

# Análise Psicométrica da escala de rastreio SQR-20 usando dados do GAPSI e Apoia USP

Gabriela Amaral (gabriela.jj.ga@usp.br)  
Brunna Quatrochi (brunnaquatrochi@usp.br)  
Jorge Luis Bazán (jlbazan@icmc.usp.br)  
Emanuela Pap da Silva (emanuela@sc.usp.br)

Trabalho desenvolvido como parte dos projetos das edições 2023-2024 e 2024-2025 do Programa Unificado de Bolsas de Estudo para apoio à Formação de Estudantes de Graduação da Universidade de São Paulo (PUB-USP).

Orientação: Jorge Luis Bazán

Agradecimentos: Emanuela Pap da Silva, Assistente Social do Serviço de Promoção Social da USP São Carlos, e Tais Bleicher, Professora do Departamento de Psicologia da UFSCar autores dos questionário e dos estudos que deram origem aos dados estudados.

## Resumo

Neste notebook, é realizada uma análise psicométrica da escala de rastreamento SRQ-20, composta por vinte itens dicotômicos que avaliam indicadores de transtornos mentais comuns (TMC), com respostas “sim” ou “não”. Foram utilizadas bases de dados dos anos de 2021, 2022 e 2023, obtidas por meio de questionários aplicados pela GAPSI e Apoia USP, que são serviços de apoio psicossocial à saúde mental da comunidade universitária da USP, campus de São Carlos, respondidos por alunos de graduação deste campus. Os resultados indicam que o teste é confiável e unidimensional, e os escores do TMC podem ser obtidos tanto pela teoria clássica dos testes quanto pelo modelo de dois parâmetros da teoria da resposta ao item (TRI). O modelo TRI de dois parâmetros foi escolhido como o melhor modelo para reportar os escores do TMC que medem o sofrimento mental. No modelo escolhido, os itens são caracterizados pelos parâmetros de dificuldade e discriminação. A dificuldade de um item pode ser interpretada como a gravidade do sintoma abordado pela questão, enquanto o parâmetro de discriminação mede a eficácia do item em diferenciar respondentes com diferentes níveis de sofrimento mental avaliado. Uma avaliação do modelo foi realizada, e os resultados foram salvos. A análise realizada neste estudo permitiu identificar as características psicométricas dos itens, oferecendo uma compreensão mais aprofundada das propriedades do instrumento e sua adequação para avaliar a saúde mental dos estudantes.

## Introdução

O SRQ-20 é uma escala de rastreio composta por vinte itens dicotômicos, com respostas “sim” ou “não”. Desenvolvido pela Organização Mundial da Saúde (OMS), o instrumento não tem o objetivo de produzir diagnóstico, e, sim, indicar a suspeita de transtornos do humor, de ansiedade e de somatização, denominados na literatura como Transtornos Mentais Comuns (TMC) especialmente em contextos de atenção primária. Ele é composto por 20 questões com respostas sim/não, referentes aos últimos 30 dias do momento do preenchimento, sobre sintomas psíquicos e somáticos.

Sua praticidade e objetividade tornaram o SRQ-20 amplamente utilizado em diversos países, sendo empregado em inúmeras pesquisas ao redor do mundo. Veja Silveira et al. (2022).

Questionários desenvolvidos pelo GAPSÍ e pelo Apoia USP foram aplicados nos anos de 2021, 2022 e 2023, com o objetivo de explorar diferentes aspectos do estado psíquico dos estudantes da graduação e identificar tanto as dificuldades quanto as potencialidades associadas à vivência acadêmica na comunidade USP de São Carlos. A escala de rastreio SRQ-20 faz parte dos questionários e as respostas dos estudantes será usada para desenvolver uma análise psicométrica da escala de rastreio SRQ-20.

## Pacotes

Instalação dos pacotes que serão usados para analisar os dados:

```
# install.packages("ggplot2")
# install.packages("readxl")
# install.packages("foreign")
# install.packages("ltm")
# install.packages("mirt")
# install.packages("psych")
# install.packages("psychometric")
# install.packages("irttoys")
# install.packages("mirtCAT")
# install.packages("openxlsx")

library(ggplot2)
library(readxl)
require(foreign)
library(ltm)
library(mirt)
library(psych)
library(psychometric)
library(irttoys)
library(mirtCAT)
library(openxlsx)
```

## 1 Importando a Base de dados

O conjunto de dados utilizado neste estudo abrange 823 estudantes e contém 20 variáveis dicotômicas (“sim” e “não”) relacionadas aos itens do SRQ-20. Essas informações foram obtidas por meio de um formulário aplicado a alunos de graduação. As perguntas incluídas no questionário são descritas a seguir:

- **SRQ01:** Você tem dores de cabeça frequentes?
- **SRQ02:** Tem falta de apetite?
- **SRQ03:** Dorme mal?
- **SRQ04:** Assusta-se com facilidade?
- **SRQ05:** Tem tremores nas mãos?
- **SRQ06:** Sente-se nervoso(a), tenso(a) ou preocupado(a)?
- **SRQ07:** Tem má digestão?
- **SRQ08:** Tem dificuldade de pensar com clareza?
- **SRQ09:** Tem se sentido triste ultimamente?
- **SRQ10:** Tem chorado mais do que de costume?

- **SRQ11:** Encontra dificuldades para realizar com satisfação suas atividades diárias?
- **SRQ12:** Tem dificuldade para tomar decisões?
- **SRQ13:** Tem dificuldade no serviço (seu trabalho é penoso, causa-lhe sofrimento)?
- **SRQ14:** É incapaz de desempenhar um papel útil em sua vida?
- **SRQ15:** Tem perdido o interesse pelas coisas?
- **SRQ16:** Você se sente uma pessoa inútil, sem préstimo?
- **SRQ17:** Tem tido ideia de acabar com a vida?
- **SRQ18:** Sente-se cansado(a) o tempo todo?
- **SRQ19:** Você se cansa com facilidade?
- **SRQ20:** Tem sensações desagradáveis no estômago?

A escala explora sintomas não psicóticos, como insônia, fadiga, alterações no apetite, dificuldades de pensamento, alterações de humor e problemas somáticos, que são manifestações típicas dos TMC.

Primeiro, faremos a leitura dos dados e selecionaremos os itens do SRQ-20 que serão analisados.

```
#setwd(dirname(rstudioapi::getActiveDocumentContext())$path))

df1 <- read_excel("DadosNotebook.xlsx")

# Selecionando apenas as questões do SRQ
df_srq <- subset(df1, select = c(2:21))

# Renomeando as colunas do df_srq de SRQ01 a SRQ20
colnames(df_srq) <- sprintf("SRQ%02d", 1:20)

# Convertendo para número
df_srq[] <- lapply(df_srq, function(x) if(is.character(x)) as.numeric(x) else x)
```

## 2 Teoria Clássica dos Testes (TCT)

A Teoria Clássica dos Testes (TCT), também conhecido como modelo clássico dos testes, foi desenvolvido a partir de trabalhos de Spearman e Binet (Sartes & Souza-Formigoni 2013). Ela é uma abordagem da Psicometria que determina a dificuldade dos itens de um teste ou a habilidade (ou desempenho) daqueles que foram submetidos a esse teste, sendo o principal objetivo desse modelo a compreensão e melhoria da confiabilidade dos testes (Bazán, 2018).

A TCT se baseia na ideia principal de que é possível estimar a pontuação verdadeira de um indivíduo a partir do valor observado pelo teste, devido a associação entre tais variáveis. O método define o seguinte:

$$X = V + E$$

em que  $X$  é o escore (ou valor) observado,  $V$  é o escore (ou pontuação) verdadeiro e  $E$  é o erro aleatório. Temos que a esperança do escore observado é igual ao escore verdadeiro, ou seja, quanto maior a amostra observada, maior é a precisão do resultado do teste.

Neste estudo, inicialmente desenvolvemos uma análise clássica de itens com as principais estatísticas para os itens do questionário. Além disso, obtivemos as estimativas de Alfa de Cronbach do questionário, usual para analisar sua confiabilidade e a contribuição das questões nela. Ambas as análises são descritas a seguir.

## 2.1 Análise Clássica de Itens

A análise clássica de itens é um método para avaliar a qualidade dos itens de teste usando a Teoria Clássica de Testes (TCT). Esta análise é uma abordagem simples e intuitiva que usa estatísticas descritivas para avaliar a eficácia de cada item com foco na média e desvio padrão. No caso de um questionário com itens dicotômicos ou binários, a média de cada item é a proporção de resposta obtida na resposta Sim ( $p$ ) sendo interpretada como a dificuldade do item ou questão. Assim, alguns pesquisadores identificaram itens com valores menores que 0,30 como itens difíceis e aqueles maiores que 0,70 como itens fáceis. Os itens muito fáceis e os itens muito difíceis podem precisar ser revisados para obter um teste válido. Um bom teste deve ter uma variedade de itens, desde fácil, itens moderados a difíceis de acordo com Wright e Stone (1979).

Também no caso de itens dicotômicos ou binários, a maior variabilidade é obtida quando a média é  $p = 0,5$ . Neste caso, a variância do item  $i$  é  $\sigma_{x_i}^2 = p \times (1-p) = 0,25$  e o desvio padrão é  $\sigma_{x_i} = 0,5$ . Maior variabilidade é desejável para que o item consiga capturar uma maior diversidade de respostas.

A função *describe* no pacote psych destina-se a produzir as estatísticas solicitadas com mais frequência em estudos psicométricos e de psicologia, apresentando-as de forma clara em uma tabela de fácil leitura.

```
# Análise descritiva dos itens
```

```
des<-describe(df_srq)
```

```
print(des,digits=3)
```

##	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
## SRQ01	1	823	0.434	0.496	0	0.417	0	0	1	1	0.267	-1.931
## SRQ02	2	823	0.345	0.476	0	0.307	0	0	1	1	0.651	-1.579
## SRQ03	3	823	0.608	0.489	1	0.634	0	0	1	1	-0.440	-1.809
## SRQ04	4	823	0.412	0.492	0	0.390	0	0	1	1	0.357	-1.875
## SRQ05	5	823	0.300	0.459	0	0.250	0	0	1	1	0.871	-1.243
## SRQ06	6	823	0.861	0.346	1	0.951	0	0	1	1	-2.089	2.367
## SRQ07	7	823	0.350	0.477	0	0.313	0	0	1	1	0.628	-1.607
## SRQ08	8	823	0.529	0.499	1	0.536	0	0	1	1	-0.114	-1.989
## SRQ09	9	823	0.634	0.482	1	0.668	0	0	1	1	-0.557	-1.692
## SRQ10	10	823	0.316	0.465	0	0.270	0	0	1	1	0.791	-1.377
## SRQ11	11	823	0.722	0.448	1	0.777	0	0	1	1	-0.988	-1.025
## SRQ12	12	823	0.655	0.476	1	0.693	0	0	1	1	-0.651	-1.579
## SRQ13	13	823	0.409	0.492	0	0.387	0	0	1	1	0.368	-1.867
## SRQ14	14	823	0.333	0.472	0	0.291	0	0	1	1	0.708	-1.501
## SRQ15	15	823	0.605	0.489	1	0.631	0	0	1	1	-0.429	-1.818
## SRQ16	16	823	0.428	0.495	0	0.410	0	0	1	1	0.292	-1.917
## SRQ17	17	823	0.187	0.390	0	0.109	0	0	1	1	1.602	0.566
## SRQ18	18	823	0.677	0.468	1	0.721	0	0	1	1	-0.755	-1.432
## SRQ19	19	823	0.722	0.448	1	0.777	0	0	1	1	-0.988	-1.025
## SRQ20	20	823	0.416	0.493	0	0.395	0	0	1	1	0.342	-1.885
##	se											
## SRQ01	0.017											
## SRQ02	0.017											
## SRQ03	0.017											
## SRQ04	0.017											
## SRQ05	0.016											
## SRQ06	0.012											
## SRQ07	0.017											
## SRQ08	0.017											
## SRQ09	0.017											
## SRQ10	0.016											
## SRQ11	0.016											

```

## SRQ12 0.017
## SRQ13 0.017
## SRQ14 0.016
## SRQ15 0.017
## SRQ16 0.017
## SRQ17 0.014
## SRQ18 0.016
## SRQ19 0.016
## SRQ20 0.017

```

Baseado nos resultados mostrados, verifica-se que as três questões com médias mais altas, ou seja, as questões em que há mais indivíduos que responderam “Sim” para a questão, são: “Sente-se nervoso(a), tenso(a) ou preocupado(a)?” (SRQ06), “Encontra dificuldades para realizar com satisfação suas atividades diárias?” (SRQ11) e “Você se cansa com facilidade?” (SRQ19). Temos que 86%, 72% e 72%, respectivamente, dos indivíduos confirmaram ter tais características, proporções significativas. Esses itens estão principalmente ligados a aspectos emocionais e físicos relacionados ao estresse e ao bem-estar geral. Desde um ponto de vista da TCT podemos dizer que estes itens foram fáceis.

