

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Evidência para o modelo do fator geral ou do mutualismo na inteligência: composição de um fator geral e comparação de modelos de inteligência em uma coorte de 2511 crianças e adolescentes avaliadas em São Paulo e Porto Alegre.”

Matheus de Castro Siniscarchio

Lucia Pereira Barroso

São Paulo, Junho de 2024

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Evidência para o modelo do fator geral ou do mutualismo na inteligência: composição de um fator geral e comparação de modelos de inteligência em uma coorte de 2511 crianças e adolescentes avaliadas em São Paulo e Porto Alegre.”

PESQUISADORA: Marina Ferme

ORIENTADOR: Marcelo Batistuzzo

INSTITUIÇÃO: PUC-SP

FINALIDADE DO PROJETO: publicação.

RESPONSÁVEIS PELA ANÁLISE: Matheus de Castro Siniscarchio

Lucia Pereira Barroso

REFERÊNCIA DESTE TRABALHO: SINISCARCHIO, M. C.; BARROSO, L. P. **Relatório de Análise Estatística sobre o Projeto: “Evidência para o modelo do fator geral ou do mutualismo na inteligência”.** São Paulo, IME-USP, 2024. (RAE – CEA – 24P10)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

BROWN, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. 2. ed. New York: Guilford Press.

BROWNE, M. W.; CUDECK, R. (1993). Alternative ways of assessing model fit. *Testing structural equation models* (pp. 136-162). In K. A. Bollen & J. S. Long (Eds.). Newbury Park, CA: Sage.

HU, L. T.; BENTLER, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**, 1-55.

Lavaan: An R Package for Structural Equation Modeling. UCLA. *Confirmatory Factor Analysis in R: A Seminar*. Disponível em:

<https://stats.oarc.ucla.edu/r/seminars/rcfa/>. Acesso em: 15 jun. 2024.

SALUM, Giovanni Abrahão et al. **High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results**. *International Journal of Methods in Psychiatric Research*, v. 24, n. 1, p. 58-73, 2015. DOI: 10.1002/mpr.1459.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

LibreOffice Writer 24.2.2.2

R for Windows 4.3.3 (2024-02-29 ucrt)

RStudio for Windows 2023.12.1

TÉCNICAS ESTATÍSTICAS UTILIZADAS

06:080 – Análise Fatorial

ÁREA DE APLICAÇÃO

Resumo

Este trabalho apresenta uma análise estatística sobre a estrutura da inteligência em uma amostra de 2.511 crianças e adolescentes brasileiros, residentes em São Paulo e Porto Alegre. O estudo foi conduzido no contexto do Estudo de Coorte de Alto Risco para o Desenvolvimento de Distúrbios Psiquiátricos Infantis, um estudo longitudinal. Utilizando um modelo de equações estruturais com o pacote *lavaan*, investigamos a adequação do modelo do fator geral de inteligência (fator g).

A análise incluiu diversos testes cognitivos e psicoeducacionais, como o Teste de Desempenho Escolar (TDE), o Wechsler Intelligence Scale for Children (WISC) e o teste Confias, entre outros. Esses testes avaliaram uma ampla gama de habilidades cognitivas, incluindo aritmética, linguagem, fluência no desenho, memória episódica e memória operacional.

Dentre os principais desafios enfrentados durante a análise, destacam-se a presença de variáveis ausentes, a necessidade de controle da variável idade e a natureza binária de muitas variáveis. Para mitigar esses desafios, diferentes abordagens foram exploradas, incluindo a inclusão da idade como covariável, a estratificação por faixas etárias e a imputação de dados ausentes.

Os resultados indicam uma forte correlação entre diversas habilidades cognitivas, apoiando a existência de um fator geral de inteligência. No entanto, as variáveis ausentes e a necessidade de um controle rigoroso da idade introduziram complexidades adicionais na modelagem.

Este estudo contribui para a compreensão da estrutura da inteligência em crianças e adolescentes brasileiros, oferecendo evidências empíricas para teorias cognitivas e sugerindo a necessidade de abordagens robustas na análise de dados complexos. As conclusões destacam a importância de um modelo multifacetado para capturar a complexidade das habilidades cognitivas e a influência de fatores como a idade no desempenho dos testes.

Sumário

1 Introdução.....	.8
2 Objetivos.....	.8
3 Descrição do estudo.....	.9
4 Descrição das variáveis.....	.9
4.1 Variáveis da seção de Aritmética do TDE.....	.9
4.2 Variáveis da seção de Linguagem.....	.10
4.2.1. Variáveis do teste Confias.....	.10
4.2.2 Variáveis do TDE.....	.12
4.2.3 Variáveis do teste WISC.....	.13
4.3 Variáveis do teste de Fluência no desenho [3 Variáveis].....	.14
4.4 Variáveis do teste de Memória episódica.....	.14
4.4.1 Planejamento – Recuperação imediata.....	.14
4.4.2 Planejamento – Recuperação tardia.....	.15
4.5 Variáveis do teste de Memória operacional.....	.15
4.6 Variáveis do teste de Motricidade (Luria).....	.16
4.7 Variáveis do teste de Praxia construtiva.....	.17
5. Análise descritiva.....	.19
5.1 Seção de Aritmética do TDE.....	.19
5.1.1 Solução de problemas orais.....	.19
5.2 Seção de Linguagem.....	.21
5.2.1 Teste Confias.....	.21
5.2.2 Teste TDE.....	.22
5.2.3 Teste WISC.....	.23
5.3 Teste de Fluência no desenho.....	.23

5.4 Teste de Memória operacional.....	.24
5.4.1 Dígitos (WISC).....	.24
5.4.2 Blocos de Corsi.....	.24
5.5 Teste de Motricidade (Luria).....	.25
5.6 Teste de Praxia construtiva.....	.25
5.6.1 Bender.....	.25
5.6.2 Figura de Rey.....	.25
5.6.3 Cubos (WISC).....	.26
6 Modelagem.....	.26
6.1 Controle da Idade.....	.27
6.2 Variáveis Ausentes.....	.27
6.4 Ajustando o modelo.....	.29
6.5 Modelo utilizando as médias.....	.30
6.5.1 Idade ligada diretamente ao fator G.....	.31
Interpretação dos Índices de Ajuste.....	.31
Relações Importantes no Modelo.....	.32
6.5.2 Idade ligada a todas as variáveis de média.....	.32
Interpretação dos Índices de Ajuste.....	.33
Variáveis Latentes.....	.33
Regressões com Idade.....	.34
Considerações sobre Critérios de Ajuste.....	.34
7. Conclusão.....	.36
APÊNDICE A.....	.37
.....	.38

1 Introdução

A inteligência humana é um construto psicológico multifacetado que tem sido alvo de inúmeras pesquisas e debates acadêmicos. Desde o início do século XX, com a postulação do fator geral (g) da inteligência por Charles Spearman, muitos modelos foram propostos para explicar como diferentes habilidades cognitivas se correlacionam e contribuem para o que chamamos de inteligência. Spearman introduziu o uso da análise fatorial, uma técnica estatística que permite examinar as relações entre variáveis observadas e identificar conjuntos de variáveis que são altamente correlacionadas, sugerindo a existência de um ou mais fatores subjacentes comuns. Este método tem sido fundamental na pesquisa da inteligência, permitindo a identificação de estruturas fatoriais que apoiam a noção de um fator de inteligência geral, bem como fatores específicos. No entanto, o debate sobre a existência e a utilidade de um fator geral persiste. O advento de novas metodologias e modelos estatísticos, incluindo o modelo de mutualismo baseado em redes, trouxe uma perspectiva diferente para compreender a inteligência. Nesse contexto, este trabalho visa explorar a estrutura da inteligência em uma amostra de crianças brasileiras, buscando contribuir com evidências empíricas que suportem a compreensão de modelos cognitivos subjacentes à inteligência e seu desenvolvimento.

2 Objetivos

O objetivo principal deste estudo é investigar a estrutura da inteligência em uma amostra compreensiva de crianças e adolescentes brasileiros, fornecendo evidências para a teoria do fator geral da inteligência e para o modelo de mutualismo. Especificamente, busca-se:

- Ajustar um modelo fatorial e de mutualismo aos dados coletados.
- Comparar o ajuste do modelo fatorial geral e de mutualismo aos dados coletados, determinando qual deles oferece uma explicação mais robusta para o fenômeno do manifolde positivo, que é a correlação positiva entre distintas habilidades cognitivas.

- Compor um fator geral de inteligência (g) para cada indivíduo na amostra, utilizando uma ampla gama de testes que avaliam diversas funções cognitivas.

3 Descrição do estudo

A amostra de crianças brasileiras foi obtida do Estudo de Coorte de Alto Risco para o Desenvolvimento de Distúrbios Psiquiátricos Infantis (HRC), que é um estudo longitudinal e multicêntrico destinado a examinar as trajetórias de desenvolvimento de distúrbios mentais em crianças e adolescentes. O estudo inclui 2511 participantes nascidos entre 1998 e 2004, residentes em Porto Alegre e São Paulo, Brasil. As avaliações do estudo foram realizadas em várias fases e os dados analisados nesse estudo se concentram na fase posterior à triagem, onde diversos testes são aplicados nas crianças.

4 Descrição das variáveis

4.1 Variáveis da seção de Aritmética do TDE

O Teste de Desempenho Escolar (TDE) é uma ferramenta de avaliação psicoeducacional utilizada para medir a proficiência de crianças em áreas-chave do desempenho escolar.

O segmento de Aritmética do TDE é dividido em duas categorias principais: 'Solução de problemas orais' e 'Solução de problemas escritos'. Dentro destas, as variáveis são organizadas por ano escolar, avaliando as habilidades específicas dos alunos em resolver problemas matemáticos. As variáveis são as seguintes:

Solução de problemas orais [3 variáveis]:

- f_tde2a: avalia o conceito de ordenação numérica
- f_tde2b: avalia a habilidade de realizar adição.
- f_tde2c: avalia a habilidade de realizar subtração.

Solução de problemas escritos [35 variáveis]:

Para os anos do 1º ao 7º, as variáveis variam de f_tde3a1 a f_tde3a4, f_tde3b1 a f_tde3b5, f_tde3c1 a f_tde3c6, f_tde3d1 a f_tde3d5, f_tde3e1 a f_tde3e5, f_tde3f1 a

f_tde3f5, e f_tde3g1 a f_tde3g5, respectivamente. Cada conjunto de variáveis mede a competência em resolver problemas matemáticos escritos adequados para o nível de conhecimento esperado em cada ano. Além disso, dada a relevância da aritmética para a inteligência geral, a avaliação deste domínio é considerada um indicador direto, situando-se justamente abaixo do fator G na hierarquia de habilidades cognitivas analisadas pelo TDE.

