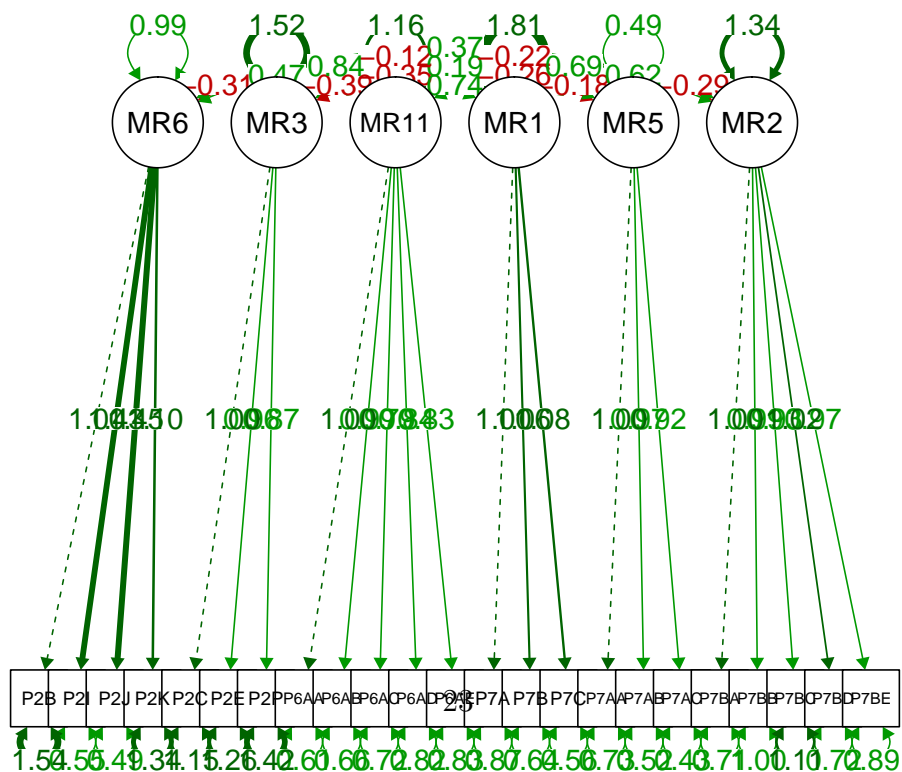




Table 5.4

<b>Modelo com carga fatorial &gt; 0,60</b>
MR6 =~ P2B + P2I + P2J + P2K
MR3 =~ P2C + P2E + P2P
MR11 =~ P6AA + P6AB + P6AC + P6AD + P6AE
MR1 =~ P7A + P7B + P7C
MR5 =~ P7AA + P7AB + P7AC
MR2 =~ P7BA + P7BB + P7BC + P7BD + P7BE

## 5.1 Modelagem por Equações Estruturais (SEM)

Após a análise fatorial confirmatória (CFA), realizamos a modelagem por equações estruturais. Os fatores são nomeados de acordo com as perguntas do questionário. Observe que algumas perguntas possuem sentimento positivo, como por exemplo “Este evento significou muito para mim”, enquanto que outras perguntas possuem sentimento negativo, como em “me senti aborrecido”. A coerência nas respostas a estas perguntas reflete na correlação negativa entre os fatores MR5 e MR2, MR5 e MR11, e MR1 e MR3.

- **MR1 - Identificação com o evento**

- P7A: Este evento significou muito para mim
- P7B: Estive muito apegado(a) à Copa do Mundo de 2014
- P7C: Identifiquei-me muito com este evento

- **MR2 - Transparência do governo**

- P7BA: As obras da Copa do Mundo no Brasil foram EXECUTADAS de forma transparente
- P7BB: Todo o processo da Copa do Mundo no Brasil foi DIVULGADO pelo Governo e pela mídia de forma clara
- P7BC: Os moradores locais puderam ver e acompanhar claramente o progresso e a situação das obras da Copa do Mundo no Brasil
- P7BD: As obras da Copa do Mundo no Brasil foram REALIZADAS com total transparência
- P7BE: O Governo divulgou informações suficientes aos moradores locais sobre o andamento das obras da Copa do Mundo no Brasil

- **MR3 - Insatisfação**

- P2C: Me senti Aborrecido
- P2E: Me senti Insatisfeito

– P2P: Eu me senti Descontente

- **MR5 - Incompetência e corrupção**

- P7AA: Houve irregularidades na construção das obras da Copa do Mundo no Brasil
- P7AB: Houve empresas que não cumpriram os padrões exigidos nos projetos da Copa do Mundo no Brasil
- P7AC: Há uma sensação de que houve desvio de fundos públicos e superfaturamento nas obras da Copa no Brasil

- **MR6 - Compaixão**

- P2B: Me senti Inspirado
- P2I: Eu senti Amor
- P2J: Eu senti Carinho
- P2K: Eu senti Compaixão