Em contrapartida, verifica-se que as três questões com médias mais baixas são: “Tem tremores nas mãos?” (SRQ05), “Tem chorado mais do que de costume?” (SRQ10) e “Tem tido ideia de acabar com a vida?” (SRQ17). Observa-se que 30%, 32% e 19% dos indivíduos apresentam esses sintomas. Embora sejam as proporções mais baixas entre as perguntas, são números significativos e que não devem ser ignorados. Esses sintomas, embora menos prevalentes, podem estar associados a questões graves de saúde mental, como transtornos de ansiedade e depressão, e merecem atenção especial. Desde um ponto de vista da TCT podemos dizer que estes itens são difíceis.

Analisando os desvios padrão das questões, nota-se que seus valores são semelhantes com valores que variam entre 0.45 e 0.50 com exceção das perguntas “Sente-se nervoso(a), tenso(a) ou preocupado(a)?” (SRQ06) e “Tem tido ideia de acabar com a vida?” (SRQ17), que apresentam desvios padrão mais baixos, 0.35 e 0.39, respectivamente. Entretanto, na maioria dos itens há uma dispersão similar que é próxima da maior variabilidade esperada e, portanto, adequados. Já as questões SRQ06 e SRQ17 apresentam uma menor variabilidade. No caso do item SRQ06 por ser o item mais fácil e no caso do item SRQ17 por ser o mais difícil. Em ambos os casos, estes itens requerem cuidado nas interpretações.

## 2.2 Alfa de Cronbach Ordinal

O alfa de Cronbach é índice obtido para avaliar a confiabilidade de um questionário comparando a quantidade de variância compartilhada, ou covariância, entre os itens que compõem o questionário com a quantidade de variância geral. Este índice é considerado uma medida de consistência interna do questionário e é definido como

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{i=1}^k \sigma_{x_i}^2}{\sigma_X^2} \right) = \frac{k}{k-1} \left( \frac{2 \sum_{i < j} \sigma_{x_i x_j}}{\sigma_X^2} \right) = \frac{k\bar{c}}{\bar{v} + (k-1)\bar{c}}$$

em que:

$k$  é o número de itens ou questões do questionário,  $\sigma_{x_i}^2$  é a variância das respostas de cada item  $i$ ,  $\sigma_{x_i x_j}$  é a covariância entre as respostas dos itens  $i$  e  $j$ ,  $\sigma_X^2 = \sum_{i=1}^n \sigma_{x_i}^2 + 2 \sum_{i < j} \sigma_{x_i x_j}$  é a variância do escore total obtido considerando  $X = \sum_{i=1}^k x_i$ ,  $\bar{v} = \frac{\sum_{i=1}^k \sigma_{x_i}^2}{k}$  é a média das variâncias dos itens e  $\bar{c} = \frac{\sigma_{x_i x_j}}{k(k-1)}$  é a média das covariâncias entre os itens.

As expressões anteriores podem ser expressadas também em termos de correlações dos itens desde que as covariâncias são  $\sigma_{x_i x_j} = r_{x_i x_j} \times \sigma_{x_i} \times \sigma_{x_j}$ .

Neste estudo, utilizamos o Alfa de Cronbach Ordinal, que, diferentemente do Alfa de Cronbach tradicional, emprega a matriz de correlações tetracóricas em vez da matriz de covariâncias de Pearson. Essa abordagem foi escolhida porque o cálculo do Alfa de Cronbach convencional assume que os dados são contínuos, o que não corresponde ao nosso caso, já que os dados são dicotômicos. Quando essa suposição é violada, como ocorre com dados categóricos, a matriz de covariâncias de Pearson pode apresentar distorções, comprometendo a estimativa do coeficiente alfa (Gadermann et al., 2012).

O alfa de Cronbach é um coeficiente que quando mais perto de 1 sinaliza que os escores obtidos de um conjunto de itens serão mais confiáveis. Como os dados do questionário estão baseados numa escala de respostas ordinais, como demonstrado por Zumbo et al. (2017), o alfa ordinal é mais apropriado do que o alfa de Cronbach tradicional. Nunnally & Bernstein (1994) sinalizam que um valor mínimo de 0,8 é recomendável na pesquisa aplicada. Adicionalmente é fornecida a correlação do item com o resto das questões, assim como o alfa de Cronbach do questionário excluindo um determinado item o de melhor forma, se o item é eliminado. O objetivo é criar testes com o menor número de itens que ainda forneçam o nível de confiabilidade desejado.

```
#itemstats(df_srq)
#Alfa ordinal
#Correlação dados ordinais
#ord1<-polychoric(df_srq)
#Correlação dados dicotômicos
ord2<-tetrachoric(df_srq)
```

```
#ord2
psych::alpha(ord2$rho)
```

```
##
## Reliability analysis
## Call: psych::alpha(x = ord2$rho)
##
##   raw_alpha std.alpha G6(smc) average_r S/N median_r
##      0.94      0.94    0.96      0.43  15      0.42
##
##   95% confidence boundaries
##      lower alpha upper
## Feldt  0.89  0.94  0.97
##
## Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r S/N var.r med.r
## SRQ01      0.94      0.94    0.95      0.43  15 0.015  0.42
## SRQ02      0.94      0.94    0.96      0.43  15 0.015  0.42
## SRQ03      0.94      0.94    0.96      0.43  14 0.015  0.42
## SRQ04      0.94      0.94    0.96      0.44  15 0.014  0.43
## SRQ05      0.94      0.94    0.96      0.44  15 0.014  0.43
## SRQ06      0.93      0.93    0.95      0.42  14 0.015  0.39
## SRQ07      0.93      0.93    0.95      0.43  14 0.015  0.42
## SRQ08      0.93      0.93    0.95      0.42  14 0.015  0.41
## SRQ09      0.93      0.93    0.95      0.42  14 0.013  0.41
## SRQ10      0.93      0.93    0.95      0.43  14 0.015  0.42
## SRQ11      0.93      0.93    0.95      0.42  14 0.014  0.41
## SRQ12      0.93      0.93    0.95      0.43  14 0.015  0.42
## SRQ13      0.93      0.93    0.95      0.42  14 0.015  0.41
## SRQ14      0.93      0.93    0.95      0.42  14 0.014  0.41
## SRQ15      0.93      0.93    0.95      0.42  14 0.014  0.41
## SRQ16      0.93      0.93    0.95      0.42  14 0.014  0.41
```

```

## SRQ17      0.93      0.93      0.95      0.42  14 0.015  0.41
## SRQ18      0.93      0.93      0.95      0.42  14 0.015  0.39
## SRQ19      0.93      0.93      0.95      0.42  14 0.015  0.41
## SRQ20      0.93      0.93      0.95      0.43  14 0.014  0.42
##
## Item statistics
##      r r.cor r.drop
## SRQ01 0.58 0.56 0.53
## SRQ02 0.58 0.55 0.53
## SRQ03 0.60 0.57 0.55
## SRQ04 0.51 0.47 0.45
## SRQ05 0.51 0.47 0.45
## SRQ06 0.79 0.78 0.75
## SRQ07 0.63 0.63 0.58
## SRQ08 0.77 0.76 0.74
## SRQ09 0.81 0.81 0.78
## SRQ10 0.64 0.62 0.59
## SRQ11 0.74 0.73 0.71
## SRQ12 0.62 0.60 0.57
## SRQ13 0.73 0.72 0.69
## SRQ14 0.69 0.69 0.65
## SRQ15 0.71 0.70 0.68
## SRQ16 0.74 0.73 0.70
## SRQ17 0.72 0.71 0.68
## SRQ18 0.76 0.76 0.73
## SRQ19 0.71 0.70 0.67
## SRQ20 0.64 0.64 0.59

```

Para encontrar o Alfa de Cronbach do questionário, primeiro calculamos a matriz de correlações tetracóricas entre os itens dado que eles são binários (se forem ordinais teria que ser calculada a matriz de correlações polycóricas) usando a função `tetrachoric` do pacote `psych` e então usando a matriz de correlações calculada é aplicada a função `alpha` do pacote `psych` para obter o Alfa de cronbach e outras estatísticas descritas em Revelle (2024).

Primeiramente, vamos avaliar o Alfa de Cronbach Ordinal do questionário. A avaliação dessa medida se dará considerando as definições de Landis & Koch (1977) e Streiner (2003), que são explicadas por César Gonçalves.

Considerando todos os itens, obtemos um alfa igual a 0.94. Essa medida, que é superior a 0.80, segundo Landis & Koch (1977), indica que a consistência interna do teste é quase perfeita. Entretanto, considerando a afirmação de Streiner (2003) de que os valores ideais do alfa devem estar entre 0.80 e 0.90, temos que pode estar ocorrendo redundância ou duplicações de informações, ou seja, ocorrência de itens que estão medindo exatamente o mesmo elemento.

Para fazer a análise desses resultados, vamos verificar as estatísticas `raw_alpha`, `average_r` e `r.drop`, que serão detalhadas à medida que as análises forem realizadas.

Agora, vamos analisar o Alfa de Cronbach do questionário quando um determinado item é eliminado verificando se o valor de alfa é mais alto, permanece igual ou piora. Isso é reportado como `raw_alpha` na saída apresentada acima. Levando em conta estes resultados observamos que quando os itens “Você tem dores de cabeça frequentes?” (SRQ01), “Tem falta de apetite?” (SRQ02), “Dorme mal?” (SRQ03), “Assusta-se com facilidade?” (SRQ04) ou “Tem tremores nas mãos?” (SRQ05) são retirados, temos que o valor de alfa permanece igual a 0.94, ou seja, a confiabilidade do teste permanece a mesma. Assim, poderíamos propor tirar esses itens já que não contribuem para a melhoria da confiabilidade do teste, no entanto, devido à relevância dessas questões para a avaliação psicológica, estas perguntas serão mantidas.

Por outro lado, quando um dos demais itens são retirados, que incluem por exemplo “Tem se sentido triste ultimamente?” (SRQ09), “Tem chorado mais do que de costume?” (SRQ10), “Você se sente uma pessoa inútil, sem préstimo?” (SRQ16) e “Tem tido ideia de acabar com a vida?” (SRQ17), o valor de Alfa de Cronbach tem uma leve redução de 0.94 para 0.93. Desse modo, o retiro dos itens de SRQ06 a SRQ20 apresentam impacto negativo na confiabilidade do teste e por tanto devem ser mantidos, dada a sua importância para o estudo.

Agora, analisaremos os resultados de **average\_r** que se refere a média da correlação de um determinado item com os outros. Neste caso, encontramos valores acima de 0,42 em todos os itens, sinalizando o relacionamento entre cada item com os demais.

Por fim, analisaremos **r\_drop**, que mede a correlação entre o escore de um item e o escore total dos outros itens sem ele. Essa medida é importante para verificar se determinado item está medindo as mesmas características dos outros itens.

Temos que os itens com as menores correlações são 4 e 5, que se referem, respectivamente, às perguntas “Assusta-se com facilidade?” (SRQ04) e “Tem tremores nas mãos?” (SRQ05). Ambos os itens apresentam valores iguais a 0.45 para a medida em questão, indicando que possuem correlação moderada com os outros itens.

Por outro lado, os itens com as maiores correlações são “Sente-se nervoso(a), tenso(a) ou preocupado(a)?” (SRQ06), “Tem dificuldade de pensar com clareza?” (SRQ08), “Tem se sentido triste ultimamente?” (SRQ09), “Encontra dificuldades para realizar com satisfação suas atividades diárias?” (SRQ11), “Você se sente uma pessoa inútil, sem préstimo?” (SRQ16) e “Sente-se cansado(a) o tempo todo?” (SRQ18). Esses itens apresentam correlações iguais a 0.75, 0.74, 0.71, 0.70 e 0.73, respectivamente. Esses valores indicam que esses itens têm uma correlação alta com os outros itens do questionário, sugerindo que tais itens medem características semelhantes ou iguais ao restante das questões.

De modo geral, com as análises mostradas, verificou-se que os itens do questionário apresentam uma boa consistência interna.

## 2.3 Avaliação da dimensionalidade

Uma das suposições básicas dos modelos do TRI é a unidimensionalidade, que se refere à ideia de que há uma aptidão dominante (um fator ou traço latente dominante) responsável pelo desempenho de um indivíduo nos itens de um teste. É essa aptidão que o teste busca medir. Esse pré-requisito é crucial, pois, até o momento, o TRI ainda não dispõe de soluções adequadas para modelos multidimensionais. Para mais detalhar, veja Pasquali & Primi (2009).

Assim, uma análise de dimensionalidade é apresentada a seguir com o objetivo de verificar se o conjunto de questões do nosso estudo mede uma única característica, de modo a inferir se a saúde mental dos alunos de graduação é caracterizada por um único traço latente. Para isso, desenvolveremos uma análise fatorial usando o pacote *psych* e especificamente a função **fa** (Revelle, in preparation), que implementa uma análise fatorial exploratória utilizando o método MinRes (mínimo residual). Este método utiliza os Mínimos Quadrados Ordinários (OLS) para encontrar a solução residual mínima (minres), resultando em soluções muito semelhantes às obtidas pela máxima verossimilhança, mesmo para matrizes mal-comportadas. Ver Harman e Jones (1966).