Codificação da resposta (para todas as seções do TDE):

0: Indica uma resposta incorreta ou inabilidade para responder à tarefa.

1: Indica uma resposta correta ou a realização bem-sucedida da tarefa.

88: Não se aplica

99: Não sabe/Não respondeu

Para as variáveis (f_tde2a, f_tde2a, f_tde2c) temos uma pergunta matemática e a resposta correta foi transformada para 1 e errado para 0. Exemplo: qual número é maior, 9 ou 7?. Para 9 transformamos em 1 e 7 em 0.

4.2 Variáveis da seção de Linguagem

As variáveis de linguagem dentro do estudo são um conjunto de medidas detalhadas projetadas para capturar e avaliar a complexidade da linguagem e habilidades cognitivas relacionadas em crianças. Elas englobam diferentes instrumentos de avaliação, cada um focado em componentes específicos do desenvolvimento linguístico, como fonologia, vocabulário, leitura e escrita.

Através do teste "Confias", são exploradas as habilidades fonológicas, cruciais para a aquisição e o desenvolvimento da leitura e da escrita. O TDE avalia as competências em escrita e leitura, abrangendo desde o reconhecimento de palavras até a compreensão de textos. O componente de Vocabulário do Weschler Intelligence Scale for Children (WISC), mede o conhecimento e a aplicação de vocábulos, refletindo o desenvolvimento da linguagem expressiva.

4.2.1. Variáveis do teste Confias

O teste Confias é uma avaliação psicoeducacional que investiga habilidades fonológicas em crianças. O teste é estruturado para medir a consciência fonológica, que

é a habilidade de reconhecer e manipular fonemas, as unidades sonoras que compõem as palavras.

O Confias abrange diversos aspectos da consciência fonológica, divididos em dois níveis principais: o nível da sílaba e o nível do fonema. Cada nível é composto por várias tarefas que avaliam diferentes habilidades fonológicas, como síntese, segmentação, identificação e transposição, tanto de sílabas quanto de fonemas. O teste é desenhado para ser aplicado em um único momento e oferece uma visão detalhada das capacidades fonológicas individuais.

Variáveis nível da sílaba:

- Síntese (f_confias1a1, ..., f_confias1d4): combinação de sílabas para formar palavras. [4 variáveis].
- Segmentação (f_confias2a1, ..., f_confias2d4): identificação e separação das sílabas dentro de uma palavra. [4 variáveis].
- Identificação da sílaba inicial (f_confias3a1, ..., f_confias3d4) e medial (f_confias6a1, ..., f_confias6d4): reconhecimento da posição das sílabas em palavras. [8 variáveis].
- Identificação de rima (f_confias4a1, ..., f_confias4d4). [4 variáveis].
- Exclusão da sílaba (f_confias8a1, ..., f_confias8h6): capacidade de reconhecer a palavra que permanece após a remoção de uma sílaba [8 variáveis].
- Transposição da sílaba (f_confias9a1, ..., f_confias9d4): habilidade para reorganizar sílabas e formar novas palavras. [4 variáveis].
- Produção de palavra com sílaba (f_confias5a1, ..., f_confias5d4) e rima (f_confias7a1, ..., f_confias7d4): capacidade de criar palavras a partir de sílabas ou rimas dadas. [8 variáveis]

Variáveis nível do fonema:

- Produção de palavras que começam com um determinado som (f_confias10a1, f_confias10d4). [4 variáveis].

- Reconhecimento dos fonemas no início e no final de palavras: identificação do fonema inicial (f_confias11a1, ..., f_confias11d4) e final (f_confias12a1, ..., f_confias12d4) [8 variáveis].
- Exclusão do fonema (f_confias13a1, ..., f_confias13f6): identificação da palavra que resulta da remoção de um fonema específico. [6 variáveis]
- Síntese (f_confias14a1, ..., f_confias14d4): união de fonemas para criar palavras. [4 variáveis].
- Segmentação (f_confias15a1, ..., f_confias15d4): separação dos fonemas individuais de uma palavra. [4 variáveis].
- Transposição do fonema (f_confias16a1, ..., f_confias16d4): habilidade para trocar fonemas de lugar e formar novas palavras. [4 variáveis]

Codificação da resposta:

0: Errado

1: Certo

77: Recusa-se a responder

99: Não sabe/Não respondeu

4.2.2 Variáveis do TDE

O TDE é estruturado em subtestes projetados para avaliar a capacidade das crianças de transcrever palavras e textos através de ditados (escrita), decodificar e compreender passagens de leitura (leitura), bem como resolver problemas matemáticos (aritmética).

No subteste de escrita, as crianças são solicitadas a ouvir e transcrever palavras e frases. Essa parte do teste fornece uma medida direta de suas habilidades ortográficas e de processamento auditivo-linguístico, sendo crucial para identificar a fluência da escrita, a precisão ortográfica e a capacidade de composição textual.

No subteste de leitura, o foco está na rapidez e precisão com que as crianças podem ler palavras e compreender textos.

TDE escrita [34 Variáveis]:

- Ditado de palavras e/ou frases (f_tde1a,..., f_tde1hh): variáveis que representam a capacidade do indivíduo de transcrever individualmente.

TDE leitura [72 variáveis]:

- Leitura de palavras e/ou frases (f_tde4a1 - f_tde4a70): variáveis associadas à capacidade de reconhecimento e decodificação de palavras e textos.

Codificação da resposta (para todas as seções do TDE):

0: Indica uma resposta incorreta ou inabilidade para responder à tarefa.

1: Indica uma resposta correta ou a realização bem-sucedida da tarefa.

88: Não se aplica

99: Não sabe/Não respondeu

4.2.3 Variáveis do teste WISC

O teste WISC avalia a extensão e profundidade do conhecimento de vocabulário dos participantes, pontuando-os com base na precisão que fornecem para uma lista predeterminada de palavras. Essas palavras são organizadas em um gradiente de dificuldade, o que significa que nem todas as crianças conseguem responder a todos os itens. O teste consiste em 30 variáveis, denominadas de p_voc1 a p_voc30. No protocolo do teste, todas as crianças iniciam no item 1 e há um critério de interrupção que é aplicado após o participante errar seis itens.

Codificação da resposta:

0: Errou

1: Acertou parcialmente

2: Acertou

77: Não se aplica

88: Não sabe/Não respondeu

4.3 Variáveis do teste de Fluência no desenho [3 Variáveis]

Este teste avalia a fluência no desenho, que é a capacidade de gerar ideias visuais de forma rápida e diversificada.

- Desenhos únicos (p_t5p1): é a principal variável do teste e refere-se ao número de desenhos distintos e não repetidos criados pelo participante.
- Perseverações (p_t5p2): representa a quantidade de vezes em que o participante repete o mesmo desenho ou tema. Perseverações indicam dificuldade em mudar de estratégia ou em inovar, refletindo a rigidez cognitiva.
- Violações (p_t5p3): contabiliza ocorrências em que o participante não segue as instruções do teste, como desenhar elementos proibidos ou sair dos limites estipulados.

Codificação da resposta:

Quantidade de desenhos: 0, 1, 2, ...

77: Não se aplica

88: Não respondeu

99: Não sabe

4.4 Variáveis do teste de Memória episódica

O teste de memória episódica é destinado a medir a capacidade de um indivíduo de codificar, armazenar e recuperar informações.

4.4.1 Planejamento – Recuperação imediata

O planejamento da recuperação imediata é uma medida de função executiva relacionada à capacidade de organização, ao conseguir identificar e representar os grandes componentes da figura principal a ser evocada.

Reconhecimento visual de formas geométricas básicas (p_sava_ri1,...,p_sava_ri5) [5 variáveis].

4.4.2 Planejamento – Recuperação tardia

O planejamento da recuperação tardia é uma medida de função executiva relacionada à capacidade de organização, ao conseguir identificar e representar os grandes componentes da figura principal a ser evocada

Reconhecimento visual de formas geométricas básicas (p_sava_r1,...,p_sava_r5) [5 variáveis].

4.5 Variáveis do teste de Memória operacional

O teste de Memória Operacional é projetado para avaliar a capacidade de reter e manipular informações temporariamente. Este teste é composto por dois subtestes principais que medem diferentes aspectos da memória operacional:

Dígitos (WISC) :

- **Ordem direta:** os participantes ouvem uma sequência de números e devem repeti-la na mesma ordem. As variáveis (p_dgd1, p_dgd2, ..., p_dgd8) representam a habilidade de memorização e reprodução direta da informação. [8 variáveis]
- **Ordem indireta:** os participantes ouvem uma sequência de números e devem repeti-la em ordem inversa. As variáveis (p_dgi1, p_dgi2, ..., p_dgi7) medem a habilidade de manipulação de informação na memória operacional. [7 variáveis]

Blocos de Corsi:

- **Ordem direta:** nesta tarefa, blocos são tocados em uma sequência específica e o participante deve reproduzir a sequência tocando os blocos na mesma ordem. As variáveis (p_bld1, p_bld2, ..., p_bld8) avaliam a memória visuoespacial. [8 variáveis]
- **Ordem indireta:** semelhante à tarefa anterior, mas o participante deve reproduzir a sequência em ordem inversa. As variáveis (ex.: p_bli1, p_bli2, ..., p_bli8) são indicativas da capacidade de manipulação visuoespacial. [8 variáveis]

O teste dos blocos de Corsi possui variáveis dummy, que indicam o sucesso nas tentativas individuais, com valores separados para acertos na primeira e na segunda tentativa, além de combinar esses resultados em variáveis compostas.