- **MR11 - Confiança no governo**

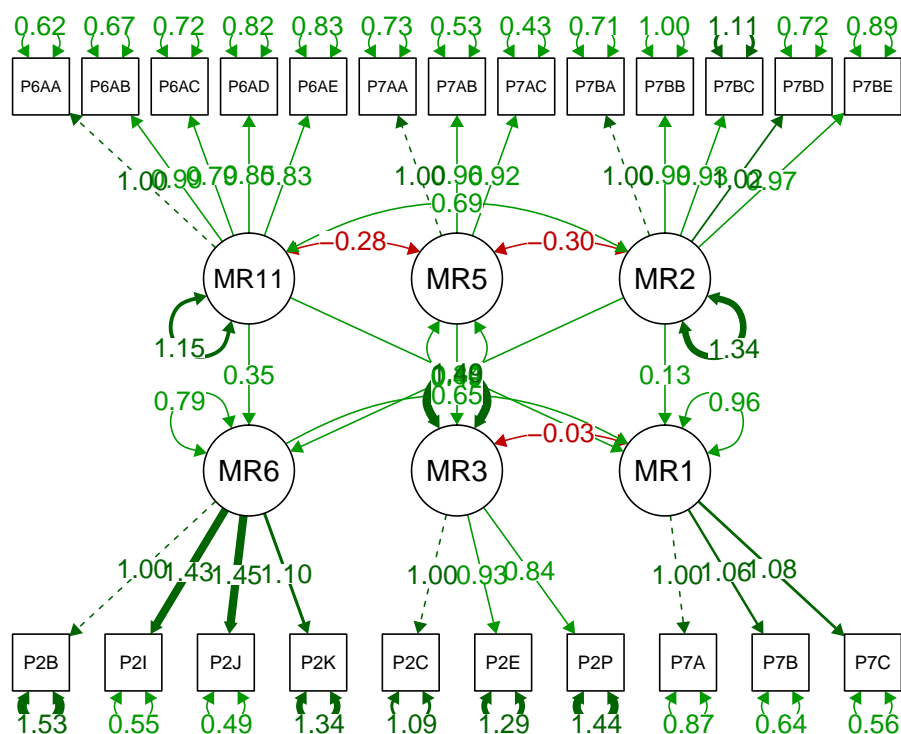
- P6AA: O Governo fez o que é correto para o bom desenvolvimento da Copa do Mundo
- P6AB: O Governo fez o que é correto para a boa organização da Copa do Mundo
- P6AC: O Governo agiu de acordo com os interesses dos moradores da cidade onde você mora
- P6AD: Eu confiei nas decisões tomadas pelo Governo sobre a Copa
- P6AE: O Governo se esforçou em incorporar os moradores da cidade no planejamento da Copa

O modelo teórico proposto

```
# Calcula os índices de fitness do CFA
fit <- lavaan::cfa mdl_sem, data= tbl)
# Diagrama
semPlot::semPaths(fit, 'par', edge.label.cex=1.0, fade=FALSE)
```

Table 5.5

Modelo final
MR3 $\sim$ MR5
MR1 $\sim$ MR6 + MR2 + MR11
MR6 $\sim$ MR2 + MR11
MR6 $\sim$ P2B + P2I + P2J + P2K
MR3 $\sim$ P2C + P2E + P2P
MR11 $\sim$ P6AA + P6AB + P6AC + P6AD + P6AE
MR1 $\sim$ P7A + P7B + P7C
MR5 $\sim$ P7AA + P7AB + P7AC
MR2 $\sim$ P7BA + P7BB + P7BC + P7BD + P7BE



```
## lavaan 0.6-7 ended normally after 44 iterations
##
```

```

##      Estimator                                ML
##      Optimization method                      NLMINB
##      Number of free parameters                56
##
##      Number of observations                    3786
##
## Model Test User Model:
##
##      Test statistic                            1650.291
##      Degrees of freedom                       220
##      P-value (Chi-square)                     0.000
##
## Model Test Baseline Model:
##
##      Test statistic                            42227.298
##      Degrees of freedom                       253
##      P-value                                  0.000
##
## User Model versus Baseline Model:
##
##      Comparative Fit Index (CFI)               0.966
##      Tucker-Lewis Index (TLI)                 0.961
##
## Loglikelihood and Information Criteria:
##
##      Loglikelihood user model (H0)             -133563.222
##      Loglikelihood unrestricted model (H1)      -132738.077
##
##      Akaike (AIC)                             267238.445
##      Bayesian (BIC)                            267587.833
##      Sample-size adjusted Bayesian (BIC)       267409.891
##
## Root Mean Square Error of Approximation:
##
##      RMSEA                                    0.041
##      90 Percent confidence interval - lower    0.040
##      90 Percent confidence interval - upper    0.043
##      P-value RMSEA <= 0.05                    1.000
##
## Standardized Root Mean Square Residual:
##
##      SRMR                                    0.052
##
## Parameter Estimates:
##

```

```

##      Standard errors
##      Information
##      Information saturated (h1) model
##      Standard
##      Expected
##      Structured
##
## Latent Variables:
##      Estimate   Std.Err   z-value   P(>|z|)
##
##      MR6 =~
##      P2B          1.000
##      P2I          1.433    0.034    42.599    0.000
##      P2J          1.452    0.034    42.859    0.000
##      P2K          1.100    0.031    35.653    0.000
##      MR3 =~
##      P2C          1.000
##      P2E          0.926    0.029    32.286    0.000
##      P2P          0.841    0.027    31.682    0.000
##      MR11 =~
##      P6AA          1.000
##      P6AB          0.993    0.019    51.518    0.000
##      P6AC          0.794    0.018    45.053    0.000
##      P6AD          0.846    0.019    45.086    0.000
##      P6AE          0.829    0.019    44.241    0.000
##      MR1 =~
##      P7A          1.000
##      P7B          1.058    0.017    62.343    0.000
##      P7C          1.081    0.017    63.609    0.000
##      MR5 =~
##      P7AA          1.000
##      P7AB          0.962    0.034    27.896    0.000
##      P7AC          0.924    0.033    28.016    0.000
##      MR2 =~
##      P7BA          1.000
##      P7BB          0.992    0.020    49.423    0.000
##      P7BC          0.934    0.020    46.300    0.000
##      P7BD          1.017    0.019    54.041    0.000
##      P7BE          0.971    0.019    50.326    0.000
##
## Regressions:
##      Estimate   Std.Err   z-value   P(>|z|)
##
##      MR3 ~
##      MR5          0.445    0.041    10.965    0.000
##      MR1 ~
##      MR6          0.651    0.027    24.460    0.000
##      MR2          0.130    0.021     6.156    0.000
##      MR11         0.296    0.025    11.911    0.000
##      MR6 ~