### 2.3.1 Análise fatorial utilizando matriz de correlação tetracórica

Para avaliar a unidimensionalidade do teste, é realizada uma análise fatorial utilizando a matriz de correlações tetracóricas. Essa abordagem foi escolhida porque, ao utilizar testes ou escalas com respostas dicotômicas, como no nosso estudo, ou politômicas, a aplicação da matriz de correlação de Pearson tradicional para a análise fatorial não é válida. No caso de dados dicotômicos, devem ser consideradas correlações tetracóricas, enquanto para dados politômicos, devem ser utilizadas correlações policóricas, conforme discutido em Revelle (2022).



Primeiramente, foi aplicado essa metodologia utilizando um único fator. Os resultados são apresentados a seguir:

```
c<-ord2
fap1 <- fa(r=c$rho) #Análise fatorial utilizando a matriz de correlação tetracórica
fap1

## Factor Analysis using method =  minres
## Call: fa(r = c$rho)
## Standardized loadings (pattern matrix) based upon correlation matrix
##          MR1    h2    u2 com
## SRQ01 0.54 0.29 0.71    1
## SRQ02 0.54 0.29 0.71    1
## SRQ03 0.57 0.32 0.68    1
## SRQ04 0.46 0.21 0.79    1
## SRQ05 0.46 0.21 0.79    1
## SRQ06 0.78 0.60 0.40    1
## SRQ07 0.59 0.35 0.65    1
## SRQ08 0.76 0.58 0.42    1
## SRQ09 0.82 0.67 0.33    1
## SRQ10 0.61 0.38 0.62    1
## SRQ11 0.74 0.55 0.45    1
## SRQ12 0.59 0.35 0.65    1
## SRQ13 0.72 0.52 0.48    1
## SRQ14 0.69 0.47 0.53    1
## SRQ15 0.71 0.50 0.50    1
## SRQ16 0.73 0.54 0.46    1
## SRQ17 0.71 0.51 0.49    1
## SRQ18 0.75 0.57 0.43    1
## SRQ19 0.69 0.48 0.52    1
## SRQ20 0.59 0.35 0.65    1
##
##          MR1
## SS loadings    8.74
## Proportion Var 0.44
##
## Mean item complexity =  1
## Test of the hypothesis that 1 factor is sufficient.
##
## df null model =  190  with the objective function =  13.15
## df of the model are 170  and the objective function was  4
##
## The root mean square of the residuals (RMSR) is  0.08
## The df corrected root mean square of the residuals is  0.08
##
## Fit based upon off diagonal values = 0.97
## Measures of factor score adequacy
##
##          MR1
## Correlation of (regression) scores with factors  0.97
## Multiple R square of scores with factors        0.95
## Minimum correlation of possible factor scores    0.89
```

Na coluna MR1, são apresentadas as cargas fatoriais de cada item assumindo um único fator, indicando o quanto cada item contribui para o fator latente subjacente aos itens. Os valores na coluna *h2* correspondem

à quantidade de variância explicada pelo fator (retido) em cada item e representa a soma dos quadrados das cargas fatoriais, também conhecida como comunalidade. Enquanto que  $u^2 := 1 - h^2$  representa a variância residual, também chamada de unicidade. Na coluna *com* é apresentada a complexidade de Hoffman de cada item, que é igual a 1 se um item carrega apenas em um fator, 2 se carrega uniformemente em dois fatores, e assim por diante. Esse valor indica o quanto um item reflete um único construto, sendo menor para itens com cargas fatoriais relativamente baixas. Consulte Pettersson e Turkheimer (2010) para mais detalhes.

Com os resultados obtidos, é possível verificar que as cargas fatoriais entre os itens e o fator está entre 0.46 e 0.82, o que indica que a associação entre as questões do teste e o fator é moderada ou forte. Também notamos que o índice de complexidade de todos os itens é igual a 1, sinalizando que uma única dimensão seria suficiente para explicar a variabilidade dos itens. Desse modo, há indícios de que os itens medem uma mesma característica.

Analisando mais detalhadamente, temos que as questões 4 e 5 são as que menos se relacionam com o primeiro fator, com ambas apresentando um valor igual a 0.46. Esses itens se referem, respectivamente, a “Assusta-se com facilidade?” (SRQ04) e “Tem tremores nas mãos?” (SRQ05). Além disso, esses itens apresentam a menor comunalidade (0.21) e a maior variância residual (0.79), indicando que estes itens são os menos associados com o fator ou dimensão subjacente aos outros itens.

Por outro lado, o item 9, “Tem se sentido triste ultimamente?” (SRQ09), apresenta um valor expressivo de carga fatorial igual a 0.82, indicando uma forte associação com o primeiro fator. Consequentemente, esse item possui a maior comunalidade (0.67) e a menor variância residual (0.33).

No sumário, são apresentadas a soma dos quadrados das cargas fatoriais, bem como a proporção de variância, que indica quanto da variância total dos itens é explicada pelo fator.

Neste caso, observa-se que o fator (uma dimensão) explica cerca de 44% da variabilidade dos dados, um valor relevante. Além disso, analisando outros resultados apresentados, encontramos uma correlação entre os escores e o fator de 0.97, um valor significativo e que reforça a sugestão de unidimensionalidade do teste.

A seguir, é avaliada a possibilidade de o teste apresentar dois fatores, ou seja, de medir duas características distintas. Para isso, realizamos a mesma análise fatorial, desta vez considerando dois possíveis fatores, e apresentamos a solução utilizando a rotação *varimax*. Em seguida, será feita uma comparação entre os resultados obtidos com um único fator e com dois fatores.

Adicionalmente, o pacote *psych* apresenta os índices de ajuste no estilo de Modelagem de Equações Estruturais (*Structural Equation Modeling*, SEM), geralmente utilizados em análises fatoriais confirmatórias, e que, neste estudo, será utilizada para representar graficamente os resultados da análise fatorial exploratória de dois fatores. Mais informações podem ser encontradas na documentação do R com `?factor.stats`.

```
fap2 <- fa(r=c$rho,nfactors=2,rotate="varimax")
fap2
```

```
## Factor Analysis using method = minres
## Call: fa(r = c$rho, nfactors = 2, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1  MR2  h2  u2 com
## SRQ01 0.25 0.57 0.39 0.61 1.4
## SRQ02 0.41 0.35 0.29 0.71 2.0
## SRQ03 0.43 0.36 0.32 0.68 1.9
## SRQ04 0.27 0.40 0.24 0.76 1.8
## SRQ05 0.28 0.39 0.23 0.77 1.8
## SRQ06 0.55 0.56 0.62 0.38 2.0
## SRQ07 0.20 0.76 0.62 0.38 1.1
## SRQ08 0.61 0.46 0.58 0.42 1.9
## SRQ09 0.77 0.33 0.71 0.29 1.3
## SRQ10 0.48 0.37 0.37 0.63 1.9
```

```

## SRQ11 0.75 0.23 0.62 0.38 1.2
## SRQ12 0.50 0.32 0.35 0.65 1.7
## SRQ13 0.64 0.35 0.53 0.47 1.6
## SRQ14 0.76 0.14 0.60 0.40 1.1
## SRQ15 0.73 0.21 0.57 0.43 1.2
## SRQ16 0.75 0.22 0.62 0.38 1.2
## SRQ17 0.69 0.27 0.55 0.45 1.3
## SRQ18 0.57 0.49 0.57 0.43 2.0
## SRQ19 0.56 0.40 0.48 0.52 1.8
## SRQ20 0.11 0.91 0.83 0.17 1.0
##
##
##          MR1  MR2
## SS loadings      6.14 3.92
## Proportion Var    0.31 0.20
## Cumulative Var    0.31 0.50
## Proportion Explained 0.61 0.39
## Cumulative Proportion 0.61 1.00
##
## Mean item complexity = 1.5
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model = 190 with the objective function = 13.15
## df of the model are 151 and the objective function was 2.77
##
## The root mean square of the residuals (RMSR) is 0.05
## The df corrected root mean square of the residuals is 0.06
##
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##
##          MR1  MR2
## Correlation of (regression) scores with factors 0.95 0.94
## Multiple R square of scores with factors        0.90 0.89
## Minimum correlation of possible factor scores    0.80 0.77

```

É possível verificar que as cargas fatoriais entre os itens e o primeiro fator variam entre 0.11 e 0.76, enquanto no segundo fator estão entre 0.14 e 0.91. Considerando cargas fatoriais superiores a 0.40, podemos afirmar que os itens 2, 3, e 9, 10, 11, 12, 13, 14, 15, 16 e 17 se associam com o fator 1, enquanto os itens 1, 4, 7 e 20 se associam com o fator 2. Ainda, os itens 6, 8, 18 e 19 estão associados com ambos os fatores, ao passo que o item 5 está associado a nenhum fator. Essa análise pode ser visualizada na figura que apresenta a solução de dois fatores.

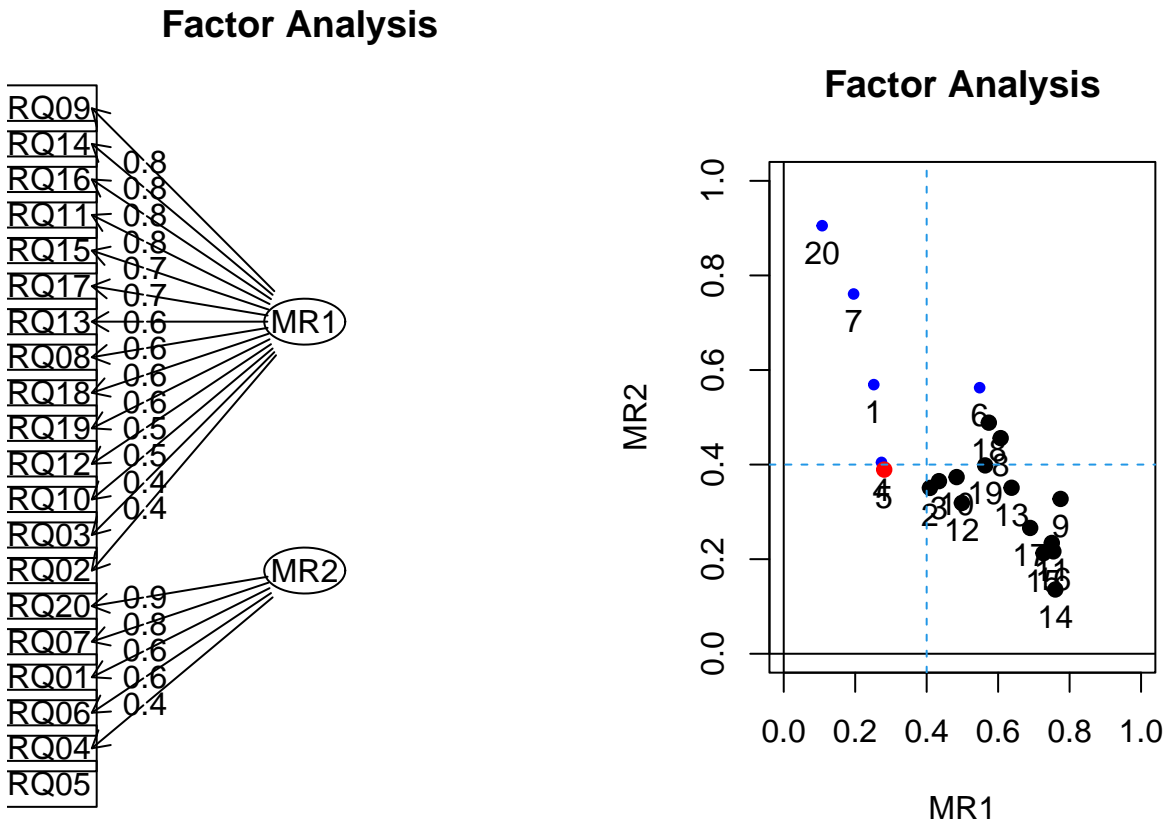
Assim como na solução de um fator, observamos que as menores comunalidades, e consequentemente as maiores variâncias residuais, ocorrem nos itens 4 e 5. Além disso, encontramos que a complexidade do item é igual a 2 para os itens 2, 6 e 18, o que justifica, possivelmente, uma solução de dois fatores para esses itens. No entanto, como a média da complexidade dos itens é 1.5, concluímos que, de maneira geral, não é viável uma solução de dois fatores.

Analisando os resultados de forma geral, temos que o primeiro fator explica 31% da variância dos dados, enquanto o segundo fator explica 20%. Assim, em conjunto os fatores explicam cerca de 50% da variância dos dados, um valor que não se distancia significativamente de 44%, valor obtido anteriormente com a análise fatorial de fator único.

Na solução de dois fatores, o fator 1 apresenta uma correlação entre os escores e o fator igual a 0.95, enquanto o fator 2 apresenta correlação igual a 0.94, valores que são relevantes, porém menores que o valor de 0.97 encontrado na análise de fator único. Além disso, observa-se que o valor de RMSR é 0.05, um pouco inferior ao valor de 0.08 obtido na solução de um único fator.