Codificação da resposta:

2: Acerto na primeira tentativa

1: Acerto na segunda tentativa

0: Erro em ambas as tentativas

88: Não se aplica

99: Não sabe/Não respondeu

4.6 Variáveis do teste de Motricidade (Luria)

O teste de Motricidade avalia a função motora e a coordenação neuromuscular. Este teste, inspirado nas avaliações neuropsicológicas de Luria, pode incluir diversas tarefas que medem a velocidade, fluência e precisão dos movimentos, bem como a habilidade de manter a simetria e a coordenação, e o número de acertos. Cada aspecto da motricidade é quantificado através das seguintes variáveis:

- **Tempo (p_ludp1, p_luo1, p_lupe1):** registram o tempo necessário para completar 4 ciclos de uma tarefa específica em segundos. [3 variáveis]
- **Acertos (p_ludp2, p_luo2, p_lupe2):** avaliar a precisão com que o indivíduo consegue repetir uma sequência de movimentos motores. [3 variáveis]
- **Fluência (p_ludp3, p_luo3, p_lupe3):** avalia a suavidade e a eficiência dos movimentos, indicando a capacidade de executar ações de maneira fluida e sem hesitação. [3 variáveis]
- **Precisão (p_ludp4, p_luo4 p_lupe4):** mede quão exatamente os movimentos atendem aos objetivos da tarefa. [3 variáveis]
- **Simetria (p_ludp5, p_luo5, p_lupe5):** avalia se os movimentos são simétricos, principalmente em tarefas que requerem ação bilateral ou movimentos que devem ser espelhados entre os lados do corpo. [3 variáveis]
- **Coordenação (p_ludp6, p_luo6, p_lupe6):** refere-se à capacidade de sincronizar movimentos de maneira eficiente. [3 variáveis]

Codificação das respostas das variáveis:

2: O indivíduo completa quatro ciclos sem erros após o examinador apresentar o modelo inicial. Isso indica uma execução bem-sucedida e autônoma da tarefa.

1: O indivíduo consegue realizar quatro ciclos corretamente, mas requer estímulo verbal ou um modelo externo. Isso sugere que o indivíduo pode realizar a tarefa, mas ainda depende de alguma forma de assistência ou lembrete.

0: O indivíduo comete erros mesmo com estímulo verbal ou modelo externo. Isso pode indicar dificuldades na execução da tarefa motora ou problemas de compreensão/atenção.

99: O indivíduo não sabe ou não respondeu. Esta categoria é usada quando não há dados disponíveis sobre a execução da tarefa.

77: O indivíduo recusa-se a realizar a tarefa, o que pode ser por várias razões, incluindo falta de compreensão, falta de vontade ou outros fatores externos.

Para as outras variáveis temos respostas binárias (sim e não), além das categorias de não saber e recusa.

4.7 Variáveis do teste de Praxia construtiva

O teste de Praxia construtiva avalia a habilidade de um indivíduo para construir ou desenhar objetos, figuras ou formas em resposta a uma demonstração visual ou por cópia. Este teste geralmente mede as habilidades visuoespaciais e motoras. O teste pode ser composto por várias tarefas que incluem a reprodução de figuras geométricas e a montagem de objetos em três dimensões:

Bender:

Avalia a habilidade de copiar figuras geométricas, cada variável representa uma figura específica dentro do Teste Gestáltico Visomotor de Bender, avaliando precisão, organização espacial e reprodução de detalhes. (p_bd1,..., p_bd8) [8 variáveis]

Codificação das respostas

0: 0 pontos.

1: 1 ponto.

2: 2 pontos.

3: 3 pontos (observados em uma linha).

99: Indica que o respondente não sabe ou não respondeu.

77: Indica que o respondente se recusou a responder.

Figura de Rey (Cópia):

- **Base Rectangle (p_sava_c1):** Esta variável é avaliada da seguinte maneira:

2: O retângulo base é desenhado corretamente sem ser representado como uma unidade fragmentada, indicando uma boa integridade visual e a capacidade de reproduzir formas geométricas com precisão.

0: O retângulo base está ausente ou fragmentado, sugerindo dificuldades na percepção visual ou na coordenação motora para reproduzir a figura.

88: O indivíduo não sabe, o que pode indicar incerteza ou falta de confiança na tarefa.

99: O indivíduo não respondeu, o que pode ocorrer por uma variedade de razões, incluindo possível hesitação ou recusa em participar.

- **Outras variáveis (p_sava_c2, p_sava_c3, p_sava_c4, p_sava_c5):** essas variáveis seguem um esquema similar, mas sem a categoria que indica uma execução sem erros:

Codificação das respostas

1: A figura é desenhada corretamente sem ser representada como uma unidade fragmentada.

0: A figura está ausente ou fragmentada.

88: O indivíduo não sabe.

99: O indivíduo não respondeu.

Cubos (WISC):

Essas variáveis se referem à tarefa de construção de cubos, um subteste da Escala Wechsler de Inteligência para Crianças (WISC). Avaliam a capacidade de compreender e reproduzir padrões tridimensionais, coordenação visuomotora e raciocínio espacial.

Cube Model 1 - Points (p_cb1): Registra a pontuação baseada na precisão da montagem do modelo de Cubo 1. "2" é atribuído se o modelo foi corretamente montado

na primeira tentativa, "1" se foi corretamente montado na segunda tentativa, "0" se houve erro em ambas as tentativas, com as outras codificações seguindo o esquema anterior.

Para os demais Modelos de cubos (2 a 12), a estrutura das variáveis é similar.

5. Análise descritiva

Devido ao grande número de variáveis, optou-se por uma abordagem de análise mais generalista, focada nas correlações e nas distribuições de frequências. Essas correlações são fundamentais para a análise fatorial, que é um dos principais objetivos deste projeto. Para calcular as correlações, utilizamos métodos policóricos e tetracóricos. A correlação policórica é empregada para variáveis ordinais e é calculada entre variáveis latentes subjacentes que se supõe serem contínuas. Já a correlação tetracórica é usada para variáveis dicotômicas. Para o cálculo das correlações as categorias "não respondeu" e "recusa-se a responder" foram consideradas como respostas incorretas. Quando o item era "não aplicável", ele foi excluído da análise.

Vale ressaltar que os dados ausentes foram codificados como NA.

5.1 Seção de Aritmética do TDE

5.1.1 Solução de problemas orais

A distribuição percentual das respostas para as variáveis f_tde2a , f_tde2b , e f_tde2c do subteste de aritmética oral é a seguinte (Figura A.1):

- Para a variável f_tde2a , relacionada à comparação de grandezas numéricas, observa-se uma prevalência de respostas corretas de 73% (código 1, representado em verde).
- As variáveis f_tde2b e f_tde2c , que envolvem problemas de soma e subtração, mostram uma ligeira redução na frequência de respostas corretas em comparação com f_tde2a , tendo f_tde2b com aproximadamente 67% e f_tde2c com aproximadamente 65%.

- Para f_tde2a, cerca de 11% das respostas estão categorizadas como 'não sabe/não respondeu' (código 99, em azul) e há uma proporção idêntica de dados ausentes (NA).
- A análise de dados ausentes revela que estes são observados apenas para f_tde2a.

A Figura A.2 exibe a matriz de correlações entre as três variáveis do subteste de solução de problemas orais do TDE, evidenciando fortes inter-relações, com todas as correlações excedendo 0,9. Estes valores elevados de correlação implicam que o desempenho dos alunos em uma tarefa está fortemente ligado ao seu desempenho nas outras tarefas relacionadas dentro deste subteste.

5.1.2 Solução de problemas escritos

Na análise das variáveis da solução dos problemas escritos, as quais são categorizadas por ano escolar e, conseqüentemente por idade, decidiu-se conduzir a análise percentual por categoria e as análises de correlações separadas para cada ano do ensino. A Figura A.3 apresenta a distribuição de frequências de cada variável deste segmento do teste. É particularmente importante observar a proporção de respostas classificadas como 'não se aplica' (99), já que essas precisam ser excluídas para permitir uma comparação percentual adequada entre as categorias.

As Figuras A.4 a A.10 mostram a distribuição de frequências das variáveis avaliadas para cada ano escolar. Elas revelam uma tendência interessante: à medida que aumenta o ano escolar, observa-se uma diminuição na frequência de alunos que acertam as questões e um aumento daqueles que erram. Isso ocorre principalmente porque os alunos mais velhos respondem a todas as perguntas matemáticas escritas, enquanto os mais novos respondem apenas às questões apropriadas para suas idades. Esse padrão é evidenciado nos histogramas de idades dos alunos que responderam às perguntas designadas para o primeiro ano (Figura A.11) e para o sétimo ano (Figura A.12). Portanto, para uma análise mais detalhada deste subteste, será essencial controlar a variável idade.

Observa-se que, nas análises de correlação, à medida que o nível escolar aumenta, as correlações também tendem a crescer, como demonstrado nas Figuras A.13 até A.19. Este fenômeno ocorre porque, nos últimos anos escolares, as respostas

são mais consistentemente direcionadas para faixas etárias específicas. Portanto, ao fazer a análise por idade, as correlações naturalmente parecem mais fortes. Essa consistência acentuada sugere que a idade é um fator significativo que influencia as correlações observadas nos dados.

5.2 Seção de Linguagem

5.2.1 Teste Confias

Nível da sílaba

Observa-se, conforme ilustrado na Figura A.20, uma expressiva proporção de respostas corretas nas variáveis avaliadas, indicando um desempenho geralmente positivo para a maioria das crianças. Notavelmente, os percentuais atribuídos às categorias '77: recusa-se a responder' e '99: não sabe/não respondeu' foram inferiores a 1%.

Além disso, todas as correlações registradas foram positivas, como se pode verificar na Figura A.21. Uma inspeção visual das correlações revela que a grande maioria delas é consideravelmente alta (maior que 0,5), destacando-se um conjunto de correlações particularmente fortes entre as variáveis `f_confias1a1` até `f_confias2d4`. Esta observação enfatiza a possível interdependência das habilidades avaliadas nessas tarefas, como síntese e segmentação de sílabas.

Nível do fonema

No nível do fonema, conforme observado na Figura A.22, verifica-se uma maior proporção de erros e de respostas '99: não sabe/não respondeu', em comparação com o nível da sílaba, indicando possíveis dificuldades adicionais nas tarefas desse nível. Já a proporção de respostas '77: recusa-se a responder' mantém o comportamento do nível da sílaba.

As correlações para todas as variáveis desse nível são positivas, como se pode ver na Figura A.23. Destacam-se blocos de correlações muito altas (maiores que 0,9) nas variáveis de `f_confias1a1` a `f_confias2d4`, reforçando a consistência das habilidades fonêmicas avaliadas.

5.2.2 Teste TDE

TDE escrita

A análise da Figura A.24 revela que a categoria 1 (acertos) é a mais frequente na maioria das colunas, indicando uma prevalência significativa dessa categoria nos dados. A categoria 0 (vermelho) apresenta uma variação considerável, dominando algumas variáveis, como *f_tde1ee* e *f_tde1ff*, onde sua frequência ultrapassa 50%. A categoria 99 (não respondeu) está presente em todas as colunas, mas sempre em proporções menores comparadas às categorias 0 e 1. A categoria NA (cinza) aparece consistentemente em todas as colunas com uma porcentagem de 11%, indicando uma quantidade uniforme de dados faltantes.