```

```

##      MR2              0.097      0.018      5.293      0.000
##      MR11             0.352      0.021     16.431      0.000
##
## Covariances:
##              Estimate   Std.Err   z-value   P(>|z|)
##      MR11 ~~
##      MR5             -0.276      0.018    -15.748      0.000
##      MR2              0.689      0.028     24.700      0.000
##      MR5 ~~
##      MR2             -0.297      0.019    -15.818      0.000
##      .MR3 ~~
##      .MR1            -0.029      0.025     -1.133      0.257
##
## Variances:
##              Estimate   Std.Err   z-value   P(>|z|)
##      .P2B             1.535      0.038     40.669      0.000
##      .P2I             0.550      0.022     25.010      0.000
##      .P2J             0.491      0.022     22.810      0.000
##      .P2K             1.335      0.034     39.527      0.000
##      .P2C             1.090      0.049     22.422      0.000
##      .P2E             1.290      0.047     27.711      0.000
##      .P2P             1.444      0.045     32.377      0.000
##      .P6AA            0.617      0.019     31.984      0.000
##      .P6AB            0.665      0.020     32.973      0.000
##      .P6AC            0.719      0.019     37.337      0.000
##      .P6AD            0.815      0.022     37.322      0.000
##      .P6AE            0.833      0.022     37.707      0.000
##      .P7A            0.869      0.026     33.677      0.000
##      .P7B            0.637      0.023     27.961      0.000
##      .P7C            0.558      0.022     25.251      0.000
##      .P7AA           0.731      0.023     32.127      0.000
##      .P7AB           0.528      0.018     28.717      0.000
##      .P7AC           0.433      0.016     26.943      0.000
##      .P7BA           0.710      0.021     33.063      0.000
##      .P7BB           0.998      0.028     36.232      0.000
##      .P7BC           1.113      0.029     37.757      0.000
##      .P7BD           0.718      0.022     32.830      0.000
##      .P7BE           0.891      0.025     35.699      0.000
##      .MR6            0.787      0.039     20.338      0.000
##      .MR3            1.485      0.067     22.235      0.000
##      MR11            1.152      0.040     28.595      0.000
##      .MR1            0.958      0.035     27.435      0.000
##      MR5             0.489      0.027     17.983      0.000
##      MR2             1.338      0.046     28.895      0.000

```

## 5.2 Índices de adequação

No SEM, existem vários índices de adequação (fitness) que refletem o quão adequado é o modelo aos dados disponíveis. No entanto, não há acordo entre os pesquisadores sobre quais índices de adequação usar. [Hair et al. \(2000\)](#) recomendam o uso de pelo menos um índice de adequação de cada categoria de ajuste do modelo. Existem três categorias de ajuste de modelo, nomeadamente Ajuste Absoluto, Ajuste Incremental e Ajuste Parcimonioso.

A escolha do índice para escolher em cada categoria para relatar depende de qual literatura está sendo referenciada. As informações sobre a categoria de ajuste do modelo e seu nível de aceitação são apresentados na tabela a seguir.

Índice	Nome do Índice	Nível de aceitação
Chi-quadrado	Discrepancy Chi Square	p-valor > 0,05
<b>RMSEA</b>	<b>Root Mean Square of Error Approximation</b>	<b>RMSEA &lt; 0,08</b>
<b>GFI</b>	<b>Goodness of Fit Index</b>	<b>GFI &gt; 0,90</b>
AGFI	Adjusted Goodness of Fit	AGFI > 0,90
<b>CFI</b>	<b>Comparative Fit Index</b>	<b>CFI &gt; 0,90</b>
TLI	Tucker-Lewis Index	TLI > 0,90
NFI	Normed Fit Index	NFI > 0,90
<b>Chi-quadrado/df</b>	<b>Chi Square/Degrees of Freedom</b>	<b>Chi-quadrado/df &lt; 3,0</b>

Os índices em negrito são recomendados, já que são frequentemente reportados na literatura ([Zainudin, 2012](#)).

Os valores de corte aceitáveis relatados pelos pesquisadores podem variar dependendo do suporte da literatura que eles estão se referindo. Os índices calculados para o modelo proposta está na tabela [5.7](#).

Table 5.7

<b>Índice</b>	<b>Valor</b>	<b>Comentário</b>
Chi-quadrado (p-valor)	0	O nível recomendado NÃO FOI alcançado.
RMSEA	0.041	O nível recomendado FOI alcançado.
GFI	0.962	O nível recomendado FOI alcançado.
AGFI	0.953	O nível recomendado FOI alcançado.
CFI	0.966	O nível recomendado FOI alcançado.
TLI	0.961	O nível recomendado FOI alcançado.
NFI	0.961	O nível recomendado FOI alcançado.
Chi-quadrado/df	7.5	O nível recomendado NÃO FOI alcançado.