Na seguinte figura, é mostrado graficamente a solução de dois fatores.

```
par(mfrow=c(1,2))
fa.diagram(fap2,cut=0.4)
factor.plot(fap2, cut=0.4,ylim=c(0,1),xlim=c(0,1))
abline(h=0.4,col=4,lty=2)
abline(v=0.4,col=4,lty=2)
```



Em geral, podemos observar que a solução de dois fatores não parece ser suficientemente convincente quando comparada com a solução de um único fator. Isso se deve ao fato de que o possível segundo fator está formado por poucos itens, e o item 5 não se encaixaria em nenhum dos fatores.

Além da análise estatística, Bazán (2022) enfatiza que, dado que a unidimensionalidade é a suposição central em modelos baseados na Teoria de Resposta ao Item (TRI), a decisão sobre a violação dessa suposição não deve ser feita com base em um único critério. Pelo contrário, é crucial levar em conta múltiplos critérios ao avaliar a unidimensionalidade. Alguns autores defendem que um modelo pode ser considerado unidimensional quando pelo menos um desses critérios é atendido, enquanto é classificado como multidimensional quando vários critérios de unidimensionalidade são simultaneamente violados.

Além disso, Bazán (2022) aponta que a análise fatorial pode indicar a presença de um ou mais fatores, mas a existência de mais de um fator não implica necessariamente que o teste seja multidimensional. Nesse contexto, a literatura sugere que um teste é considerado unidimensional se ao menos dois dos três seguintes critérios forem atendidos:

- O primeiro fator explica pelo menos 40% da variância (Carmines & Zeller, 1979);
- O número de itens com cargas fatoriais superiores a 0,35 no primeiro fator deve ser superior a 80% (Orlando et al., 2000);

- A raiz do erro quadrático médio (Root Mean Square Residual, RMSR) deve ser próxima de 0.

Considerando esses critérios, observamos que a solução de um fator satisfaz os dois primeiros critérios, enquanto a solução de dois fatores atende apenas ao último critério.

Portanto, considerando as análises realizadas, temos que os dados fornecem evidências de que o questionário SRQ20 é unidimensional, desse modo, o teste mede um único traço latente dos indivíduos.

### 3 Teoria de Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI), também conhecida como modelo de resposta ao item na Psicometria, apresenta como principal interesse saber se um indivíduo assinala um determinado item (correto ou não) em um teste, ao invés de considerar apenas a sua pontuação total. A TRI tem o interesse em determinar como a variável latente de um indivíduo (como habilidade, conhecimento, aptidão ou proficiência) está relacionada com as respostas dadas nos itens de um teste. Isso é feito por meio de modelos probabilísticos, que tentam descrever a relação entre esse traço latente e a probabilidade de uma pessoa responder corretamente a um item (Bazán 2018). Em outras palavras, o modelo tenta explicar como algo não diretamente mensurável (traço latente) influencia o que pode ser medido ou observado (respostas do teste). Ainda, o TRI diz que a relação entre o desempenho no item e o conjunto dos traços latentes pode ser descrita pela equação monotônica Curva Característica do Item (CCI), na qual indivíduos de maior proficiência terão maior probabilidade de responder corretamente o item (Pasquali & Primi 2009).

#### 3.1 Modelo Geral TRI

Utilizando a definição de Bazán et al. (2010), considere  $n$  sujeitos avaliados em um teste com  $k$  itens. Um modelo da Teoria de Resposta ao Item (TRI), ou simplesmente modelo TRI, binário unidimensional (2L), é um sistema no qual, para cada sujeito  $i$ , tem-se um modelo de variável latente monotônico unidimensional  $(\mathbf{Y}, \theta_i)$ , definido pelas expressões:

$$\begin{aligned} Y_{ij} \mid \theta_i, \eta_j &\sim \text{Bernoulli}(p_{ij}) \\ p_{ij} &= P(Y_{ij} = 1 \mid \theta_i, \eta_j) = F(m_{ij}) \\ m_{ij} &= a_j(\theta_i - b_j), \\ i &= 1, \dots, n, \quad j = 1, \dots, k \end{aligned}$$

onde:

- $Y_{ij}$  é a variável manifesta que modela a resposta binária do sujeito  $i$  ao item  $j$ .
- $\eta_j = (a_j, b_j)$  são os parâmetros que representam, respectivamente, a discriminação ( $a_j$ ) e a dificuldade ( $b_j$ ) do item  $j$ .
- $\theta_i$  é o valor da variável latente para o sujeito  $i$ , interpretada como a habilidade do sujeito  $i$ .
- $p_{ij}$  é a probabilidade condicional, dada  $\theta_i$ .
- $F$  é denominada a curva característica do item (CCI).
- $m_{ij}$  é um preditor latente linear em relação à habilidade do sujeito  $i$ .

Em um modelo TRI binário, a densidade conjunta do vetor de respostas multivariadas  $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)'$ , com  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik})$ , dado o vetor de variáveis latentes  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$  e o vetor de parâmetros dos itens  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)'$ , pode ser escrita como:

$$f(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^k F(m_{ij})^{y_{ij}} (1 - F(m_{ij}))^{1-y_{ij}}$$

em que  $\theta_i$  representa o valor da variável latente para o indivíduo  $i$ , sendo um escore originalmente expresso em valores contínuos em uma escala aproximada de -3 a +3.

Neste estudo, podemos interpretar o escore  $\theta_i$  como uma medida do sofrimento mental. O parâmetro de dificuldade  $b_j$  reflete a severidade do sintoma abordado pelo item  $j$ , enquanto o parâmetro de discriminação  $a_j$  indica a capacidade do item  $j$  de diferenciar respondentes com diferentes níveis de sofrimento mental.

## 3.2 Casos Particulares

O modelo geral admite diversas formulações, que dependem basicamente de como a CCI é considerada. Na sua versão mais simples, podemos tomar  $a_j = 1$  e considerar uma CCI da forma:

$$P(Y_{ij} = 1 \mid \theta_i, b_j) = F(\theta_i - b_j)$$

Este é chamado de modelo TRI de um parâmetro (1L).

De forma mais geral, podemos considerar uma CCI da forma:

$$P(Y_{ij} = 1 \mid \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) F(a_j (\theta_i - b_j))$$

onde o parâmetro  $c_j \in [0, 1]$  (também representado como  $g_j$ ), conhecido como parâmetro de *guessing*, representa a probabilidade de um indivíduo acertar ou endossar um item  $j$  por acaso, e  $F$  é uma função de distribuição. Este é conhecido como o modelo TRI de três parâmetros (3L). Se considerarmos  $c_j = 0$ , o modelo se reduz ao modelo TRI de dois parâmetros (2L). Adicionalmente, se assumirmos  $a_j = 1$ , obtemos o modelo TRI de um parâmetro (1L), previamente apresentado.

## 3.3 Aplicação dos modelos TRI 1L, 2L e 3L

Como dito anteriormente, os modelos de teoria de resposta (TRI) têm como suposição a unidimensionalidade dos itens. Dessa forma, como o questionário foi considerado unidimensional, podemos aplicar os modelos TRI de um parâmetro (1L), dois parâmetros (2L) e três parâmetros (3L).

### 3.3.1 Modelo 1L

Primeiramente ajustamos o modelo 1L usando o pacote *mirt*.

```
#Estimation One Logistic IRT Model 1L usando mirt
model1L <- mirt(df_srq,model = 1, itemtype = "Rasch", SE = TRUE)
```

```
print(model1L)
```

```
##
## Call:
## mirt(data = df_srq, model = 1, itemtype = "Rasch", SE = TRUE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 16 EM iterations.
## mirt version: 1.41
## M-step optimizer: nlminb
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
```

```
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix = 9.045158
##
## Log-likelihood = -8803.857
## Estimated parameters: 21
## AIC = 17649.71
## BIC = 17748.69; SABIC = 17682
## G2 (1048554) = 6902.06, p = 1
## RMSEA = 0, CFI = NaN, TLI = NaN
```

Inicialmente, são apresentadas as estatísticas de ajuste do modelo, que corresponde a uma análise fatorial de itens com informação completa, utilizando o algoritmo EM para estimar os parâmetros dos itens, assumindo normalidade para o traço latente. São fornecidos detalhes sobre a convergência do modelo (convergiu dentro da tolerância de  $1 \times 10^{-4}$ ), o procedimento de aceleração do algoritmo EM (16 iterações), o número de pontos de quadratura usados para obter a verossimilhança marginal integrando sobre o traço latente (61), o método empregado para estimar a matriz de informação (Oakes), além do valor da verossimilhança (-8803.857).

Também são reportados o número de parâmetros estimados (21), bem como as estatísticas de comparação de modelos: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), G2 (Goodness-of-Fit Chi-Square) e RMSEA (Root Mean Square Error of Approximation, ou Erro Quadrático Médio da Aproximação), com os respectivos valores de 17649.71, 17748.69, 6902.06 e 0.

Essas medidas também serão apresentadas para os demais modelos.

A seguir, as estimativas dos parâmetros dos itens.

```
coef(model1L,simplify = TRUE, IRTpar = TRUE)$items
```

```
##      a      b g u
## SRQ01 1 0.3871226 0 1
## SRQ02 1 0.9037510 0 1
## SRQ03 1 -0.5964636 0 1
## SRQ04 1 0.5115714 0 1
## SRQ05 1 1.1839718 0 1
## SRQ06 1 -2.4798236 0 1
## SRQ07 1 0.8744046 0 1
## SRQ08 1 -0.1447368 0 1
## SRQ09 1 -0.7550995 0 1
## SRQ10 1 1.0835196 0 1
## SRQ11 1 -1.3126748 0 1
## SRQ12 1 -0.8807146 0 1
## SRQ13 1 0.5254872 0 1
## SRQ14 1 0.9778451 0 1
## SRQ15 1 -0.5822293 0 1
## SRQ16 1 0.4215613 0 1
## SRQ17 1 2.0074265 0 1
## SRQ18 1 -1.0172607 0 1
## SRQ19 1 -1.3126748 0 1
## SRQ20 1 0.4907338 0 1
```

Neste modelo, como  $a_j = 1$ , este parâmetro não é estimado, sendo estimados os parâmetros  $b_j$ , que correspondem à dificuldade do item. Note também que os valores  $g_j = 0$  e  $u_j = 1$  são apresentados. O parâmetro

$g$ , denominado *guessing* no pacote *mirt*, corresponde ao parâmetro  $c$  definido anteriormente, que para o modelo 1L é igual a 0. O outro parâmetro,  $u$ , está relacionado com a assíntota superior da curva da CCI. Nos modelos 1L, 2L e 3L, o valor de  $u$  é igual a 1, o que indica que a probabilidade de responder corretamente a qualquer item pode atingir 1. Para mais detalhes sobre o pacote *mirt*, consulte Chalmers (2012).

A seguir, utilizamos um segundo pacote para estimar o modelo 1L: o pacote *ltm* (Rizopoulos, 2006)[9], que realiza a estimativa dos parâmetros dos itens por meio de máxima verossimilhança marginal utilizando o algoritmo EM, com pontos de quadratura Gauss-Hermite.

```
#Estimation One Logistic IRT Model 1L usando ltm
fit1L <- ltm::rasch(df_srq, IRT.param = TRUE)
summary(fit1L)
```

```
##
## Call:
## ltm::rasch(data = df_srq, IRT.param = TRUE)
##
## Model Summary:
##      log.Lik      AIC      BIC
## -8803.916 17649.83 17748.8
##
## Coefficients:
##              value std.err  z.vals
## Dffclt.SRQ01  0.2650  0.0665   3.9856
## Dffclt.SRQ02  0.6141  0.0696   8.8282
## Dffclt.SRQ03 -0.3994  0.0682  -5.8603
## Dffclt.SRQ04  0.3492  0.0670   5.2137
## Dffclt.SRQ05  0.8033  0.0724  11.1025
## Dffclt.SRQ06 -1.6715  0.0968 -17.2760
## Dffclt.SRQ07  0.5943  0.0693   8.5739
## Dffclt.SRQ08 -0.0943  0.0662  -1.4244
## Dffclt.SRQ09 -0.5065  0.0693  -7.3063
## Dffclt.SRQ10  0.7357  0.0713  10.3236
## Dffclt.SRQ11 -0.8834  0.0753 -11.7356
## Dffclt.SRQ12 -0.5913  0.0704  -8.3977
## Dffclt.SRQ13  0.3587  0.0671   5.3503
## Dffclt.SRQ14  0.6643  0.0702   9.4597
## Dffclt.SRQ15 -0.3901  0.0681  -5.7309
## Dffclt.SRQ16  0.2884  0.0666   4.3286
## Dffclt.SRQ17  1.3592  0.0851  15.9786
## Dffclt.SRQ18 -0.6835  0.0718  -9.5236
## Dffclt.SRQ19 -0.8833  0.0753 -11.7339
## Dffclt.SRQ20  0.3352  0.0669   5.0108
## Dscrmn        1.4806  0.0470  31.4796
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.065
## quasi-Newton: BFGS
```

Embora a formulação padrão do modelo *Rasch* assuma que o parâmetro de discriminação seja fixado em



1, a função `rasch()` o estima. Para obter valores equivalentes às estimativas das dificuldades obtidas pelo *mirt*, deve-se multiplicar o valor estimado da discriminação, que neste caso é  $a = 1.481$ , pelo valor estimado da dificuldade. Por exemplo, a dificuldade estimada no *ltm* para o item 1 é  $1.481 \times 0.265 = 0.392$ , o que é próximo do valor da dificuldade estimada para o item 1 no *mirt*, que é 0.387.