Este é o primeiro teste em que observamos variáveis apresentando correlações tanto baixas quanto negativas. No entanto, a maioria das variáveis exibe correlações positivas, com alguns conjuntos específicos destacando-se por suas altas correlações, conforme ilustrado na Figura A.25. Essas variáveis, identificadas de *f_tde1a* a *f_tde1hh*, representam a capacidade individual de transcrição de palavras e/ou frases.

TDE leitura

Existe uma predominância de respostas corretas, conforme indicado pela Figura A.26. Isso sugere que a maioria das crianças apresenta uma boa capacidade de leitura dentro da amostra examinada.

Além disso, foi observada uma correlação muito alta entre as variáveis, como ilustrado na Figura A.27. Isso pode indicar uma forte consistência interna nas tarefas de leitura, em que as habilidades de reconhecimento de palavras e a compreensão de frases e textos estão fortemente interligadas, reforçando a ideia de que as habilidades de leitura podem se desenvolver de maneira complementar.

A Figura A.28 apresenta o histograma da distribuição do tempo de ditado, com os tempos agrupados em intervalos de 5 minutos, variando de 0 a 60 minutos. Observa-se que a frequência mais alta ocorre no primeiro intervalo (0-5 minutos), indicando que a maioria das sessões de ditado dura menos de 5 minutos. A frequência diminui progressivamente até o intervalo de 20-30 minutos e, então, oscila levemente, com uma nova queda notável no último intervalo 55 minutos. Isso sugere que sessões mais longas são menos comuns.

5.2.3 Teste WISC

Ao analisar o teste, cujas palavras são organizadas por graus crescentes de dificuldade, percebe-se, conforme ilustrado na Figura A.29, uma tendência das crianças a não progredirem para os estágios mais avançados do teste. Isso é evidenciado pelo aumento da categoria “77 – não se aplica” à medida que avançamos para variáveis com perguntas de maior dificuldade.

Na tentativa de calcular a correlação, a exclusão dos dados marcados como “não se aplica” resultaria em uma redução significativa para apenas 193 observações. Assim, a decisão foi por calcular as correlações apenas até a variável p_voc18, que precede o maior salto na porcentagem de “não se aplica”. Ao manter as observações das variáveis até p_voc18, o número de observações válidas sobe para 2124.

A inspeção do gráfico de correlações (Figura A.30) revela que, embora a maioria das correlações seja positiva, elas tendem a ser baixas. Isso pode ser atribuído ao fato de que as observações constituem sequências.

Uma consideração que surge é a possibilidade de somar as respostas das variáveis para este teste, devido à sua natureza sequencial.

5.3 Teste de Fluência no desenho

O histograma da variável p_t5p1 (Figura A.31) mostra uma distribuição assimétrica à direita, com a maioria dos valores concentrados entre 5 e 20, e um pico de frequência entre 10 e 15. A frequência diminui progressivamente à medida que os valores aumentam, com poucas observações além de 30, indicando possíveis valores atípicos.

A Figura A.32 apresenta o histograma da variável p_t5p2. Este histograma mostra uma distribuição altamente concentrada à esquerda, com a grande maioria dos participantes apresentando poucas ou nenhuma perseveração.

A Figura A.33 apresenta o histograma da variável p_t5p3. Similar à Figura A.32, este histograma também é altamente concentrado à esquerda, indicando que a maioria dos participantes seguiu as instruções e teve poucas ou nenhuma violação.

A Figura A.34 apresenta o gráfico de barras da porcentagem de dados ausentes (NA) para cada variável. Parece que todas as três colunas têm uma pequena porcentagem de dados ausentes.

5.4 Teste de Memória operacional

5.4.1 Dígitos (WISC)

A exemplo do teste WISC, este teste também é organizado por grau de dificuldade, o que novamente nos apresenta um desafio ao calcular as correlações. Ao excluir as observações rotuladas como "não se aplica", a quantidade de observações válidas diminui consideravelmente, como evidenciado na Figura A.35. Assim, optamos por remover as variáveis com um alto índice de respostas "não se aplica" antes de calcular as correlações.

É importante notar (Figura A.36) que as variáveis demonstram uma correlação satisfatória quando analisadas em relação à próxima variável ou quando comparadas com suas contrapartes diretas e indiretas, como por exemplo, `p_dgd1` e `p_dgi1`.

Assim como no teste de vocabulário, as correlações são todas positivas, embora não sejam em sua maioria altas.

Surge também aqui a possibilidade de somar as respostas das variáveis para este teste, devido à sua natureza sequencial.

5.4.2 Blocos de Corsi

Os blocos de Corsi seguem uma estrutura semelhante ao teste WISC e dígitos (WISC), organizados por grau de dificuldade. Mais uma vez, ao calcular as correlações, enfrentamos o desafio de lidar com as observações rotuladas como 'não se aplica', o que reduz significativamente o número de observações válidas, conforme demonstrado na Figura A.37. Decidimos, portanto, remover as variáveis com um alto índice de respostas 'não se aplica' antes de prosseguir com o cálculo das correlações.

Ao observar os resultados (Figura A.38), podemos notar um padrão semelhante de correlações, porém, com valores ligeiramente mais elevados em comparação com o teste de dígitos (WISC).

Assim como no teste WISC e dígitos (WISC), surge a possibilidade de somar as respostas das variáveis para este teste, devido à sua natureza sequencial.

5.5 Teste de Motricidade (Iuria)

Pela observação da Figura A.39, podemos notar que todas as variáveis apresentam uma alta taxa de acertos. Além disso, nas variáveis relacionadas ao aspecto 'acertos', observamos uma categoria adicional, denominada '2, quatro ciclos sem erros após o examinador fazer o modo inicial'. Notavelmente, essa categoria também registrou uma frequência significativa.

Uma análise visual da Figura A.40 revela que as correlações são predominantemente positivas, com a maioria delas situando-se em torno de 0,5. Além disso, algumas correlações apresentam valores mais altos, aproximadamente 0,75.

5.6 Teste de Praxia construtiva

5.6.1 Bender

Pela análise da Figura A.41, observa-se que apenas as variáveis p_bd1 e p_bda apresentam uma predominância de respostas na categoria de zero pontos. Nas demais variáveis, nota-se uma predominância de respostas com um ponto, seguido por duas respostas com dois pontos. Apenas na variável p_bd6 observa-se respostas com pontuação igual a 3.

As correlações, conforme ilustrado na Figura A.42, são todas positivas, com a maioria situando-se na faixa de 0,3 a 0,5.

5.6.2 Figura de Rey

É importante ressaltar que, neste teste, apenas a variável p_sava_c1 apresenta a categoria 2 como resposta. Observa-se que nesta variável, assim como nas variáveis p_sava_c2 e p_sava_c4, há uma predominância da categoria 0. Por outro lado, nas variáveis p_sava_c3 e p_sava_c5, nota-se uma maior frequência da categoria 1 (Figura A.43).

As correlações entre as variáveis são positivas, porém nenhuma se destaca como sendo muito alta (Figura A.44).

5.6.3 Cubos (WISC)

A Figura A.45 apresenta a distribuição de frequência das variáveis dos Cubos (WISC).

As variáveis p_cb1 a p_cb3 apresentam alta frequência de pontuações completas (2), indicando que a maioria dos respondentes consegue montar os modelos corretamente na primeira tentativa. Esses modelos podem ser considerados relativamente mais fáceis ou mais aptos à faixa etária dos participantes.

As variáveis p_cb4 a p_cb12 apresentam aumento na diversidade de pontuações (4 a 7) e elevação significativa da frequência da categoria 77 ("não se aplica"), sugerindo maior dificuldade ou inaplicabilidade para alguns respondentes.

Para calcular as correlações (Figura A.46), utilizou-se a mesma abordagem empregada anteriormente, com o objetivo de preservar o maior número possível de observações. Optou-se por excluir as variáveis com a maior incidência de "não se Aplica", especificamente p_cb10, p_cb11 e p_cb12. As correlações observadas entre as variáveis restantes são predominantemente positivas, destacando-se as elevadas correlações entre as primeiras variáveis (p_cb1 a p_cb3). Esta tendência sugere uma forte relação entre os desempenhos nessas tarefas iniciais.

6 Modelagem

Com o ordenograma fornecido pelos pesquisadores (Figura A.47), o objetivo é ajustar um modelo fatorial confirmatório para investigar se a estrutura de inteligência proposta pode ser decomposta em um fator geral de inteligência (fator G).

A estrutura proposta possui um alto nível de complexidade, com 4 níveis fatoriais, 41 fatores latentes e 334 variáveis. Devido à complexidade do modelo, optou-se por ajustá-lo progressivamente, trabalhando com um constructo de inteligência por vez. Esse método permitiu verificar a convergência do modelo a cada etapa e adicionar novos constructos gradualmente. Caso contrário, o modelo provavelmente não convergiria.

Antes mesmo da modelagem, alguns problemas principais surgiram como pontos de discussão.

6.1 Controle da Idade

Todos os testes foram realizados em crianças de diferentes faixas etárias, sendo evidente que crianças mais velhas possuem uma vantagem significativa em relação às crianças mais novas na resposta aos testes. Portanto, é necessário implementar um controle sobre a variável idade. Foram abordadas três alternativas para o controle da idade.

Incluindo Idade como covariável

A idade das crianças foi incluída como uma covariável no modelo de Análise Fatorial Confirmatória (CFA). Essa abordagem permite ajustar os efeitos dos fatores latentes sobre as variáveis observadas, considerando a idade. Ao incluir a idade como covariável, é possível controlar seu impacto nas relações entre os fatores latentes e as variáveis observadas.

Estratificação por Faixas etárias

A amostra foi dividida em diferentes faixas etárias, e análises separadas foram realizadas para cada grupo. Essa abordagem possibilita verificar se a estrutura fatorial se mantém consistente entre as diferentes faixas etárias.

Interação entre Idade e o Fator Geral de Inteligência

Foi incluída a interação entre a idade e o fator geral de inteligência (fator G). Essa técnica ajuda a identificar o impacto da idade sobre o fator geral, permitindo um ajuste mais preciso no modelo.

Essas estratégias foram exploradas para garantir que a idade não introduzisse vieses nos resultados da análise fatorial, proporcionando uma compreensão mais precisa das relações entre os fatores latentes e as variáveis observadas.