# Chapter 6

## Confiabilidade

Confiabilidade é uma avaliação do grau de consistência entre múltiplas medidas de uma variável (Hair et al., 2000). A avaliação da confiabilidade de um modelo de medição pode ser feita usando os seguintes critérios (Zainudin, 2012):

- Alfa de Cronbach
- Confiabilidade Composta (CR)
- Variância Média Extraída (AVE)

**Alfa de Cronbach** é uma medida de confiabilidade que avalia a consistência da escala inteira. O limite inferior para o alfa de Cronbach geralmente aceito é de 0,70, apesar de poder diminuir para 0,60 em pesquisa exploratória. Uma questão na avaliação do alfa de Cronbach é sua relação positiva com o número de itens na escala. Valores mais restritos devem ser usados com maior número de itens, já que o valor de confiabilidade aumenta mesmo com grau igual de intercorrelação. (Hair et al., 2000)

Uma regra prática é que um alfa de 0,70 indica confiança aceitável, 0,80 ou mais indica boa confiança, e 0,90 ou superior é extremamente bom (Ian Ruginski, 2019).

Os valores de alfa de Cronbach estão acima de 0,70 para todos os fatores, indicando que o modelo é confiável, conforme mostra a tabela 6.1.

A **Confiabilidade Composta (CR)** indica a confiabilidade e a consistência interna de um construto latente. Um valor de CR > 0,6 é necessário para atingir a confiabilidade composta para um construto (Zainudin, 2012).

A **Variância Média Extraída (AVE)** indica a porcentagem média de variação explicada pelos itens de medição para um construto latente. Um AVE > 0,5 é necessário para cada construto (AVE é calculado usando a fórmula fornecida).

$$AVE = \frac{\sum K^2}{n}$$

$$CR = \frac{(\sum K)^2}{(\sum K)^2 + (\sum 1 - K^2)}$$

Table 6.1

Reliability	MR6	MR3	MR11	MR1	MR5	MR2
alpha	0.854	0.758	0.861	0.895	0.71	0.879
omega	0.863	0.76	0.863	0.896	0.707	0.879
omega2	0.863	0.76	0.863	0.896	0.707	0.879
omega3	0.864	0.76	0.863	0.893	0.703	0.879
avevar	0.617	0.515	0.559	0.741	0.445	0.593

onde  $K$  é a carga fatorial de cada item, e  $n$  é o número de itens no fator.

A confiabilidade composta e variância média extraída estão com valores desejáveis em todos os fatores, como mostra a tabela 6.2.

Table 6.2: Confiabilidade Composta e Variância Média Extraída.

Fator	CR	AVE	Comentário
MR1	0.918	0.661	CR aceitável, AVE aceitável.
MR2	1.034	1.245	CR aceitável, AVE aceitável.
MR3	0.950	0.813	CR aceitável, AVE aceitável.
MR5	0.737	0.499	CR aceitável, AVE inaceitável.
MR6	0.978	0.887	CR aceitável, AVE aceitável.
MR11	0.949	0.759	CR aceitável, AVE aceitável.

## 6.1 Validade discriminante

Validade discriminante é o grau em que um construto é verdadeiramente diferente dos demais. Logo, validade discriminante elevada oferece evidência de que um construto é único e captura alguns fenômenos que outras medidas não conseguem (Hair et al., 2000).

A validade discriminante pode ser avaliada por meio da razão heterotração-monotração (HTMT) das correlações proposto por Henseler et al. (2015). Especificamente, ele avalia a correlação média entre os indicadores entre os construtos (ou seja, correlações de heterotração-heterométrico), em relação à correlação média entre os indicadores dentro do mesmo construto (ou seja, correlações monotração-heterométrico). Os valores HTMT

resultantes são interpretados como estimativas de correlações entre construtos. Valores absolutos das correlações são recomendados para calcular a matriz HTMT ([Jorgensen et al., 2020](#)).

O uso de HTMT como critério envolve compará-lo a um limite predefinido. Se o valor do HTMT for superior a esse limite, pode-se concluir que não há validade discriminante ([Henseler et al., 2015](#)). O nível de limite exato do HTMT é discutível; afinal, “quando uma correlação é próxima de um”? Alguns autores sugerem um limite de 0,85 ([Clark and Watson, 1995](#), [Kline \(2005\)](#)), enquanto outros propõem um valor de 0,90 ([Gold et al., 2001](#), [Teo et al. \(2009\)](#)).

A avaliação da validade discriminante usando a razão Heterotração-Monotração indica que o modelo é válido, já que todos os valores estão abaixo de 0,80.

```
semTools::htmt mdl, tbl, sample.cov = NULL, missing = 'listwise',
               ordered = NULL, absolute = TRUE)
```

```
##      MR6   MR3   MR11  MR1   MR5   MR2
## MR6   1.000
## MR3   0.248 1.000
## MR11  0.476 0.296 1.000
## MR1   0.679 0.223 0.523 1.000
## MR5   0.185 0.219 0.355 0.185 1.000
## MR2   0.353 0.157 0.565 0.408 0.358 1.000
```