As diferenças entre os valores estimados devem-se aos diferentes métodos utilizados no processo de otimização e convergência. Para mais detalhes, consulte Rizopoulos e Moustaki (2008).

Esse mesmo resultado, porém com o sinal invertido, pode ser obtido a seguir.

```
fit1La<- ltm::rasch(df_srq, IRT.param = FALSE)
summary(fit1La)
```

```
##
## Call:
## ltm::rasch(data = df_srq, IRT.param = FALSE)
##
## Model Summary:
##      log.Lik      AIC      BIC
## -8803.916 17649.83 17748.8
##
## Coefficients:
##           value std.err  z.vals
## SRQ01 -0.3924  0.0982 -3.9973
## SRQ02 -0.9092  0.1007 -9.0253
## SRQ03  0.5914  0.0994  5.9524
## SRQ04 -0.5171  0.0986 -5.2456
## SRQ05 -1.1893  0.1032 -11.5283
## SRQ06  2.4748  0.1275 19.4060
## SRQ07 -0.8799  0.1005 -8.7525
## SRQ08  0.1396  0.0978  1.4267
## SRQ09  0.7499  0.1003  7.4772
## SRQ10 -1.0893  0.1022 -10.6570
## SRQ11  1.3080  0.1054 12.4079
## SRQ12  0.8755  0.1012  8.6517
## SRQ13 -0.5312  0.0986 -5.3853
## SRQ14 -0.9835  0.1013 -9.7084
## SRQ15  0.5776  0.0993  5.8174
## SRQ16 -0.4269  0.0983 -4.3447
## SRQ17 -2.0124  0.1153 -17.4470
## SRQ18  1.0119  0.1023  9.8886
## SRQ19  1.3077  0.1054 12.4059
## SRQ20 -0.4963  0.0985 -5.0384
## z      1.4806  0.0470 31.4796
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.065
## quasi-Newton: BFGS
```

O resultado mostra os valores estimados do parâmetro de dificuldade com sinal modificado.

Outra possibilidade recomendada na documentação da função *ltm* é usar um contraste, de modo a assumir que a discriminação é igual a 1. Essa abordagem é apresentada a seguir.

```
fit1Lb=rasch(df_srq,constraint=cbind(length(df_srq)+1,1))
summary(fit1Lb)
```

```
##
## Call:
## rasch(data = df_srq, constraint = cbind(length(df_srq) + 1, 1))
##
## Model Summary:
##      log.Lik      AIC      BIC
## -8882.635 17805.27 17899.53
##
## Coefficients:
##              value std.err  z.vals
## Dffclt.SRQ01  0.3540  0.0875   4.0468
## Dffclt.SRQ02  0.8357  0.0901   9.2745
## Dffclt.SRQ03 -0.5608  0.0886  -6.3276
## Dffclt.SRQ04  0.4700  0.0879   5.3466
## Dffclt.SRQ05  1.0972  0.0926  11.8535
## Dffclt.SRQ06 -2.3012  0.1162 -19.7991
## Dffclt.SRQ07  0.8084  0.0899   8.9919
## Dffclt.SRQ08 -0.1413  0.0871  -1.6213
## Dffclt.SRQ09 -0.7079  0.0896  -7.9034
## Dffclt.SRQ10  1.0034  0.0916  10.9547
## Dffclt.SRQ11 -1.2243  0.0946 -12.9380
## Dffclt.SRQ12 -0.8243  0.0905  -9.1129
## Dffclt.SRQ13  0.4829  0.0880   5.4905
## Dffclt.SRQ14  0.9049  0.0907   9.9780
## Dffclt.SRQ15 -0.5476  0.0886  -6.1831
## Dffclt.SRQ16  0.3860  0.0876   4.4077
## Dffclt.SRQ17  1.8659  0.1047  17.8138
## Dffclt.SRQ18 -0.9508  0.0916 -10.3810
## Dffclt.SRQ19 -1.2242  0.0946 -12.9376
## Dffclt.SRQ20  0.4505  0.0878   5.1302
## Dscrmn        1.0000      NA      NA
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.023
## quasi-Newton: BFGS
```

O resultado mostra que os valores estimados para o parâmetro de dificuldade são semelhantes aos obtidos pelo *mirt*, porém com diferenças em algumas casas decimais. Considerando essas variações na forma de apresentar os resultados no *ltm*, preferimos basear nossas interpretações na saída do *mirt*.

Ao ajustar o modelo 1L, observa-se que as estimativas dos parâmetros de dificuldade (*b*) das questões do questionário SQR usando *mirt* variaram de -2.480 a 2.007, correspondendo, respectivamente, aos itens 6 e 17. Em geral, podemos pensar que o parâmetro de dificuldade se refere ao grau de severidade do sintoma,

quando negativo significa que o sintoma é menos severo e quanto mais positivo significa que o sintoma é mais severo. Assim, o item 6, que trata da questão “Sente-se nervoso(a), tenso(a) ou preocupado(a)?” (SRQ06), representa um sintoma de menor severidade entre os estudantes. Em contraste, o item 17, que aborda a questão “Tem tido ideia de acabar com a vida?” (SRQ17), reflete o sintoma de maior severidade.

### 3.3.2 Modelo 2L

No modelo 2L, além do parâmetro de dificuldade ( $b$ ) também existe o parâmetro discriminação ( $a$ ). O parâmetro de discriminação ( $a > 0$ ) mede a eficácia de um item em diferenciar respondentes com diferentes níveis do traço latente avaliado pela questão.

A seguir usamos o *mirt* para ajustar o modelo de três parâmetros.

```
#Estimation two Logistic IRT Model 2L usando mirt
model2L<- mirt(df_srq, itemtype = "2PL", SE = TRUE)
```

```
print(model2L)
```

```
##
## Call:
## mirt(data = df_srq, itemtype = "2PL", SE = TRUE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 26 EM iterations.
## mirt version: 1.41
## M-step optimizer: BFGS
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix = 21.56433
##
## Log-likelihood = -8702.742
## Estimated parameters: 40
## AIC = 17485.48
## BIC = 17674; SABIC = 17546.98
## G2 (1048535) = 6699.83, p = 1
## RMSEA = 0, CFI = NaN, TLI = NaN
```

```
coef(model2L,simplify = TRUE, IRTpar = TRUE)$items
```

```
##           a           b g u
## SRQ01 1.0227515 0.31490072 0 1
## SRQ02 1.0895522 0.72503982 0 1
## SRQ03 1.1657768 -0.47641706 0 1
## SRQ04 0.8328085 0.49026943 0 1
## SRQ05 0.8702593 1.12422110 0 1
## SRQ06 2.2364971 -1.37371601 0 1
## SRQ07 1.2407232 0.64394137 0 1
## SRQ08 2.0267448 -0.09631608 0 1
```

```
## SRQ09 2.5283602 -0.41514910 0 1
## SRQ10 1.3536200 0.76084166 0 1
## SRQ11 2.0237793 -0.76858983 0 1
## SRQ12 1.2614669 -0.66250835 0 1
## SRQ13 1.8157509 0.30894110 0 1
## SRQ14 1.7208512 0.60197531 0 1
## SRQ15 1.7934201 -0.36794103 0 1
## SRQ16 1.9101398 0.23960013 0 1
## SRQ17 1.7721683 1.23539158 0 1
## SRQ18 2.0790520 -0.59327926 0 1
## SRQ19 1.7583556 -0.81755025 0 1
## SRQ20 1.1991894 0.36236291 0 1
```

Os resultados apresentam as estimativas dos parâmetros de discriminação e dificuldade.

Adicionalmente, as estimativas também podem ser obtidas utilizando o *ltm*, como mostrado a seguir.

```
#Estimation two Logistic IRT Model 2L usando ltm
fit2L=ltm(df_srq~z1,IRT.param=TRUE)
summary(fit2L)
```

```
##
## Call:
## ltm(formula = df_srq ~ z1, IRT.param = TRUE)
##
## Model Summary:
##      log.Lik      AIC      BIC
## -8702.289 17484.58 17673.1
##
## Coefficients:
##              value std.err  z.vals
## Dffc1t.SRQ01  0.3198  0.0862   3.7109
## Dffc1t.SRQ02  0.7294  0.0971   7.5094
## Dffc1t.SRQ03 -0.4732  0.0833  -5.6796
## Dffc1t.SRQ04  0.4950  0.1069   4.6315
## Dffc1t.SRQ05  1.1305  0.1439   7.8580
## Dffc1t.SRQ06 -1.3693  0.0937 -14.6090
## Dffc1t.SRQ07  0.6503  0.0856   7.5980
## Dffc1t.SRQ08 -0.0911  0.0564  -1.6169
## Dffc1t.SRQ09 -0.4101  0.0546  -7.5083
## Dffc1t.SRQ10  0.7663  0.0854   8.9708
## Dffc1t.SRQ11 -0.7621  0.0686 -11.1010
## Dffc1t.SRQ12 -0.6568  0.0851  -7.7202
## Dffc1t.SRQ13  0.3141  0.0611   5.1431
## Dffc1t.SRQ14  0.6055  0.0689   8.7901
## Dffc1t.SRQ15 -0.3622  0.0620  -5.8396
## Dffc1t.SRQ16  0.2443  0.0587   4.1645
## Dffc1t.SRQ17  1.2356  0.0963  12.8365
## Dffc1t.SRQ18 -0.5886  0.0626  -9.4005
## Dffc1t.SRQ19 -0.8137  0.0755 -10.7780
## Dffc1t.SRQ20  0.3684  0.0786   4.6854
## Dscrmn.SRQ01  1.0236  0.1064   9.6164
## Dscrmn.SRQ02  1.0909  0.1131   9.6479
## Dscrmn.SRQ03  1.1584  0.1143  10.1325
```

```
## Dscrmn.SRQ04 0.8335 0.0972 8.5711
## Dscrmn.SRQ05 0.8688 0.1037 8.3766
## Dscrmn.SRQ06 2.2332 0.2476 9.0209
## Dscrmn.SRQ07 1.2364 0.1221 10.1250
## Dscrmn.SRQ08 2.0240 0.1750 11.5633
## Dscrmn.SRQ09 2.5320 0.2307 10.9737
## Dscrmn.SRQ10 1.3522 0.1314 10.2905
## Dscrmn.SRQ11 2.0333 0.1906 10.6700
## Dscrmn.SRQ12 1.2633 0.1236 10.2202
## Dscrmn.SRQ13 1.8182 0.1606 11.3247
## Dscrmn.SRQ14 1.7305 0.1597 10.8353
## Dscrmn.SRQ15 1.8001 0.1608 11.1943
## Dscrmn.SRQ16 1.9217 0.1713 11.2195
## Dscrmn.SRQ17 1.7850 0.1883 9.4803
## Dscrmn.SRQ18 2.0758 0.1897 10.9440
## Dscrmn.SRQ19 1.7525 0.1646 10.6441
## Dscrmn.SRQ20 1.1942 0.1170 10.2041
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Convergence: 0
## max(|grad|): 0.014
## quasi-Newton: BFGS
```

Neste caso, ambos pacotes fornecem estimativas similares. Novamente, iremos basear nossas interpretações na saída do *mirt*.

Encontramos que as estimativas dos parâmetros de dificuldade ( $b$ ) do SQR-20 estão entre -1.3737 e 1.2354, que correspondem aos mesmos itens do modelo 1L, 6 e 17, respectivamente, possuindo a mesma interpretação.

Já, para o parâmetro de discriminação ( $a$ ), que avaliam quão bem uma questão discrimina (ou diferencia) entre indivíduos com pontuação alta e baixa do sofrimento mental, encontramos os valores entre 0.8328 e 2.5284, referente às questões 4 (“Assusta-se com facilidade?” (SRQ04)) e 9 (“Tem se sentido triste ultimamente?” (SRQ09)), respectivamente. Logo, os itens 4 e 9 podem ser considerados as perguntas menos e mais significativas do questionário para a identificação da gravidade do estado de saúde mental dos indivíduos.