6.2 Variáveis Ausentes

Em um modelo fatorial, a presença de variáveis ausentes impede o ajuste adequado do modelo. Portanto, temos duas alternativas possíveis: a imputação de dados ou a remoção das observações com dados ausentes. No entanto, na nossa amostra, nenhuma criança respondeu a todos os itens de todos os testes, tornando a remoção de dados ausentes inviáveis. Além disso, algumas observações foram

rotuladas como "não se aplica", indicando que a criança não tinha a idade adequada para aquele teste específico. Quando essas observações são tratadas como dados ausentes, a quantidade de dados faltantes aumenta significativamente e além disso estamos introduzindo um viés nos dados ausentes, que é a relação com a idade.

Essas observações rotuladas como "não se aplica" estão diretamente relacionadas à variável idade. Portanto, é necessário desenvolver uma solução que aborde ambos os problemas de maneira integrada.

Uma possível solução que já foi comentada e poderia resolver os dois problemas é a estratificação por faixas etárias. Dessa forma, teríamos menos chances de ter observações ausentes dentro das mesmas faixas. Entretanto, surgem outros problemas: quais seriam as faixas ideais? Essas faixas servem para todos os testes (alguns testes podem ser mais influenciados pela idade do que outros)?

Outra opção seria a imputação de dados. No entanto, nota-se que, pela forma como o teste foi conduzido, as variáveis ausentes têm uma forte relação com a idade. A maioria dos métodos de estimação pressupõe que as variáveis ausentes tenham uma distribuição aleatória, o que não é o nosso caso.

Portanto, ao considerar a imputação de dados, é importante selecionar métodos que possam lidar com dados não aleatórios.

Para um aprofundamento das técnicas e formas de lidar com dados ausentes, recomenda-se a consulta ao Capítulo 9 de Brown (2015).

6.3 Variáveis Binárias

Além dos problemas mencionados, a maioria das variáveis em nosso estudo são binárias. Isso resulta em muitas correlações iguais ou próximas de 1, como observado no teste de TDE leitura (Figura A.26). Correlações próximas de 1 representam um problema de multicolinearidade, que pode dificultar o cálculo da inversa de uma matriz (BROWN, 2015, p. 162), potencialmente levando a erros no ajuste do modelo.

Além disso, ao trabalhar com dados categóricos, devem ser utilizadas técnicas específicas para essa natureza de dados. Para um aprofundamento das técnicas e formas de lidar com dados categóricos, recomenda-se a consulta ao Capítulo 9 de Brown (2015).

6.4 Ajustando o modelo

Neste trabalho, realizamos diversas tentativas de ajuste e destacamos aqui as principais.

A primeira tentativa foi ajustar um modelo fatorial sem modificações, exatamente como proposto no diagrama fornecido pelos pesquisadores (Figura A.47). Utilizamos o método WLSMV para a estimação. O WLSMV é particularmente útil para variáveis dependentes categóricas ou ordinais, onde os métodos tradicionais de Mínimos Quadrados Gerais (GLS) ou Máxima Verossimilhança (ML) podem não ser apropriados. Quando ajustado para um único fator, a convergência foi muito boa. Entretanto, conforme adicionamos mais fatores, surgiram problemas de convergência e um grande recurso computacional foi necessário. Os testes foram realizados em um sistema Windows 11 com processador I7, e a alocação de memória variou entre 2 GB e 10 GB. Além disso, a adição de mais fatores resultou em uma redução significativa na quantidade de observações, chegando a não ter observações suficientes quando as variáveis ausentes eram excluídas (método "listwise").

Cabe observar que o pacote **lavaan** possui métodos para lidar com variáveis ausentes, sendo o método "pairwise" um dos mais populares. Este método calcula correlações ou covariâncias utilizando todos os casos disponíveis para cada par de variáveis, maximizando o uso dos dados disponíveis. No entanto, esse método não é indicado para nosso problema, pois pode resultar em matrizes de covariância ou correlação inconsistentes, já que diferentes pares de variáveis podem ter diferentes números de casos, introduzindo um viés significativo devido à idade, dado que os participantes mais velhos tendem a ter mais variáveis respondidas.

A segunda tentativa foi semelhante à primeira, mas dividindo a amostra por faixas etárias, com o objetivo de reduzir a quantidade de variáveis ausentes dentro de cada faixa etária. No entanto, essa abordagem enfrentou os mesmos problemas da primeira tentativa.

Essas duas primeiras tentativas foram focadas no fator de linguagem. Em nenhuma das duas foi possível ajustar todos os fatores sob o fator linguagem devido a problemas de convergência. Se essa estratégia já apresentava problemas com fatores

que tinham um maior número de respondentes, seria ainda mais difícil ajustar um modelo com essa estratégia quando se avançasse para fatores com rótulos de "não se aplica". Controlar a idade também se tornaria extremamente complexo, pois seria necessário relacionar a idade com todos os testes ou itens do modelo.

Com o objetivo de simplificar nosso modelo e resolver os problemas de controle de idade e variáveis ausentes, adotamos uma terceira abordagem: calcular as médias dos itens de cada teste.

6.5 Modelo utilizando as médias

A terceira tentativa foi pensada para resolver três principais problemas identificados no modelo:

1. **Controle das idades e rótulos de “não se aplica”**
2. **Variáveis ausentes**
3. **Alta complexidade do modelo**

É importante lembrar que, em alguns testes, os participantes mais velhos responderam a todos os itens, enquanto os mais novos responderam apenas aos itens apropriados para suas idades. Quando um participante deixa de responder por aquele item não ser adequado à sua faixa etária, a observação é rotulada como “não se aplica”, o que estamos substituindo por dado ausente.

Ao calcular a média dos itens, a criança que não respondeu aos itens mais difíceis não será penalizada. Dessa forma, estamos tentando resolver o problema de "não se aplica" e dados ausentes, ao mesmo tempo em que tentamos controlar a idade.

Ao calcular a média dos testes, reduzimos um nível de fator de todo o modelo e, além disso, diminuímos significativamente a quantidade de variáveis de 334 para 35 variáveis. Esta abordagem simplifica o modelo, tornando-o mais gerenciável e menos exigente em termos de recursos computacionais.

Foram testadas duas possibilidades do modelo:

1. **Idade ligada diretamente ao fator geral:**
 - Esta abordagem não controla efetivamente a idade, mas nos diz o quanto a idade influencia o fator geral (fator G).
2. **Idade ligada a todas as variáveis de média:**

- Esta abordagem tem como objetivo controlar a idade de forma mais precisa, garantindo que a média de cada teste seja considerada em relação à faixa etária apropriada.

6.5.1 Idade ligada diretamente ao fator G

Foi ajustado um modelo de equações estruturais utilizando o pacote lavaan em uma amostra de 2.511 observações. Após remover as observações com dados ausentes, restaram 2.044 observações. O Ajuste pode ser observado na (Figura A.48). O estimador utilizado foi a Máxima Verossimilhança (ML), com os seguintes resultados:

- **Número de Parâmetros:** 117
- **Número de Iterações:** 380
- **Estatística Qui-quadrado:** 5183.282 ($p < 0.001$)
- **Graus de Liberdade:** 582
- **CFI:** 0.871
- **TLI:** 0.861
- **GFI:** 0.859
- **RMSEA:** 0.062

Interpretação dos Índices de Ajuste

Os índices de ajuste fornecem uma avaliação da adequação do modelo aos dados:

1. CFI (Comparative Fit Index) e TLI (Tucker-Lewis Index):

- Ambos os índices estão próximos, mas ligeiramente abaixo do valor de corte de 0.90, indicando um ajuste moderado. O CFI e o TLI são usados para comparar o modelo ajustado com um modelo nulo, onde valores acima de 0.90 indicam um bom ajuste.
- Hu e Bentler (1999) sugerem que CFI e TLI valores próximos de 0.95 ou maiores indicam um bom ajuste, enquanto valores entre 0.90 e 0.95 podem ser considerados aceitáveis. No entanto, valores abaixo de 0.90 indicam que o modelo deve ser suspeito ou rejeitado.

2. GFI (Goodness-of-Fit Index):

- O valor de GFI é 0.859, que está ligeiramente abaixo do valor de corte de 0.90. Isso sugere que o modelo pode ser melhorado. O GFI mede a proporção da variância-covariância nos dados que é explicada pelo modelo.

3. RMSEA (Root Mean Square Error of Approximation):

- O valor de RMSEA é 0.062, indicando um ajuste razoável, mas com margem para melhorias. Valores de RMSEA abaixo de 0.05 indicam um bom ajuste, enquanto valores até 0.08 indicam um ajuste aceitável.

Relações Importantes no Modelo

Uma relação significativa encontrada no modelo foi entre a idade e o fator geral (fator_g):

- **fator_g ~ idade:**
 - A idade tem um efeito positivo significativo sobre o fator geral, com um coeficiente de 0.013 ($p < 0.001$). Isso indica que, conforme a idade aumenta, o fator geral também tende a aumentar de forma significativa.

Os resultados indicam que o modelo ajustado oferece um ajuste moderado aos dados, com índices CFI e TLI próximos, mas abaixo do ideal, e um RMSEA que sugere um ajuste razoável. A relação significativa entre idade e o fator geral (Figura A.49), com uma correlação de 0.69, destaca a importância de considerar a idade como uma variável influente no modelo. Isso também indica que a estratégia das médias não é eficaz para controlar a idade, provavelmente devido ao fato de apenas alguns modelos terem o controle de rotular os itens das faixas de idades adequadas.

6.5.2 Idade ligada a todas as variáveis de média

Foi ajustado um modelo de equações estruturais utilizando o pacote **lavaan** em uma amostra de 2044 observações (Figura A.51). Acertar o número da figura no pdf. O estimador utilizado foi a Máxima Verossimilhança (ML), com os seguintes resultados:

- **Número de Parâmetros:** 117
- **Número de Iterações:** 380
- **Estatística Qui-quadrado:** 4748.593 ($p < 0.001$)

- **CFI (Comparative Fit Index):** 0.882
- **TLI (Tucker-Lewis Index):** 0.865
- **GFI (Goodness of Fit Index):** 0.869
- **RMSEA:** 0.061

Interpretação dos Índices de Ajuste

Os índices de ajuste fornecem uma avaliação da adequação do modelo aos dados:

1. **CFI (Comparative Fit Index) e TLI (Tucker-Lewis Index):**

- Ambos os índices estão próximos, mas ligeiramente abaixo do valor de corte de 0.90, indicando um ajuste moderado. O CFI e o TLI são usados para comparar o modelo ajustado com um modelo nulo.
- Hu e Bentler (1999) sugerem que CFI e TLI valores próximos de 0.95 ou maiores indicam um bom ajuste, enquanto valores entre 0.90 e 0.95 podem ser considerados aceitáveis. No entanto, valores abaixo de 0.90 indicam que o modelo deve ser suspeito ou rejeitado.