## 6.2 Validade convergente e discriminante

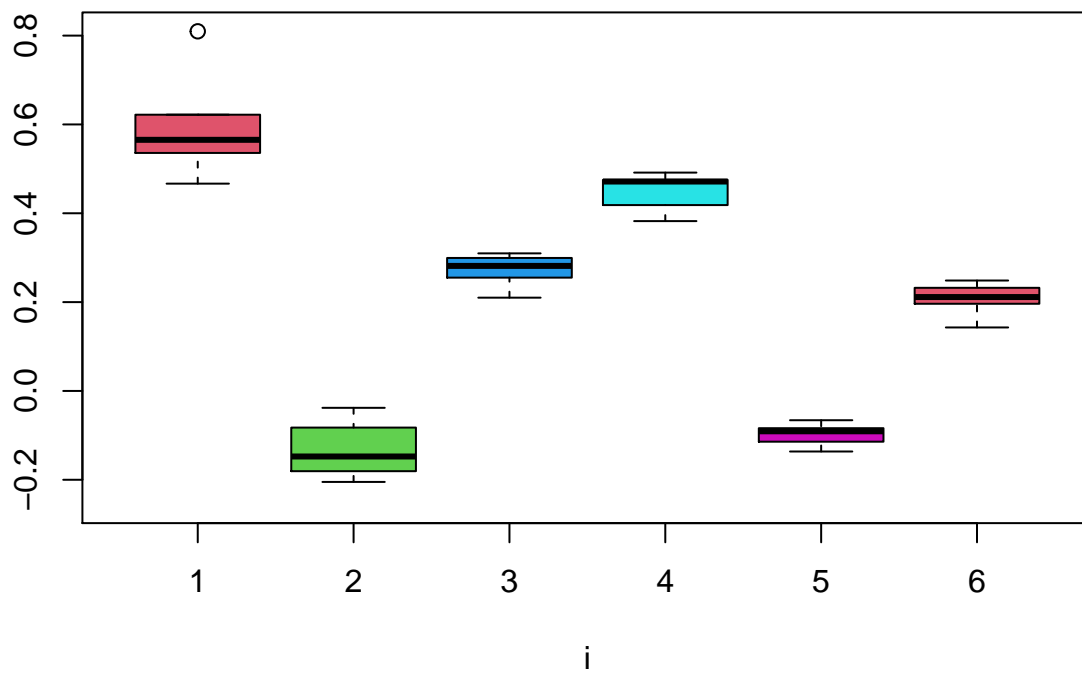
Abordagem Multitrait Multimétodo de validação de escala

Essa função é destinada a avaliar a validade convergente e discriminante das subescalas de uma determinada escala. Os itens pertencentes à mesma subescala devem estar altamente correlacionados entre si. Os itens pertencentes a diferentes subescalas não devem estar altamente correlacionados. Essa abordagem é mais simples e mais robusta do que a análise fatorial confirmatória (CFA). Pode ser interessante verificar (pelo menos aproximadamente) a estrutura proposta de um instrumento existente em uma nova população. A maioria dos psicometristas, entretanto, prefere o CFA ([Falissard, 2012](#)).

```
vars <- list()
for (i in 1:length(mdl)){
  fator <- str_extract_all(mdl, '[a-zA-Z0-9]+')[[i]][1]
  vars[[i]] <- str_extract_all(mdl, '[a-zA-Z0-9]+')[[i]][-1]
}
y <- psy::mtmm(tbl, vars, color=TRUE, itemTot=TRUE, graphItem=FALSE,
               stripChart=FALSE, namesDim=NULL)
```

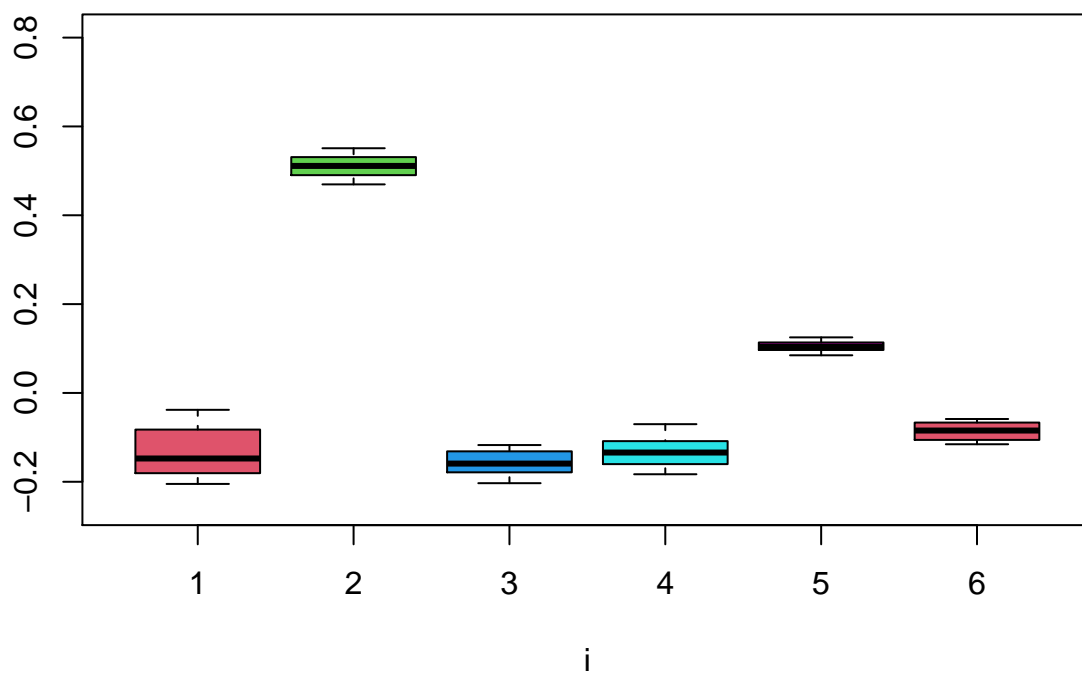
Correlation of Items of Scale i with Items of Scale 1