### 3.3.3 Modelo 3L

No modelo 3L, além dos parâmetros de dificuldade ( $b$ ) e discriminação ( $a$ ), há a inclusão do parâmetro de pseudochute ( $c$ ), que representa a probabilidade de um respondente acertar ou endossar um item por acaso, especialmente quando não possui o traço latente avaliado pela questão. Vale lembrar que o parâmetro ( $g$ ) no pacote *mirt*, denominado *guessing*, corresponde ao parâmetro de pseudochute ( $c$ ).

```
#Estimation three Logistic IRT Model 3L usando mirt
#model3L <- mirt.model("rotation = 1 - 20")
#model3L.result <- mirt::mirt(df_srq,model3L, itemtype = "3PL", SE = TRUE)
model3L<- mirt(df_srq, itemtype = "3PL", SE = TRUE)
```

```
print(model3L)
```

```
##
## Call:
## mirt(data = df_srq, itemtype = "3PL", SE = TRUE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 109 EM iterations.
## mirt version: 1.41
## M-step optimizer: BFGS
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix = 399310.4
##
## Log-likelihood = -8700.626
## Estimated parameters: 60
## AIC = 17521.25
## BIC = 17804.03; SABIC = 17613.49
## G2 (1048515) = 6695.59, p = 1
## RMSEA = 0, CFI = NaN, TLI = NaN
```

```
coef(model3L,simplify = TRUE, IRTpar = TRUE)$items
```

```
##          a          b          g u
## SRQ01 1.1980247 0.48327443 7.338843e-02 1
## SRQ02 1.0952237 0.72453251 2.482779e-05 1
## SRQ03 1.1695310 -0.47170826 7.646943e-05 1
## SRQ04 0.8406493 0.48880914 3.393665e-05 1
## SRQ05 0.8811721 1.13186631 3.924969e-03 1
## SRQ06 2.2210007 -1.37722436 2.851864e-04 1
## SRQ07 1.2493199 0.64353809 1.929422e-05 1
## SRQ08 2.1747671 -0.04797848 2.575418e-02 1
## SRQ09 2.5220167 -0.40775574 3.281585e-05 1
## SRQ10 1.5251607 0.80100215 2.909521e-02 1
## SRQ11 2.1116176 -0.70829103 3.776531e-02 1
## SRQ12 1.4869401 -0.37333747 1.402676e-01 1
## SRQ13 1.8310546 0.31236599 1.471527e-04 1
## SRQ14 1.8840332 0.62952404 1.953031e-02 1
## SRQ15 1.9058772 -0.29810318 3.817858e-02 1
## SRQ16 1.9228964 0.24374115 3.032638e-05 1
## SRQ17 1.7853462 1.22972392 2.830119e-06 1
## SRQ18 2.0774363 -0.58711178 6.390992e-05 1
## SRQ19 1.7457461 -0.81563499 4.321781e-05 1
## SRQ20 1.2045877 0.36453589 9.114762e-05 1
```

Os resultados apresentam as estimativas dos parâmetros de discriminação, dificuldade e pseudochute.

Essas estimativas também podem ser obtidas utilizando o pacote *ltm*, como apresentado a seguir.

```
#Estimation three Logistic IRT Model 3L usando ltm
fit3L=tpm(df_srq,type="latent.trait",IRT.param=TRUE,max.guessing=1)
summary(fit3L)
```

```

##
## Call:
## tpm(data = df_srq, type = "latent.trait", max.guessing = 1, IRT.param = TRUE)
##
## Model Summary:
##      log.Lik      AIC      BIC
## -8700.155 17520.31 17803.09
##
## Coefficients:
##           value std.err  z.vals
## Gussng.SRQ01  0.0749  0.0772  0.9709
## Gussng.SRQ02  0.0000  0.0001  0.0024
## Gussng.SRQ03  0.0000  0.0004  0.0027
## Gussng.SRQ04  0.0000  0.0007  0.0094
## Gussng.SRQ05  0.0005    NaN    NaN
## Gussng.SRQ06  0.0001  0.0060  0.0124
## Gussng.SRQ07  0.0000  0.0000  0.0011
## Gussng.SRQ08  0.0248  0.0358  0.6911
## Gussng.SRQ09  0.0000  0.0000  0.0002
## Gussng.SRQ10  0.0296  0.0300  0.9868
## Gussng.SRQ11  0.0482  0.1078  0.4476
## Gussng.SRQ12  0.1369  0.1144  1.1967
## Gussng.SRQ13  0.0000  0.0003  0.0016
## Gussng.SRQ14  0.0195  0.0254  0.7692
## Gussng.SRQ15  0.0378  0.0637  0.5945
## Gussng.SRQ16  0.0000  0.0000  0.0003
## Gussng.SRQ17  0.0000  0.0000  0.0051
## Gussng.SRQ18  0.0000  0.0002  0.0012
## Gussng.SRQ19  0.0000  0.0000  0.0003
## Gussng.SRQ20  0.0000  0.0006  0.0032
## Dffclt.SRQ01  0.4909  0.1899  2.5848
## Dffclt.SRQ02  0.7290  0.0966  7.5446
## Dffclt.SRQ03 -0.4697  0.0835 -5.6284
## Dffclt.SRQ04  0.4932  0.1059  4.6579
## Dffclt.SRQ05  1.1324  0.1137  9.9604
## Dffclt.SRQ06 -1.3748  0.0948 -14.4958
## Dffclt.SRQ07  0.6500  0.0849  7.6518
## Dffclt.SRQ08 -0.0448  0.0802 -0.5583
## Dffclt.SRQ09 -0.4036  0.0550 -7.3384
## Dffclt.SRQ10  0.8072  0.0941  8.5785
## Dffclt.SRQ11 -0.6875  0.1713 -4.0135
## Dffclt.SRQ12 -0.3760  0.2471 -1.5218
## Dffclt.SRQ13  0.3173  0.0607  5.2269
## Dffclt.SRQ14  0.6334  0.0759  8.3418
## Dffclt.SRQ15 -0.2939  0.1215 -2.4186
## Dffclt.SRQ16  0.2483  0.0584  4.2536
## Dffclt.SRQ17  1.2298  0.0957 12.8571
## Dffclt.SRQ18 -0.5838  0.0630 -9.2681
## Dffclt.SRQ19 -0.8135  0.0763 -10.6638
## Dffclt.SRQ20  0.3702  0.0782  4.7316
## Dscrmn.SRQ01  1.2050  0.2517  4.7875
## Dscrmn.SRQ02  1.0954  0.1138  9.6226
## Dscrmn.SRQ03  1.1608  0.1144 10.1479
## Dscrmn.SRQ04  0.8405  0.0979  8.5890

```

```
## Dscrmn.SRQ05 0.8699      NaN      NaN
## Dscrmn.SRQ06 2.2164 0.2455  9.0294
## Dscrmn.SRQ07 1.2441 0.1233 10.0921
## Dscrmn.SRQ08 2.1614 0.2595  8.3294
## Dscrmn.SRQ09 2.5231 0.2301 10.9651
## Dscrmn.SRQ10 1.5255 0.2282  6.6839
## Dscrmn.SRQ11 2.1486 0.3434  6.2569
## Dscrmn.SRQ12 1.4790 0.2702  5.4733
## Dscrmn.SRQ13 1.8297 0.1619 11.3037
## Dscrmn.SRQ14 1.8937 0.2705  6.9998
## Dscrmn.SRQ15 1.9074 0.2628  7.2574
## Dscrmn.SRQ16 1.9333 0.1728 11.1849
## Dscrmn.SRQ17 1.7992 0.1903  9.4542
## Dscrmn.SRQ18 2.0706 0.1891 10.9509
## Dscrmn.SRQ19 1.7369 0.1634 10.6277
## Dscrmn.SRQ20 1.1988 0.1178 10.1743
##
## Integration:
## method: Gauss-Hermite
## quadrature points: 21
##
## Optimization:
## Optimizer: optim (BFGS)
## Convergence: 0
## max(|grad|): 0.018
```

Neste caso, ambos os pacotes fornecem estimativas semelhantes. No entanto, para alguns itens e parâmetros, o pacote *ltm* não calcula as estimativas dos desvios padrão. Novamente, iremos basear nossas interpretações na saída do *mirt*.

Por fim, para o modelo 3L temos que os valores de discriminação ( $a$ ) estão entre 0.8406 e 2.5220, correspondendo, respectivamente, às questões 4, “Assusta-se com facilidade?” (SRQ04), e 9, “Tem se sentido triste ultimamente?” (SRQ09). Logo, a questão 4 é o sintoma menos severo entre os estudantes de graduação, enquanto a questão 9 é o mais severo.

Sobre o parâmetro de dificuldade ( $b$ ), os valores se encontram entre -1.3772 e 1.2297, que correspondem, respectivamente, aos itens 6 e 17, assim como nos modelos 1L e 2L, possuindo a mesma interpretação.

Os valores do parâmetro  $g$  foram muito próximos de zero, com exceção dos itens 1 (Você tem dores de cabeça frequentes? (SRQ01)) e 12 (Tem dificuldade para tomar decisões? (SRQ12)), que possuem valores ligeiramente mais elevados. Isso sugere que, em geral, as respostas dos estudantes não apresentam evidências significativas do comportamento de pseudochute.

### 3.4 Escolha do modelo

A seguir, desenvolvemos um teste ANOVA para escolher o melhor modelo para os dados.

```
#Estimation three Logistic IRT Model 3L usando ltm
anova(fit1L,fit2L)

##
## Likelihood Ratio Table
##      AIC      BIC log.Lik    LRT df p.value
## fit1L 17649.83 17748.8 -8803.92
## fit2L 17484.58 17673.1 -8702.29 203.25 19 <0.001
```



```
anova(fit2L,fit3L)
```

```
##
## Likelihood Ratio Table
##           AIC          BIC   log.Lik  LRT df p.value
## fit2L 17484.58 17673.10 -8702.29
## fit3L 17520.31 17803.09 -8700.15 4.27 20      1
```

```
mirt::anova(model1L,model2L,model3L)
```

```
##           AIC      SABIC      HQ      BIC    logLik      X2 df p
## model1L 17649.72 17682.00 17687.68 17748.69 -8803.857
## model2L 17485.48 17546.98 17557.81 17674.00 -8702.742 202.231 19 0
## model3L 17521.25 17613.49 17629.74 17804.03 -8700.626   4.232 20 1
```

Os resultados indicam que há uma diferença significativa entre os modelos 1L e 2L. Contudo, os modelos 2L e 3L não apresentam diferenças estatisticamente significativas.

Com base nos valores do AIC e BIC, optamos pelo modelo 2L, pois ele apresenta os menores valores para essas métricas, indicando melhor ajuste aos dados.

## 4 Resultados do modelo escolhido

Para as seguintes etapas das análises, é mostrado os resultados obtidos usando o modelo 2L.

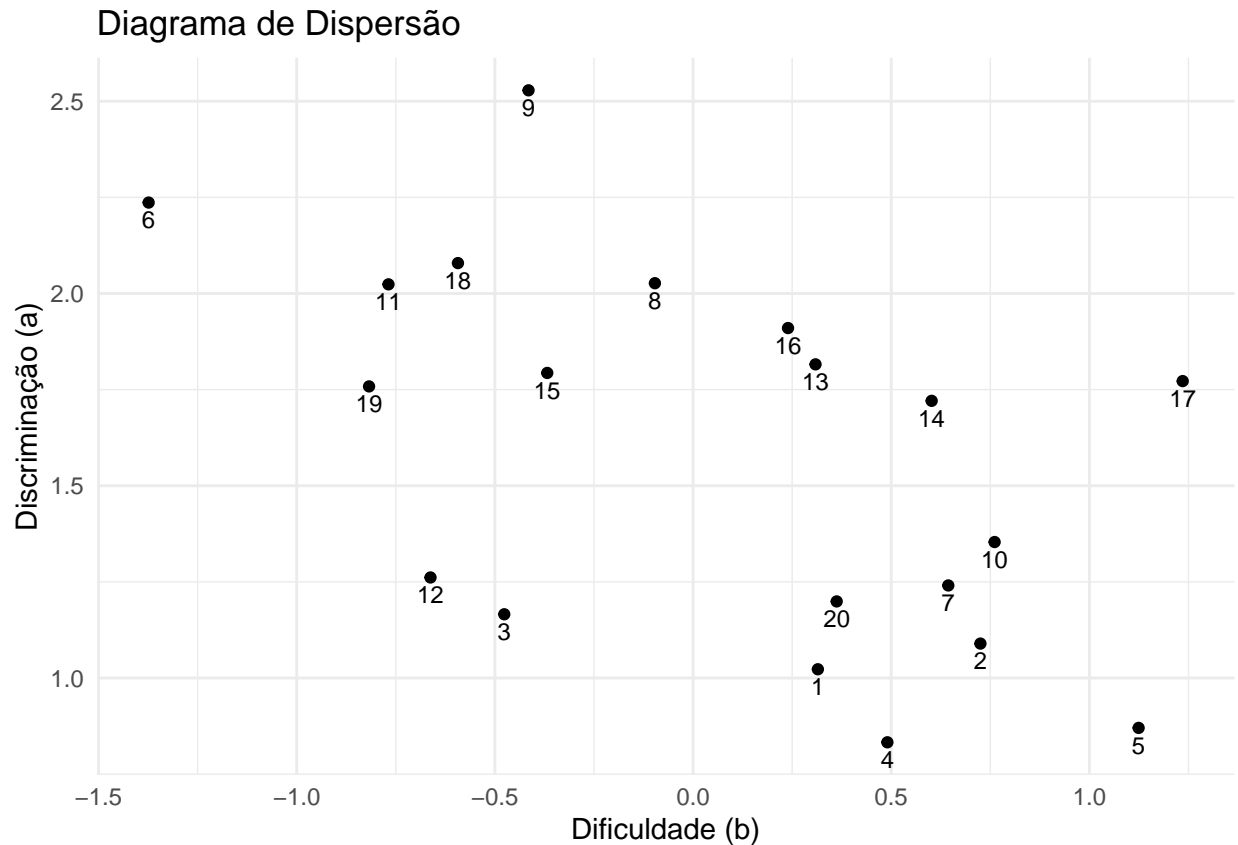
```
coef <- coef(model2L,simplify = TRUE, IRTpar = TRUE)$items

x <- coef[, 2]
y <- coef[, 1]
z <- 1:20

df <- data.frame(x = x, y = y, z = as.character(z))

scatter_plot <- ggplot(df, aes(x = x, y = y, label = z)) +
  geom_point() +
  geom_text(size = 3, vjust = 1.5) + # Adicionando os números de z aos pontos
  labs(x = "Dificuldade (b)", y = "Discriminação (a)", title = "Diagrama de Dispersão") +
  theme_minimal()

print(scatter_plot)
```



No diagrama de dispersão apresentado, temos as questões distribuídas pelos valores de dificuldade e discriminação estimados usando mirt.