2. **GFI (Goodness-of-Fit Index):**

- O valor de GFI é 0.869, que está ligeiramente abaixo do valor de corte de 0.90. Isso sugere que o modelo pode ser melhorado. O GFI mede a proporção da variância-covariância nos dados que é explicada pelo modelo.

3. **RMSEA (Root Mean Square Error of Approximation):**

- O valor de RMSEA é 0.061, indicando um ajuste razoável, mas com margem para melhorias. Hu e Bentler (1999) recomendam que valores de RMSEA próximos a 0.06 ou menores indicam um bom ajuste, enquanto Browne e Cudeck (1993) sugerem que valores abaixo de 0.08 indicam um ajuste adequado, e valores acima de 0.10 indicam um ajuste insatisfatório.

Variáveis Latentes

- **nivel_silaba e nivel_fonema:** Têm cargas fatoriais significativas para todas as variáveis observadas, indicando uma boa representação dos fatores latentes.
- **confias:** Fortemente associado a nivel_silaba e nivel_fonema.
- **tde_ling:** Bem representado por mean_tde_escrita e mean_tde_leitura.

- **linguagem:** Fortemente associada a confias, tde_ling e mean_vocabulario.
- **tde_arit, motricidade, praxia_construtiva, digitos, corsi, mem_ope e mem_epis:** Têm cargas fatoriais significativas para todas as suas variáveis observadas.

Regressões com Idade

- Todas as médias das variáveis observadas têm coeficientes de regressão significativos com a idade ($p < 0.001$), indicando que a idade tem um impacto significativo em todas essas variáveis.
- Observa-se pela Figura A.51 que não existe correlação entre o fator geral e a idade, com a correlação de Pearson entre as duas variáveis sendo zero. Isso indica que a estratégia utilizada foi **eficaz no controle da idade em relação ao fator geral**.

Os resultados indicam que o modelo ajustado oferece um ajuste moderado aos dados, com índices CFI e TLI próximos, mas abaixo do ideal, e um RMSEA que sugere um ajuste razoável. Estes resultados sugerem que, embora o modelo forneça uma base inicial sólida, há espaço para melhorias. Ajustes adicionais, como a reavaliação das variáveis incluídas ou a consideração de possíveis interações adicionais, podem ajudar a melhorar o ajuste geral do modelo.

Considerações sobre Critérios de Ajuste

Os critérios de ajuste de modelos são fundamentais para avaliar a adequação do modelo aos dados observados. Hu e Bentler (1999) propõem que:

1. **SRMR:** Valores próximos a 0.08 ou menores indicam um bom ajuste.
2. **RMSEA:** Valores próximos a 0.06 ou menores são indicativos de bom ajuste, enquanto valores entre 0.08 e 0.10 sugerem ajuste medíocre, e valores acima de 0.10 indicam um ajuste insatisfatório.
3. **CFI e TLI:** Valores próximos a 0.95 ou maiores são considerados indicadores de bom ajuste, enquanto valores entre 0.90 e 0.95 são aceitáveis, e valores abaixo de 0.90 indicam a necessidade de revisão do modelo.

No entanto, outros metodologistas, como Browne e Cudeck (1993), sugerem que valores de RMSEA abaixo de 0,08 indicam um ajuste adequado, valores abaixo de 0,05 sugerem um ajuste muito bom, e valores acima de 0,10 indicam que o modelo deve ser rejeitado. Portanto, a interpretação dos índices de ajuste deve ser feita com cautela, levando em consideração o contexto específico do modelo e a variabilidade dos critérios de ajuste.

Variâncias Negativas

Na análise das variâncias estimadas no modelo, observa-se que a maioria das variâncias apresenta valores positivos e significativos, o que é esperado e indica uma boa captura da variabilidade dos dados pelo modelo. No entanto, algumas variâncias apresentam valores negativos próximos de zero, especificamente:

.nivel_silaba: -0.000 (p = 0.001)

.digitos: -0.000 (p = 0.940)

Interpretação de Variâncias Negativas Próximas de Zero

Embora variâncias negativas sejam matematicamente implausíveis e geralmente indicativas de problemas no modelo, uma variância de -0.000 pode ser interpretada como zero, considerando a precisão numérica limitada do software utilizado para a estimação. Isso sugere que a variabilidade explicada por esses componentes é extremamente baixa ou nula.

Possíveis Causas e Implicações

Precisão Numérica:

A variância de -0.000 pode resultar de arredondamentos numéricos durante o processo de estimação. Neste contexto, é razoável interpretar essas variâncias como sendo zero, indicando que não há variabilidade significativa associada a esses fatores específicos.

Modelo Saturado:

Em modelos altamente parametrizados, onde a quantidade de variáveis e parâmetros é muito alta, pequenas discrepâncias numéricas podem surgir. Isso pode resultar em estimativas de variância muito próximas de zero.

Especificação do Modelo:

Mesmo que a variância seja tecnicamente negativa, seu valor extremamente pequeno sugere que a variabilidade associada ao fator é insignificante. Esse resultado pode indicar que a especificação do modelo está correta, mas o fator específico não contribuiu significativamente para a variabilidade observada nos dados.

Recomendações:

Interpretação como Zero:

Dado que a variância é -0.000 , pode-se interpretá-la como zero para fins práticos. Isso indica que esses fatores não adicionam variabilidade significativa ao modelo

7. Conclusão

A análise inferencial realizada revelou a complexidade e os desafios enfrentados ao ajustar o modelo fatorial confirmatório proposto para investigar a estrutura de inteligência. A inclusão da idade como covariável, a estratificação por faixas etárias e o uso de médias dos itens foram estratégias cruciais para controlar a idade e lidar com variáveis ausentes. No entanto, a presença de variâncias negativas próximas de zero e correlações elevadas entre variáveis binárias indicou problemas de multicolinearidade no ajuste. Apesar dessas dificuldades, os resultados mostraram um ajuste moderado do modelo, sugerindo que melhorias e ajustes adicionais são necessários para aumentar a precisão e a robustez das estimativas.

APÊNDICE A

Figuras

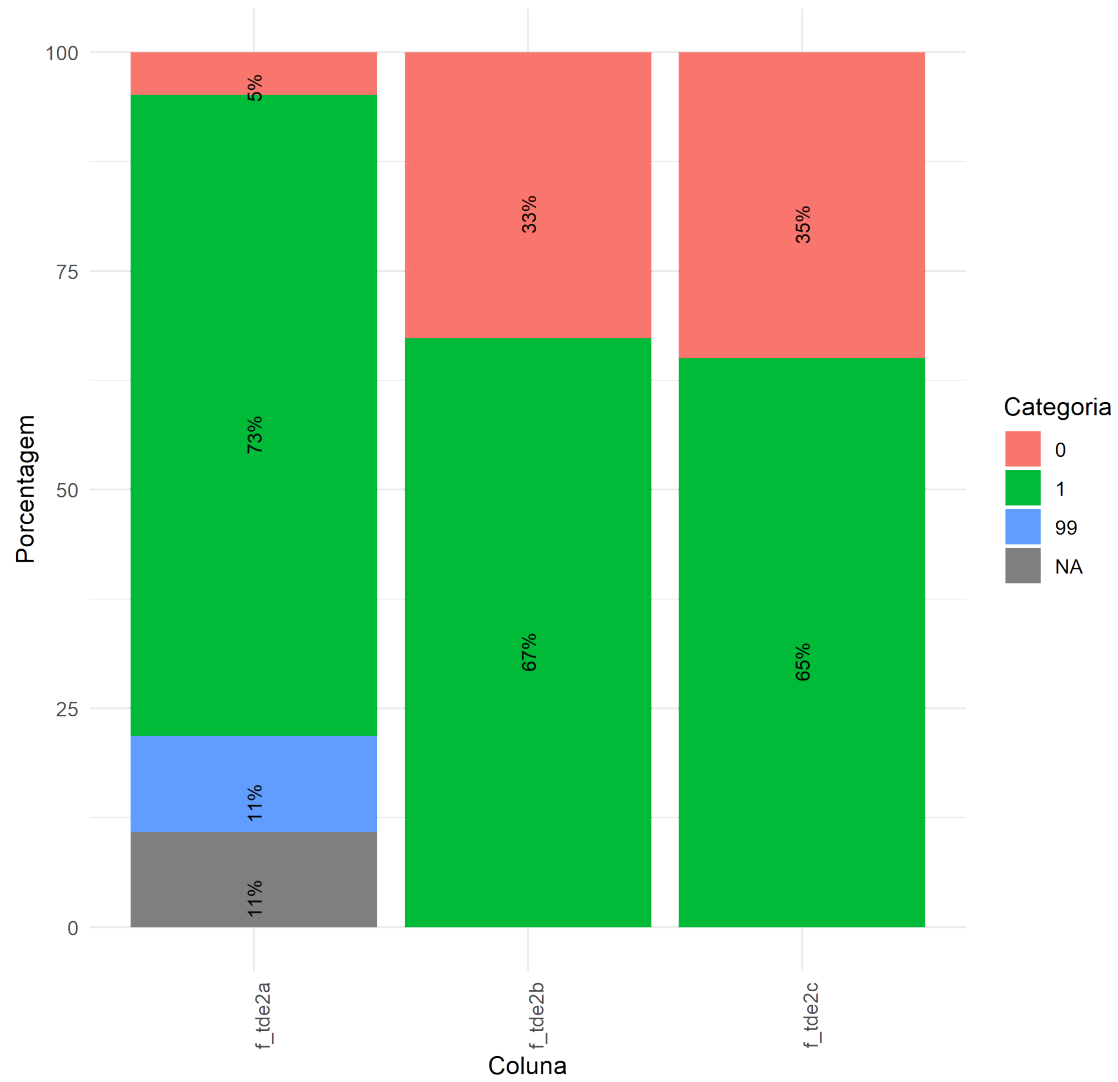


Figura A.1: Distribuição de frequências das variáveis da seção de Aritmética do teste TDE – Solução de problemas orais