## Scale 1



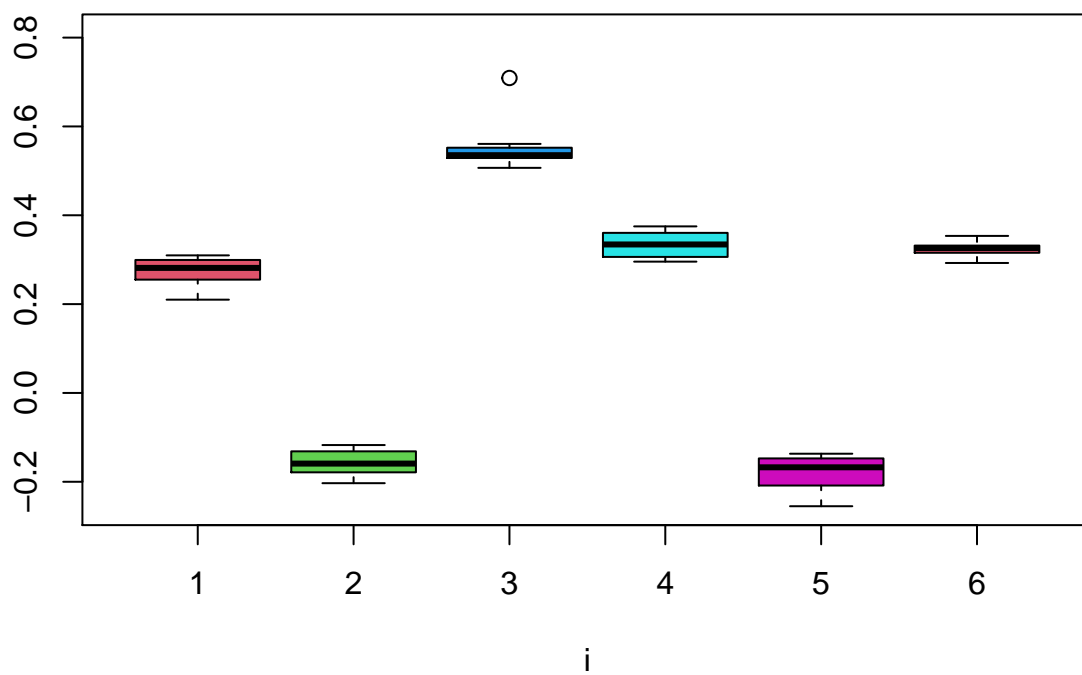
Correlation of Items of Scale i with Items of Scale 2

## Scale 2



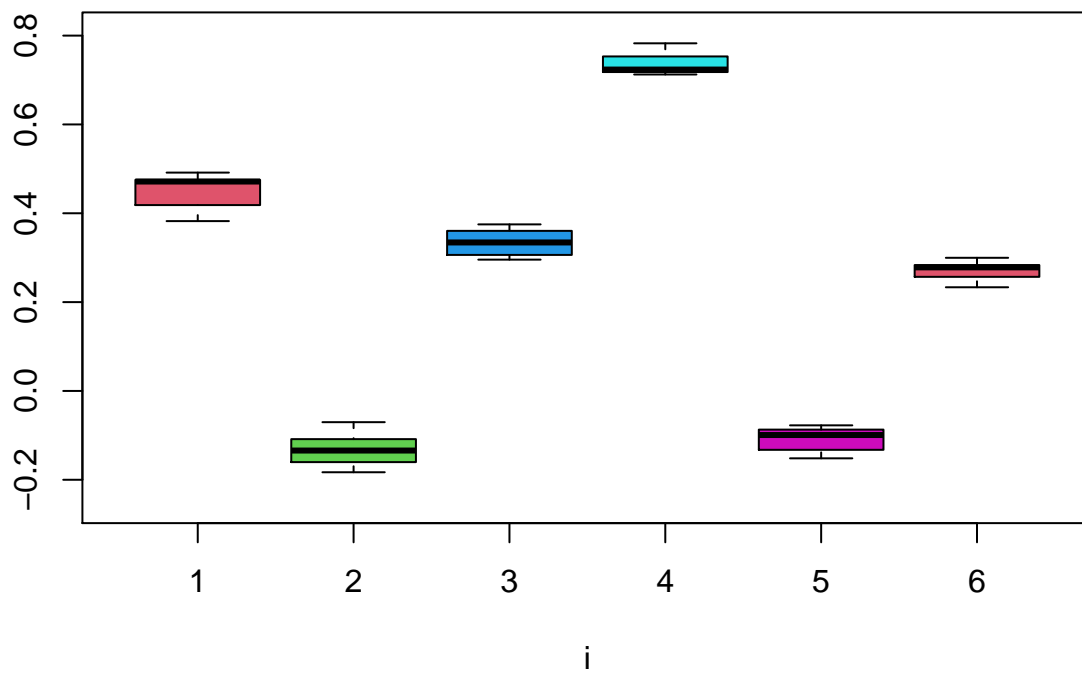
Correlation of Items of Scale i with Items of Scale 3