Pelo gráfico conseguimos ver que os itens de maior dificuldade, ou seja, possuem maior grau de gravidade são as questões 17 (“Tem tido ideia de acabar com a vida?”) e 5 (“Tem tremores nas mãos?”). Os menos severos são os itens 6 (“Sente-se nervoso(a), tenso(a) ou preocupado(a)?”) e 19 (“Você se cansa com facilidade?”).

Os itens mais discriminantes (significativos) são 6 (“Sente-se nervoso(a), tenso(a) ou preocupado(a)?”) e 9 (“Tem se sentido triste ultimamente?”). Os menos significativos são 4 (“Assusta-se com facilidade?”) e 5 (“Tem tremores nas mãos?”). Dessa forma, vemos que as questões mais ligadas ao aspecto emocional apresentam uma significância maior em comparação com as de caráter físico.

## 5 Estimação do Traço Latente

Agora, estimamos os traços latentes subjacentes aos padrões de resposta.

Primeiramente, apresentamos a distribuição dos escores no SRQ-20, ou seja, a quantidade de indivíduos que responderam “sim” a nenhuma das questões, a uma questão, a duas questões, e assim sucessivamente, até aqueles que responderam “sim” a todas as questões.

```
#escores
score<- rowSums(df1[,c(2:21)])
table(score)
```

```
## score
## 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
## 18 25 21 40 34 45 53 43 52 48 53 45 65 55 62 40 37 24 29 23 11
```

Podemos observar que a maioria dos estudantes responde de maneira equilibrada às questões com “sim” e “não”, uma vez que a quantidade de pessoas está mais concentrada entre 6 e 14 respostas afirmativas. Isso indica que a maioria dos respondentes apresenta um padrão de respostas misto, sem uma predominância de respostas afirmativas ou negativas, sugerindo que os sintomas avaliados pelo SRQ-20 não são tão frequentes ou graves para a maioria dos estudantes. Entretanto, é importante levar em consideração que 11 estudantes responderam “sim” para todas as 20 questões do questionário, indicando que esses indivíduos podem estar apresentando um quadro mais acentuado de sintomas relacionados à saúde mental, conforme avaliado pelo SRQ-20. Ainda, temos que 18 pessoas responderam “não” para todas as questões.

São apresentadas a seguir algumas estatísticas desses escores.

```
#Estatísticas dos escores
describe(score)
```

```
##      vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1      1 823 9.94 5.06    10    9.95 5.93   0 20   20 -0.05   -0.88 0.18
```

Analisando esses dados, vemos que o escore de rastreio de sofrimento mental usando TCT possui uma distribuição simétrica, conforme evidenciado pela média (9.94) e mediana (10) muito próximas, além da assimetria próxima de zero (-0.05). O desvio padrão (5.06) e a MAD (5.93) mostram que há uma dispersão significativa dos valores em torno da média. A curtose negativa (-0.88) sugere uma distribuição mais achatada e com caudas mais finas do que a normal. Com uma amostra grande (823) e um pequeno erro padrão (0.18), a média pode ser considerada como uma estimativa confiável.

Portanto, os dados têm uma distribuição simétrica e moderadamente dispersa.

Agora, estimamos os escores de rastreio de sofrimento mental usando TRI assim os traços latentes considerando o modelo TRI e uma escala de média 50 e desvio padrão 10.

```
#escores tri
fullscores <- fscores(model2L, method='MAP')
scoretri=fullscores*10+50
describe(scoretri)
```

```
##      vars   n mean   sd median trimmed  mad  min   max range  skew kurtosis   se
## X1      1 823 49.94 8.94  49.79   49.92 9.24 30.12 70.68 40.56 0.01   -0.41 0.31
```

As estatísticas sobre o score TRI indicam que a distribuição é bastante simétrica e moderadamente dispersa. A média (49.94) e a mediana (49.79) são muito próximas, o que sugere a simetria, corroborada pela assimetria quase nula (0.01). O desvio padrão (8.94) e a MAD (9.24) apontam para uma dispersão moderada dos dados. A curtose negativa (-0.41) indica uma distribuição levemente mais achatada que a normal. Com um tamanho de amostra grande (823) e um erro padrão pequeno (0.31), a média estimada é confiável e precisa.

Na seguinte figura mostramos o grau de correlação entre os escores e os escores TRI, além do Teste de Correlação de Pearson.

```
#Correlação escores e escores tri
cor(score, fullscores)
```

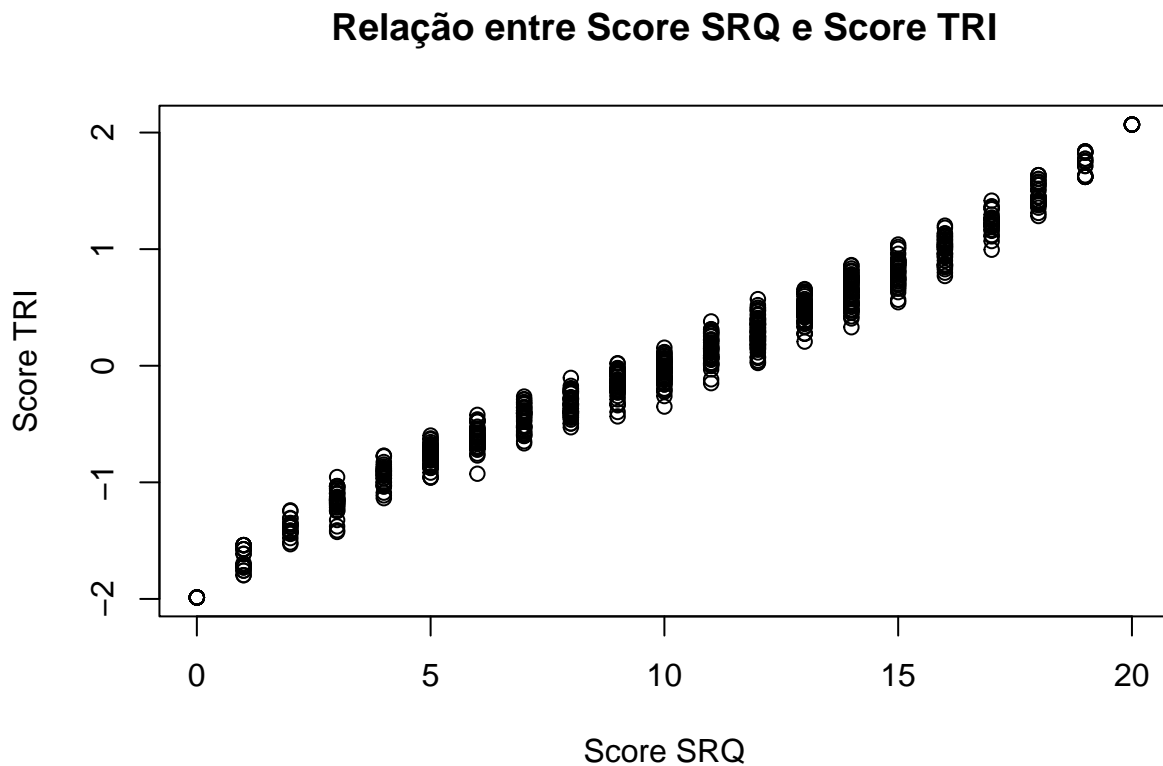
```
##              F1
## [1,] 0.989168
```

```
#Teste de Correlação de Pearson
```

```
cor.test(score, fullscores,  
         method="pearson",  
         alternative="two.sided",  
         conf.level = 0.95)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: score and fullscores  
## t = 193.09, df = 821, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9875887 0.9905472  
## sample estimates:  
## cor  
## 0.989168
```

```
plot(score, fullscores,  
     xlab = "Score SRQ",  
     ylab = "Score TRI",  
     main = "Relação entre Score SRQ e Score TRI")
```



Em resumo, a correlação entre o score SRQ-20 e o score obtido usando TRI é 0.989168, o que é evidenciado pelo gráfico entre elas, que revela uma forte relação linear positiva, indicando que as duas variáveis são altamente correlacionadas. Com o teste de Pearson, obtemos um p-valor próximo de zero ( $p\text{-value} < 2.2 \times 10^{-16}$ )

e menor que o nível de significância de 5%, o que indica que a correlação é estatisticamente significativa, ou seja, a correlação é de fato diferente de zero. Isso sugere que existe uma relação linear significativa entre as variáveis analisadas.

De acordo com os resultados e o gráfico, para um determinado escore obtido ao somar as respostas “sim” do teste, podemos obter diferentes valores nas estimativas do traço latente dos avaliados.

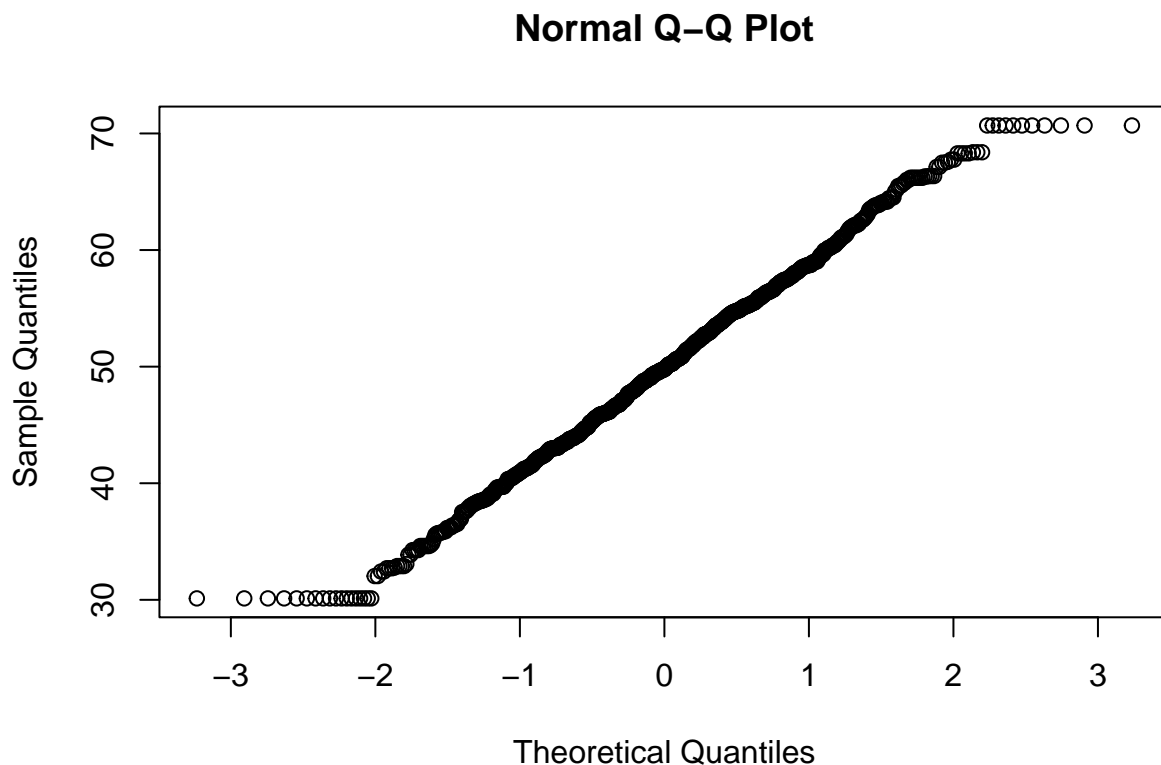
## 5.1 Avaliação da Normalidade dos escores TRI

A seguir, apresentamos algumas descrições adicionais usando uma escala transformada.

```
shapiro.test(escoretri)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  escoretri  
## W = 0.99349, p-value = 0.001202
```

```
qqnorm(escoretri)
```