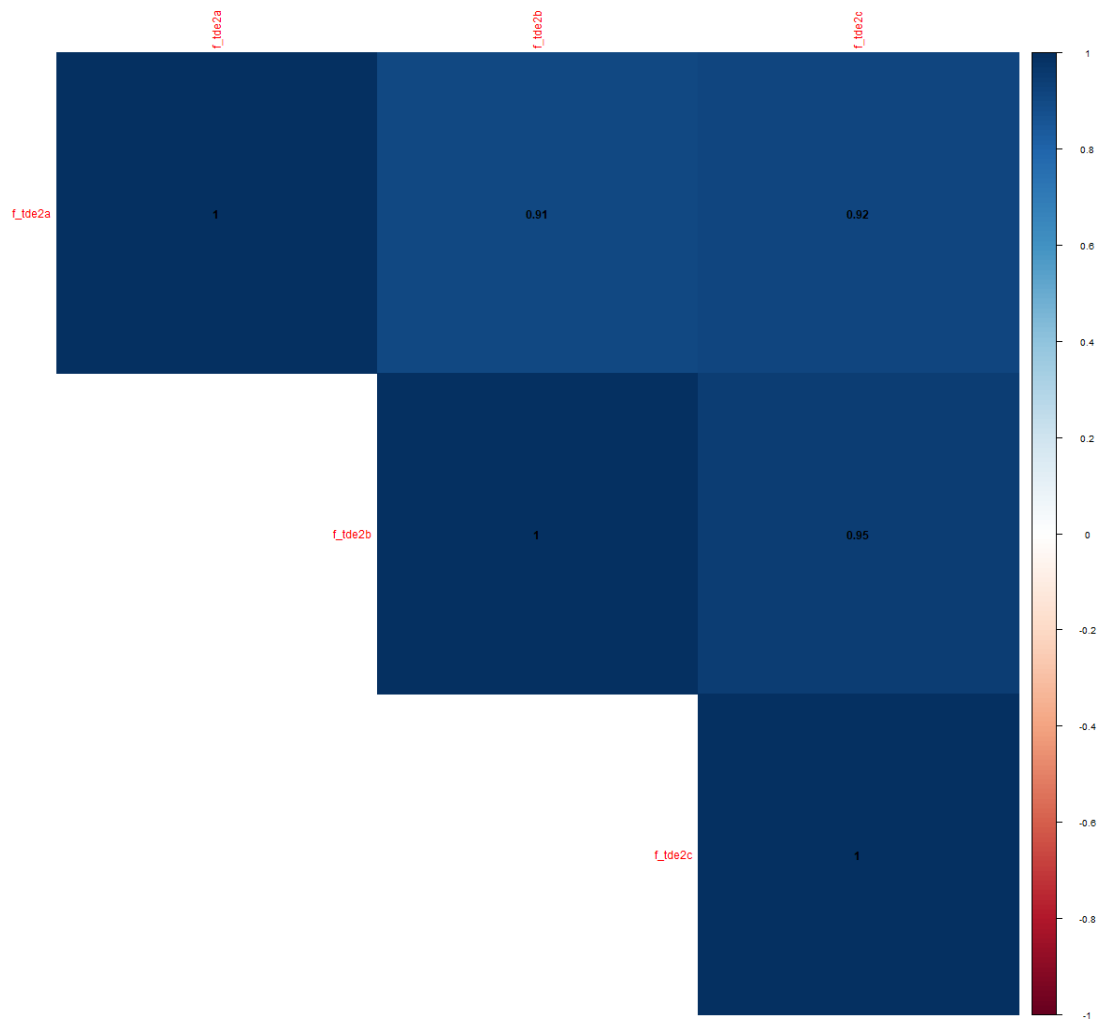


Figura A.2: Matriz de correlações entre as variáveis da seção de Aritmética do teste TDE – Solução de problemas orais

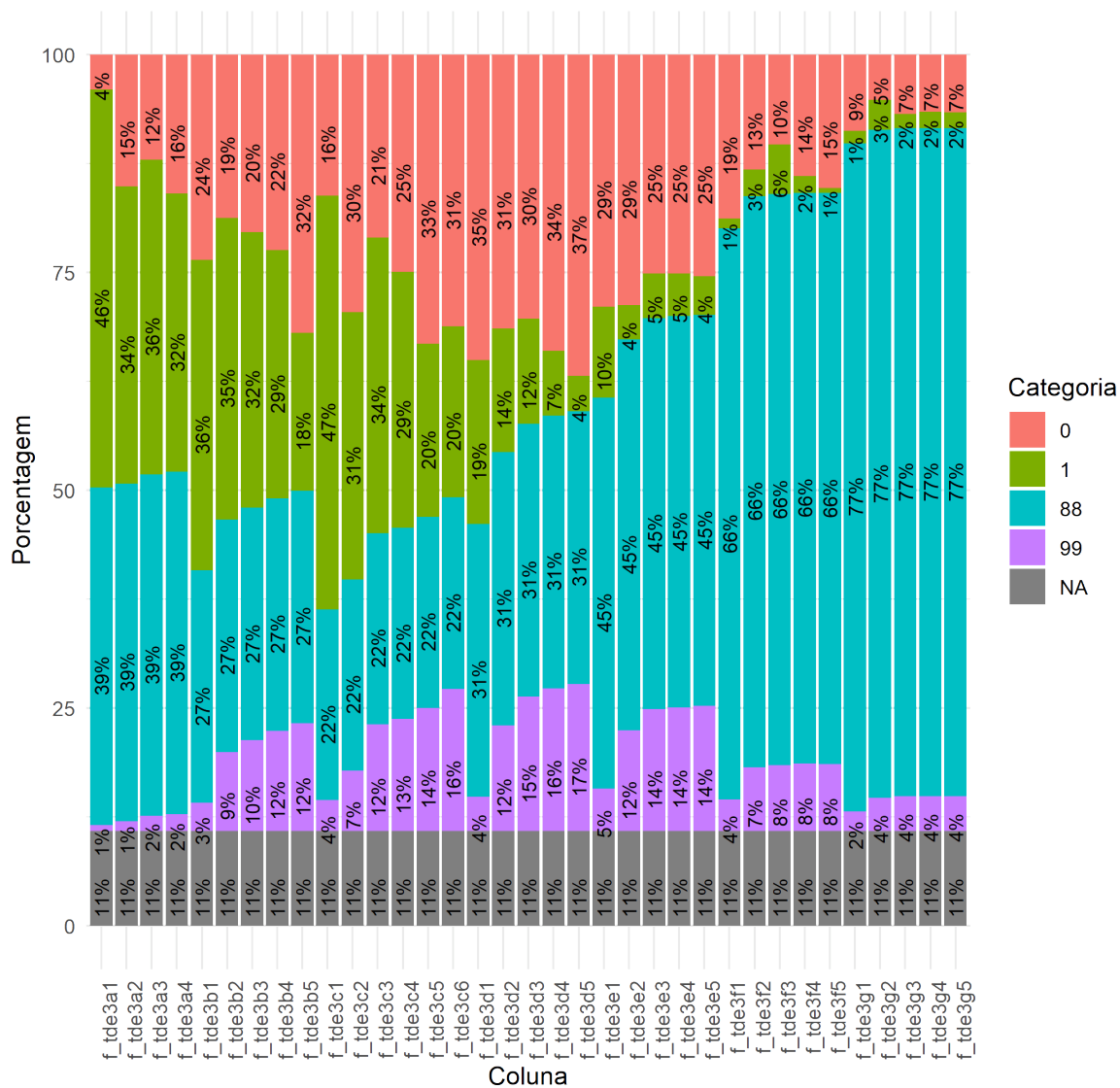


Figura A.3: Distribuição de frequências das variáveis do subtteste Solução de problemas escritos da seção de Aritmética do teste TDE

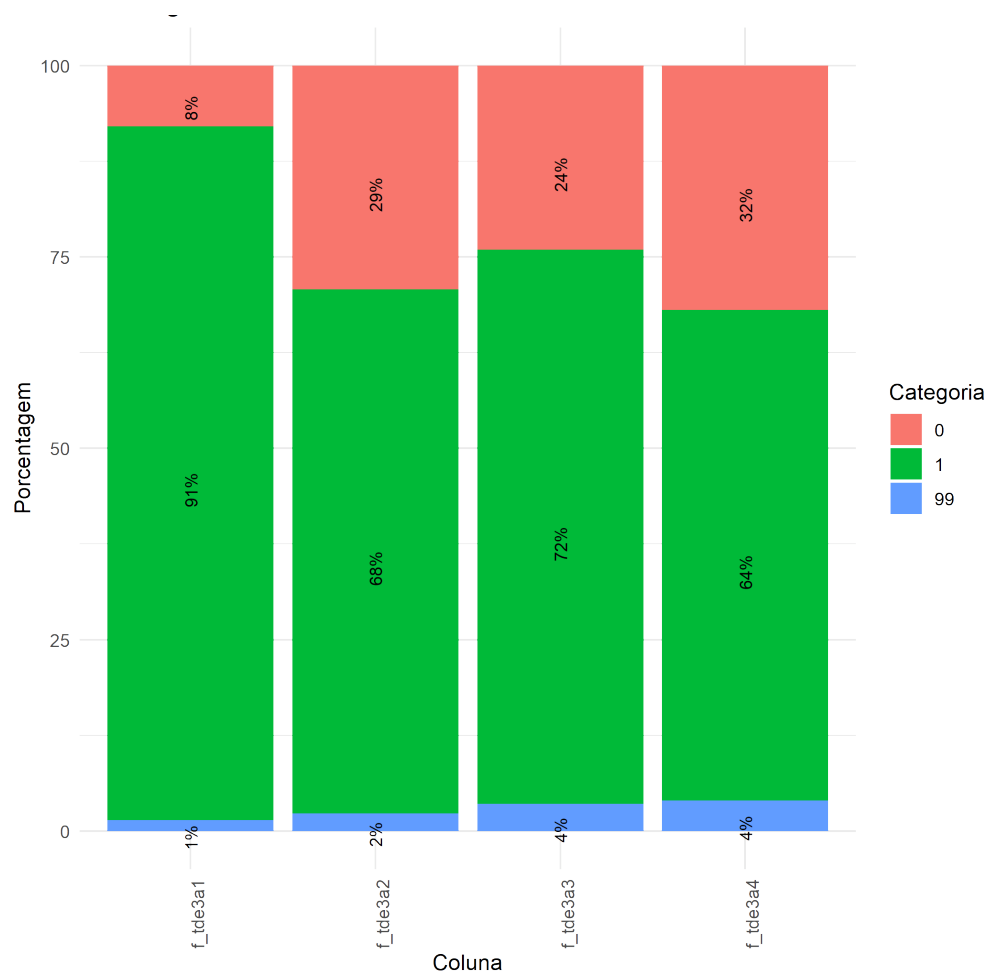


Figura A.4: Distribuição de frequências das variáveis do subtteste Solução de problemas escritos da seção de Aritmética do teste TDE, para o primeiro ano