### Scale 3



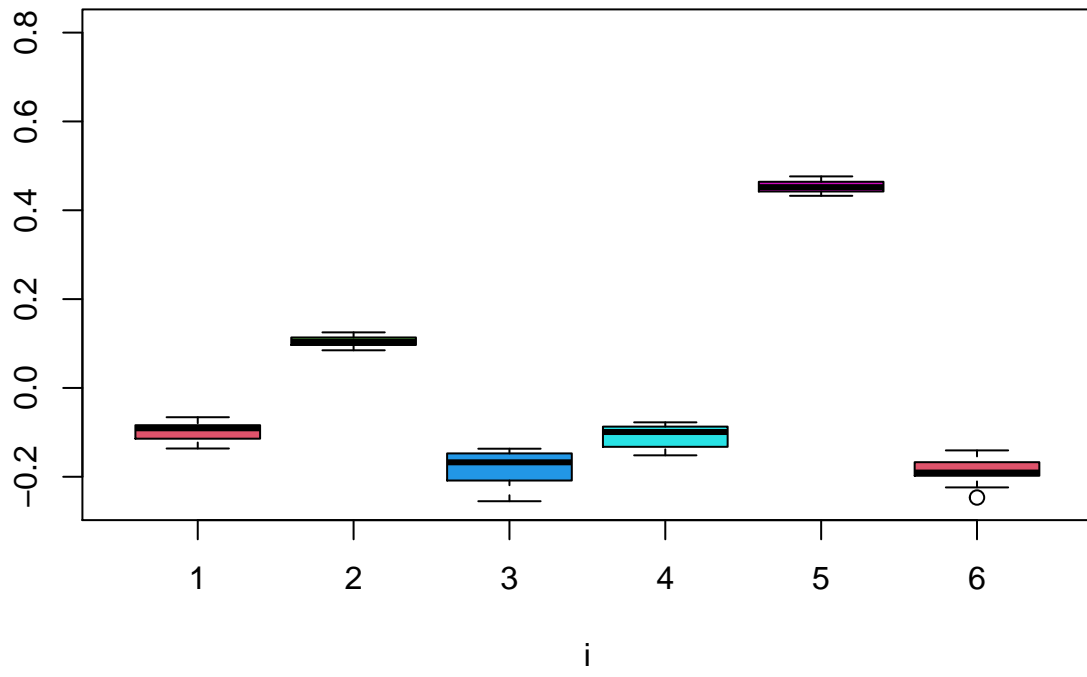
Correlation of Items of Scale i with Items of Scale 4

### Scale 4



Correlation of Items of Scale i with Items of Scale 5

## Scale 5





# Apêndice

Significado das variáveis do modelo.

Variável	Pergunta
CIDADE	CIDADE
PF1	Você mora na cidade de _____?
PF2	Tempo que mora na cidade
IDADE1	IDADE
SEXO	SEXO
P1A	Você diria que tem muito, pouco ou nenhum interesse pela Copa do Mundo de Futebol de 2014 realizada no Brasil?
P1B	De um modo geral, para você, a Copa do Mundo no Brasil, foi ótima, boa, regular, ruim ou péssima?
P1CA	A Copa aproximou a comunidade brasileira
P1CB	A Copa foi uma oportunidade para os brasileiros conhecerem novas pessoas
P1CC	A Copa promoveu o sentimento de orgulho entre os brasileiros
P1CD	A Copa serviu para estreitar os laços de coesão entre os moradores da cidade
P1CE	A Copa diminuiu a qualidade de vida dos moradores da cidade
P1CF	A Copa levou à superlotação das instalações e da infraestrutura da cidade
P1CG	A Copa aumentou os índices de violência
P1CH	A Copa aumentou os índices de prostituição

Variável	Pergunta
P1CI	A Copa aumentou a consciência ambiental
P1CJ	A Copa trouxe mais agressões ao meio ambiente
P1CK	A Copa aumentou os índices de poluição sonora, visual e do ar
P1CL	A Copa melhorou a conservação e a proteção do meio ambiente
P1CM	A Copa aumentou o lixo e a desorganização na cidade
P1CN	A Copa deu oportunidades de emprego para os moradores da cidade
P1CO	A Copa melhorou os serviços públicos e infraestrutura da cidade
P1CP	A Copa aumentou as oportunidades de negócios
P1CQ	A Copa promoveu a regeneração e o desenvolvimento das cidades
P1CR	A Copa aumentou a exposição do Brasil na mídia internacional
P1CS	A Copa melhorou a imagem internacional do Brasil
P1CT	A Copa promoveu o Brasil como um destino turístico
P1CU	A Copa trouxe aumento de impostos para os brasileiros
P1CV	Eu estava bem informado(a) sobre a Copa do Mundo de 2014
P1CW	Eu estava ciente dos possíveis impactos da Copa na minha cidade
P1CX	Eu entendi o papel do Governo na Organização da Copa
P1CY	Eu entendi o papel do Comitê Organizador na organização da Copa
P2A	Me senti fascinado
P2B	Me senti Inspirado
P2C	Me senti Aborrecido
P2D	Me senti Provocado
P2E	Me senti Insatisfeito
P2F	Me senti Maravilhado
P2G	Me senti Surpreendido
P2H	Me senti Nervoso
P2I	Eu senti Amor

Variável	Pergunta
P2J	Eu senti Carinho
P2K	Eu senti Compaixão
P2L	Eu me senti Triste
P2M	Eu me senti Feliz
P2N	Eu me senti Tenso
P2O	Eu me senti Contente
P2P	Eu me senti Descontente
P3A	Durante a Copa do Mundo, você acompanhou os jogos que ocorreram na cidade onde você mora
P3G	
P3H	
P3I	
P4A	Estou contente com o fato do Brasil ter sido sede da Copa do Mundo
P4B	O Brasil se preparou para sediar a Copa do Mundo
P4C	Eu apoiei o Brasil como sede da Copa do Mundo
P4D	Eu odiei a ideia do Brasil sediar a Copa do Mundo
P4E	Sediar a Copa do Mundo foi desperdício dos nossos recursos
P4F	Sediar a Copa do Mundo foi motivo de orgulho nacional
P4G	Sediar a Copa não foi interessante para o Brasil
P5A	A economia brasileira está com graves problemas
P5B	O governo promoveu a criação de empregos nas cidades sede da Copa
P5C	São necessários mais empregos na cidade e região
P5D	Estou disposto a pagar mais impostos desde que haja maiores retornos para a população
P5E	Na cidade onde moro, são necessários mais empregos para que a população jovem não tenha que migrar para outras áreas
P5F	Estou satisfeito em morar nesta cidade