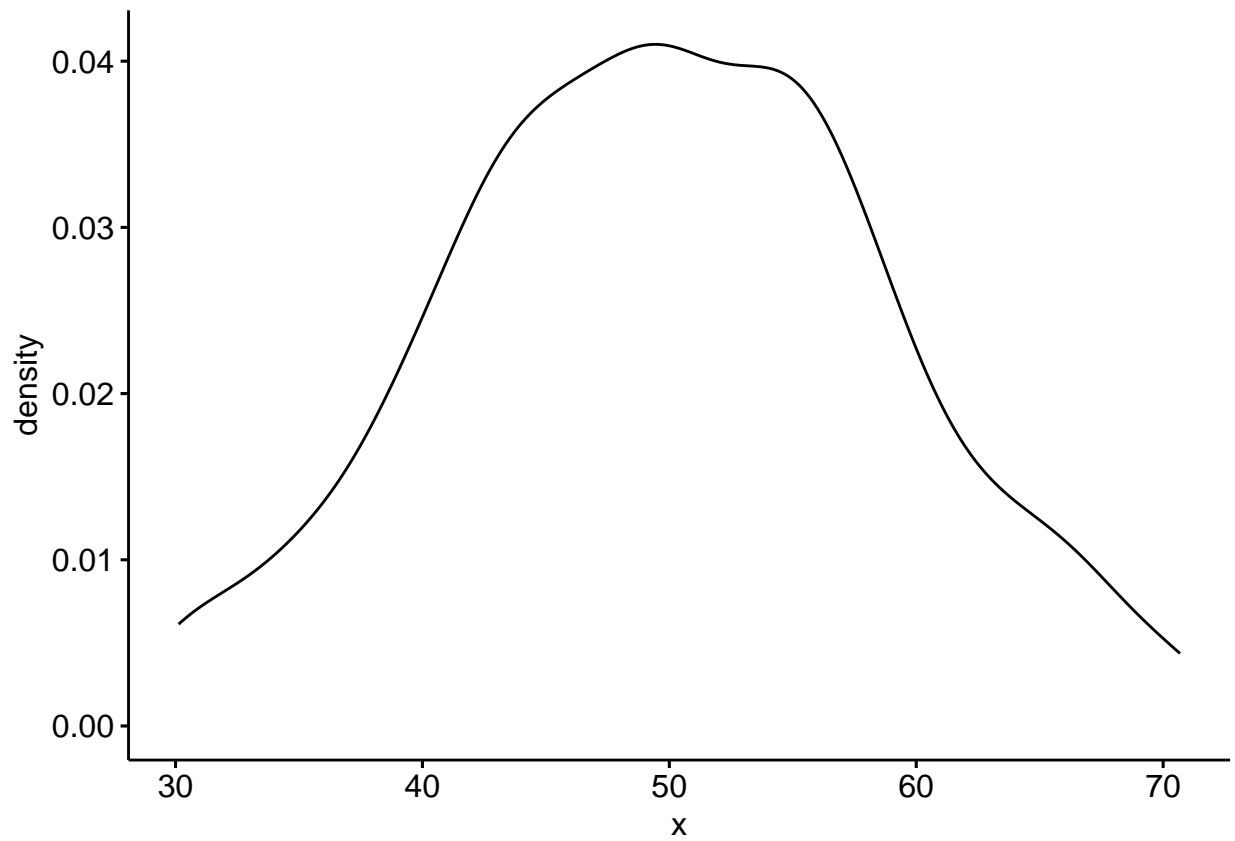


```
head(escoretri)
```

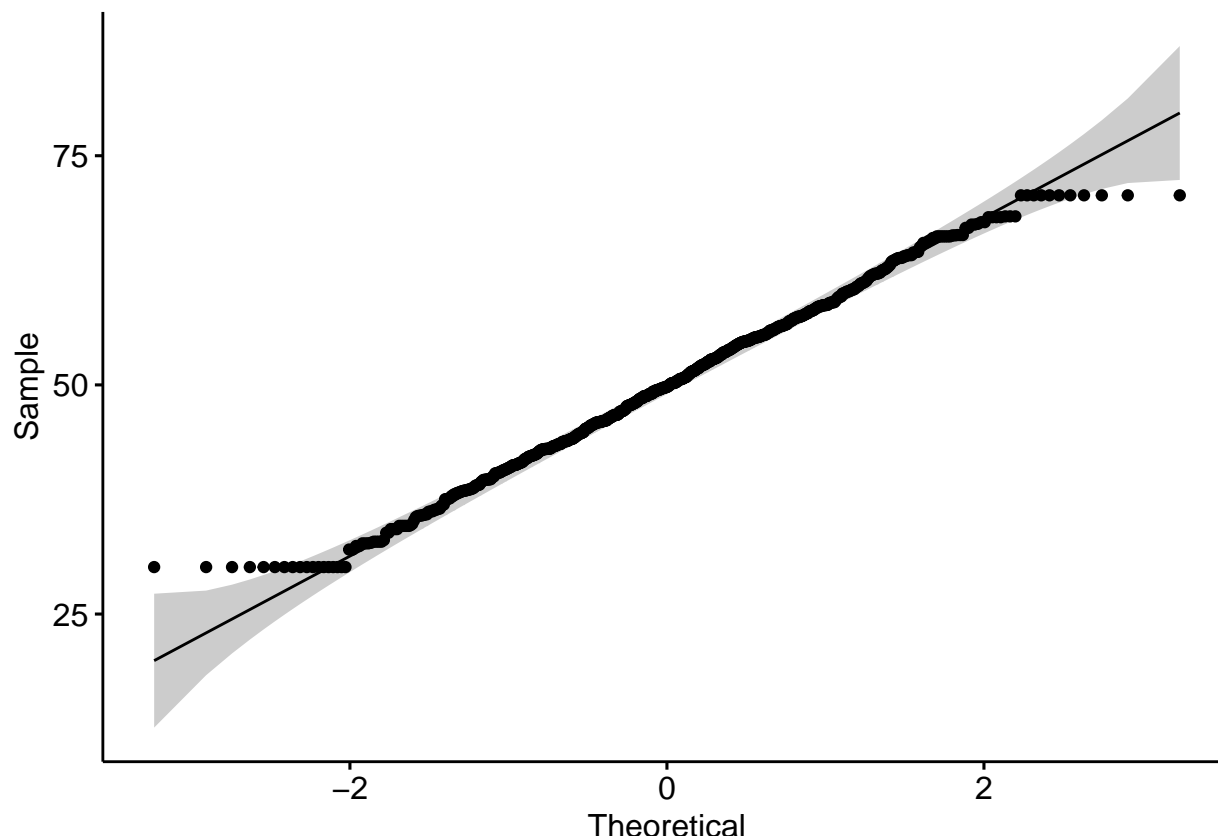
```
##          F1
```

```
## [1,] 48.18535
## [2,] 47.83775
## [3,] 46.75876
## [4,] 50.78378
## [5,] 42.50081
## [6,] 59.94951
```

```
library(ggpubr)
ggdensity(escoretri[,1])
```



```
ggqqplot(escoretri[,1])
```



O resultado do teste de normalidade de Shapiro-Wilk indica que a distribuição dos dados do escore TRI difere significativamente de uma distribuição normal, uma vez que o p-valor obtido é próximo de zero e menor que o nível de significância de 5%. Isso sugere que os dados não são normalmente distribuídos, o que pode ser notado pelos gráficos acima.

O gráfico Q-Q Plot apresenta um desvio significativo nas extremidades, afastando-se da linha diagonal. O gráfico de densidade mostra assimetrias em sua curva e não segue a forma da distribuição normal. Por fim, o gráfico de envelope exibe pontos fora de seus limites, indicando que os dados não seguem uma distribuição normal.

Finalmente, os resultados dos escores TRI são salvos na base de dados a seguir.

```
#Adiciona a nova coluna scoreTRI ao dataframe
escoretri<-as.data.frame(escoretri)
df1$scoreTRI <- escoretri$F1

#Salva o dataframe no arquivo Excel
write.xlsx(df1, "DadosNotebookTRI.xlsx")

head(df1)
```

```
## # A tibble: 6 x 22
##   ID  I01  I02  I03  I04  I05  I06  I07  I08  I09  I10  I11  I12
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1     1     0     1     1     0     1     0     0     1     0     1     0
## 2     2     0     0     0     1     0     1     0     1     1     0     1     1
## 3     3     0     0     1     0     1     1     1     0     0     0     1     1
```

```
## 4      4      1      0      1      1      1      1      0      1      0      0      0      1
## 5      5      0      0      1      0      0      0      0      0      1      0      0      1
## 6      6      1      1      1      1      1      1      1      1      1      1      1      1
## # i 9 more variables: I13 <dbl>, I14 <dbl>, I15 <dbl>, I16 <dbl>, I17 <dbl>,
## #      I18 <dbl>, I19 <dbl>, I20 <dbl>, scoreTRI <dbl>
```

## 6 Considerações Finais

A análise psicométrica da escala SRQ-20 realizada neste estudo evidenciou que o teste é confiável e unidimensional, permitindo a aplicação tanto da Teoria Clássica dos Testes (TCT), como da Teoria da Resposta ao Item (TRI) para estimar os escores de transtornos mentais comuns (TMC) e medir o sofrimento mental dos alunos de graduação da USP São Carlos. O modelo de dois parâmetros da TRI foi considerado o mais adequado para a obtenção dos resultados. A partir desse modelo, foi possível a compreensão da gravidade dos sintomas avaliados e da capacidade discriminativa de cada item. De maneira geral, verificou-se que as questões mais relacionadas ao aspecto emocional são mais significativas para indicar algum sofrimento mental em comparação com as questões de caráter físico. Esses resultados mostram que o SRQ-20 é uma ferramenta importante e capaz de auxiliar na avaliação da saúde mental dos estudantes, com base em suas respostas no questionário.

## 7 Referências

- Sartes, L. M. A. & Souza-Formigoni, M. L. O. (2013). Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item. Disponível em <https://doi.org/10.1590/S0102-79722013000200004>
- Bazán Guzmán, J. L. (2018). Psicometria e avaliação por testes: um marco metodológico. In Avaliação da educação : referências para uma primeira conversa. São Carlos, SP: EdUFSCar. Disponível em [https://edisciplinas.usp.br/pluginfile.php/7888277/mod\\_resource/content/2/Bazan2018.pdf](https://edisciplinas.usp.br/pluginfile.php/7888277/mod_resource/content/2/Bazan2018.pdf)
- Pasquali, Luiz, & Primi, Ricardo. (2003). Fundamentos da teoria da resposta ao item: TRI. Avaliação Psicológica, 2(2), 99-110. Disponível em [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712003000200002&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002&lng=pt&tlng=pt)
- Gadermann, A., Guhn, M. and Zumbo, B. (2012) Estimating Ordinal Reliability for Likert-Type and Ordinal Item Response Data: A Conceptual, Empirical, and Practical Guide. Practical Assessment, Research & Evaluation, 17, 1-13. Disponível em <https://files.eric.ed.gov/fulltext/EJ977577.pdf>
- Zumbo, Bruno D.; Gadermann, Anne M.; and Zeisser, Cornelia (2007) “Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales,” Journal of Modern Applied Statistical Methods: Vol. 6 : Iss. 1 , Article 4. DOI: 10.22237/jmasm/1177992180 Disponível em: <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/4>
- Material preparado pelo Prof. Dr. César Gonçalves de Lima – FZEA/USP. Disponível em [https://edisciplinas.usp.br/pluginfile.php/8009978/mod\\_resource/content/2/GIA5005%20%2811%29%20-%20An%C3%A1lise%20de%20itens%20%28alfa%20de%20Cronbach%29.pdf](https://edisciplinas.usp.br/pluginfile.php/8009978/mod_resource/content/2/GIA5005%20%2811%29%20-%20An%C3%A1lise%20de%20itens%20%28alfa%20de%20Cronbach%29.pdf)
- Bazán Guzmán, J. L., Valdivieso Serrano, L. H., & Calderón García, A. (2010). Enfoque Bayesiano en modelos de teoría de respuesta al ítem. Disponível em <http://argos.pucp.edu.pe/~jlbazan/download/Reportef27.pdf>
- Silveira, L. B., Kroeff, C. da R., Teixeira, M. A. P., & Bandeira, D. R. (2022). Uso do Self-Reporting Questionnaire (SRQ-20) para identificação de grupo clínico e predição de risco de suicídio. Revista Psicologia E Saúde, 13(4), 49-61. Disponível em <https://doi.org/10.20435/pssa.v13i4.1219>
- Chalmers, R., P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, 48(6), 1-29. doi: 10.18637/jss.v048.i06



- Rizopoulos, D. (2006) ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25. URL doi: 10.18637/jss.v017.i05
- Rizopoulos, D. and Moustaki, I. (2008) Generalized latent variable models with nonlinear effects. *British Journal of Mathematical and Statistical Psychology*, 61, 415–438.
- Harman, Harry and Jones, Wayne (1966) Factor analysis by minimizing residuals (minres), *Psychometrika*, 31, 3, 351-368.
- Revelle, W. (in preparation) *An Introduction to Psychometric Theory with applications in R*. Springer. Disponível em <https://personality-project.org/r/book/>
- Pettersson, E. and Turkheimer, E. (2010). Item selection, evaluation, and simple structure in personality data. *Journal of Research in Personality*, 44(4), 407–420. Disponível em <https://doi.org/10.1016/j.jrp.2010.03.002>
- Bazán, J. (2022). Una metodología para el Análisis de Unidimensionalidad de una Prueba. Nota Técnica. Disponível em <https://jorgeluisbazan.weebly.com/uploads/1/2/5/6/125695412/metodosanalisisunidimensionalidadv1.pdf>
- Revelle, W. (2022) How To: Use the psych package for Factor Analysis and data reduction. Disponível em <https://cran.r-project.org/web/packages/psychTools/vignettes/factor.pdf>
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Mesa Press.