Variável	Pergunta
P6AA	O Governo fez o que é correto para o bom desenvolvimento da Copa do Mundo
P6AB	O Governo fez o que é correto para a boa organização da Copa do Mundo
P6AC	O Governo agiu de acordo com os interesses dos moradores da cidade onde você mora
P6AD	Eu confiei nas decisões tomadas pelo Governo sobre a Copa
P6AE	O Governo se esforçou em incorporar os moradores da cidade no planejamento da Copa
P6BA	O Comitê Organizador fez o que é correto para o bom desenvolvimento da Copa do Mundo
P6BB	O Comitê Organizador fez o que é correto para a boa organização da Copa do Mundo
P6BC	O Comitê Organizador agiu de acordo com os interesses dos moradores da cidade onde você mora
P6BD	Eu confiei nas decisões tomadas pelo Comitê Organizador sobre a Copa
P7A	Este evento significou muito para mim
P7B	Estive muito apegado(a) à Copa do Mundo de 2014
P7C	Identifiquei-me muito com este evento
P7D	Tive uma ligação especial com o evento e com as pessoas que ajudaram em sua organização
P7E	Preferi participar deste evento esportivo do que de outros eventos
P7F	Não trocaria este evento esportivo por outros eventos de lazer ou de entretenimento
P7G	Eu me preocupei em receber bem os turistas que vieram à cidade assistir os jogos da Copa
P7AA	Houve irregularidades na construção das obras da Copa do Mundo no Brasil

Variável	Pergunta
P7AB	Houve empresas que não cumpriram os padrões exigidos nos projetos da Copa do Mundo no Brasil
P7AC	Há uma sensação de que houve desvio de fundos públicos e superfaturamento nas obras da Copa no Brasil
P7AD	Algumas obras feitas para a da Copa do Mundo no Brasil eram desnecessárias
P7AE	Houve favoritismo e amiguismo na seleção de empresas para construir as obras da Copa no Brasil
P7BA	As obras da Copa do Mundo no Brasil foram EXECUTADAS de forma transparente
P7BB	Todo o processo da Copa do Mundo no Brasil foi DIVULGADO pelo Governo e pela mídia de forma clara
P7BC	Os moradores locais puderam ver e acompanhar claramente o progresso e a situação das obras da Copa do Mundo no Brasil
P7BD	As obras da Copa do Mundo no Brasil foram REALIZADAS com total transparência
P7BE	O Governo divulgou informações suficientes aos moradores locais sobre o andamento das obras da Copa do Mundo no Brasil
P8A	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8B	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona

Variável	Pergunta
P8C	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8D	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8E	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8F	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8G	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P8H	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona
P9A	pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:

Variável	Pergunta
P9B	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9C	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9D	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9E	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9F	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9G	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
P9H	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:

Variável	Pergunta
P9I	Pensando nessa escala de 1 a 5, onde você se posiciona quando pensa sobre qual tipo de pessoa você é. As características descritas são opostas, por favor, utilizando essa escala, me diga onde você se posiciona:
ESTCIVIL	ESTADO CIVIL
PEA	OCUPAÇÃO PRINCIPAL
ESCOLA	ESCOLARIDADE
RENDAF	RENDAMENTO FAMILIAR MENSAL
COR	COR



# Referências

- Child, D. (2006). *The Essentials of Factor Analysis*. Bloomsbury Academic.
- Clark, L. A. and Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3):309–319.
- Comrey, A. and Lee, H. (2013). *A First Course in Factor Analysis*. Taylor & Francis.
- Costello, A. and Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*, 10:1–9.
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, 11(2):213–228. Publisher: Instituto Brasileiro de Avaliação Psicológica. UFRGS.
- Falissard, B. (2012). Various procedures used in psychometry.
- Field, A. P. (2009). *Discovering statistics using SPSS: and sex, drugs and rock 'n' roll*. SAGE Publications, Los Angeles, 3rd ed edition.
- Gold, A. H., Malhotra, A., and Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, 18(1):185–214. Publisher: Routledge \_eprint: <https://doi.org/10.1080/07421222.2001.11045669>.
- Hair, J. J. F., Black, W. C., and Sant’Anna, A. S. (2000). *Análise multivariada de dados (6a. ed.)*. Grupo A - Bookman. OCLC: 923756114.
- Henseler, J., Ringle, C. M., and Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1):115–135.
- Ian Ruginski (2019). Structural Equation Modeling in R Tutorial 6: Confirmatory Factor Analysis using lavaan in R.
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2020). *textttsemTools: Useful tools for structural equation modeling*.
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4):401–415.
- Kline, P. (1994). *An Easy Guide to Factor Analysis*. Routledge.

- Kline, R. (2005). Principles and Practice of Structural Equation Modeling (2nd Edition).
- Lorenzo-Seva, U., Timmerman, M., and Kiers, H. (2011). The Hull Method for Selecting the Number of Common Factors. *Multivariate Behavioral Research*, 46:340–364.
- Pasquali, L. and Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação em Psicologia*, 2:99 – 110. Publisher: scielopecs.
- Revelle, W. R. (2017). *psych: Procedures for Personality and Psychological Research*.
- Tabachnick, B. G. and Fidell, L. S. (2006). *Using Multivariate Statistics (5th Edition)*. Allyn & Bacon, Inc., USA.
- Teo, T., Srivastava, S., and Jiang, L. (2009). Trust and Electronic Government Success: An Empirical Study. *J. of Management Information Systems*, 25:99–132.
- Waternaux, C. M. (1976). Asymptotic Distribution of the Sample Roots for a Nonnormal Population. *Biometrika*, 63(3):639–645. Publisher: [Oxford University Press, Biometrika Trust].
- Yong, A. and Pearce, S. (2013). A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, 9:79–94.
- Zainudin, A. (2012). *A handbook on SEM: Structural equation modelling using amos graphics*. Universiti Sultan Zainal Abidin, 2 edition.