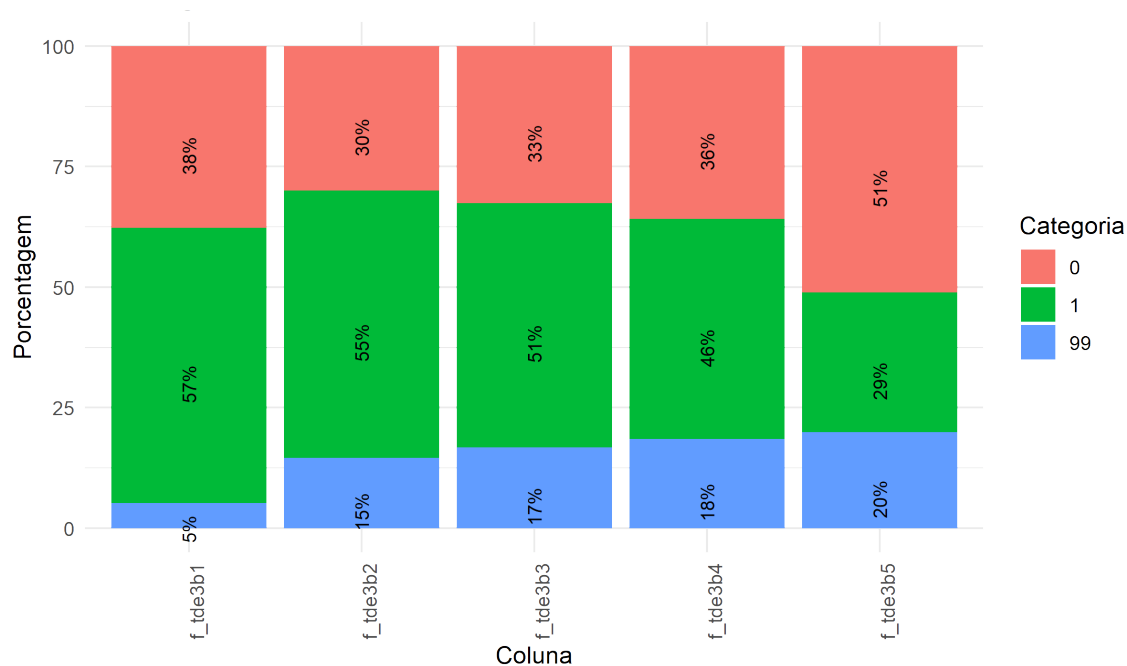


Figura A.5: Distribuição de frequências das variáveis do subteste Solução de problemas escritos da seção de Aritmética do teste TDE, para o segundo ano.

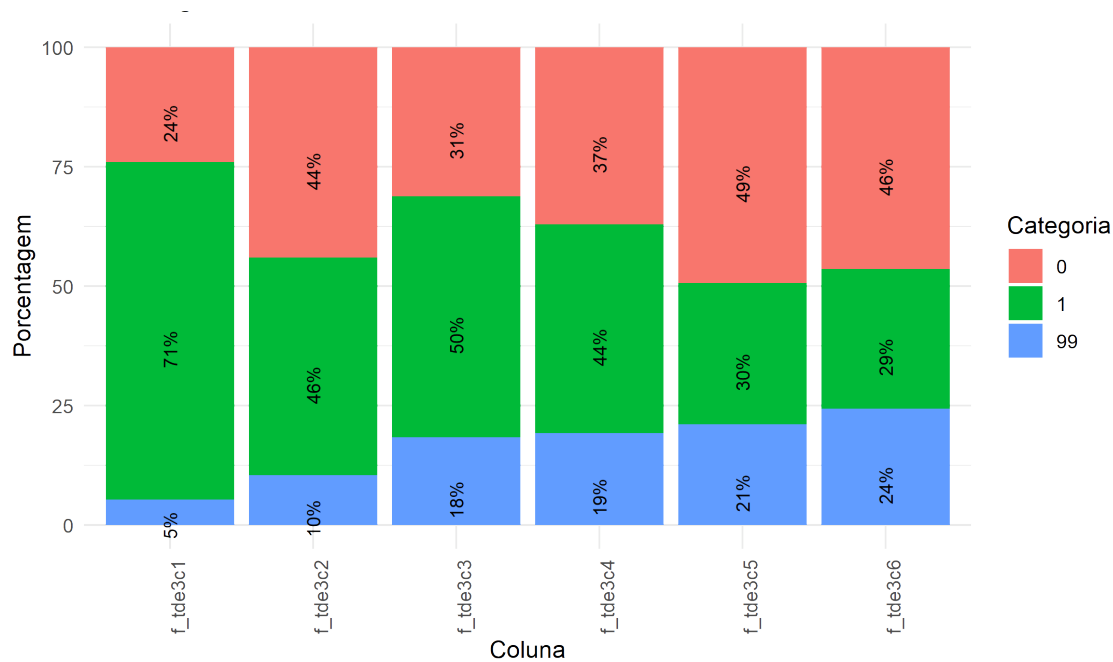


Figura A6: Percentual de categorias do subteste escrito de aritmética do TDE para o terceiro ano

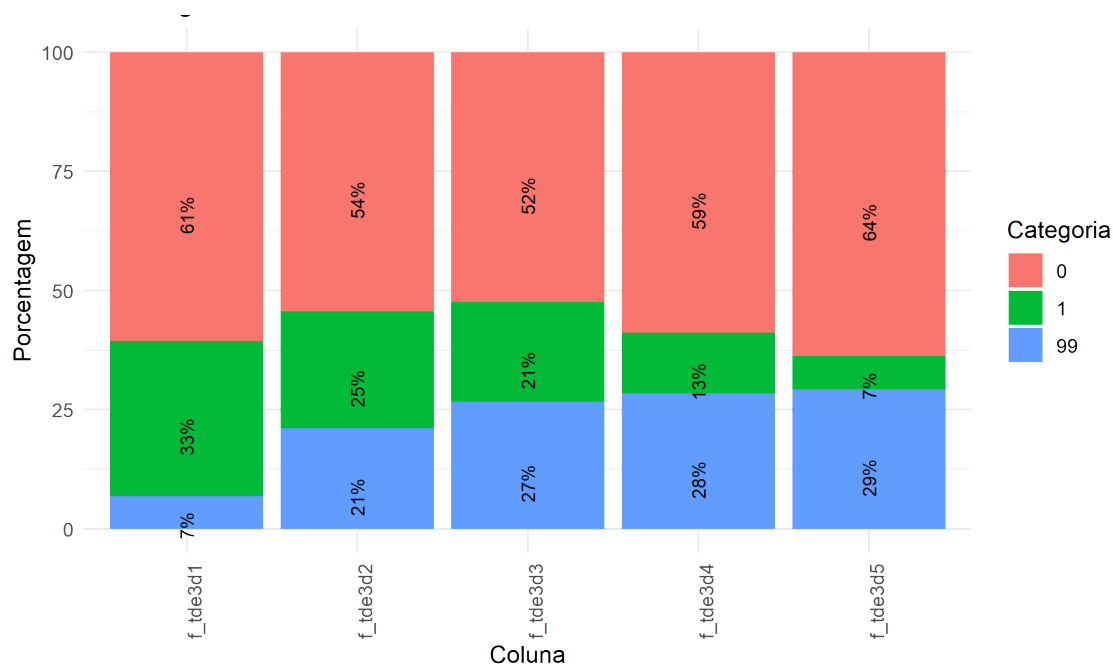


Figura A7: Percentual de categorias do substeste escrito de aritmética do TDE para o quarto ano

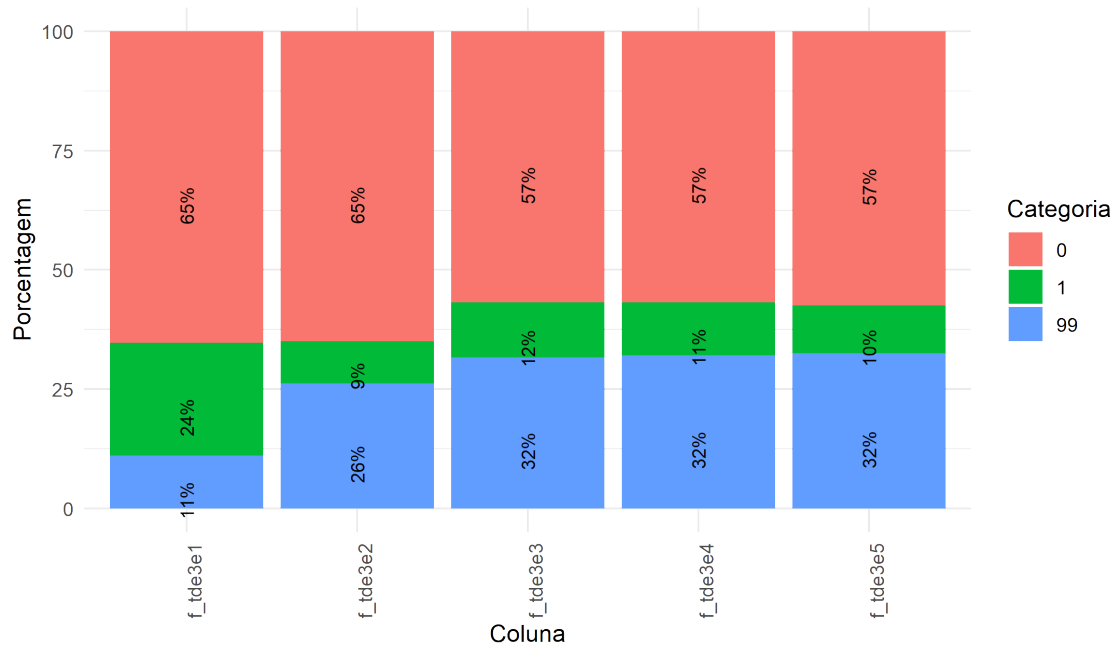


Figura A.8 : Percentual de categorias do substeste escrito de aritmética do TDE para o quinto ano

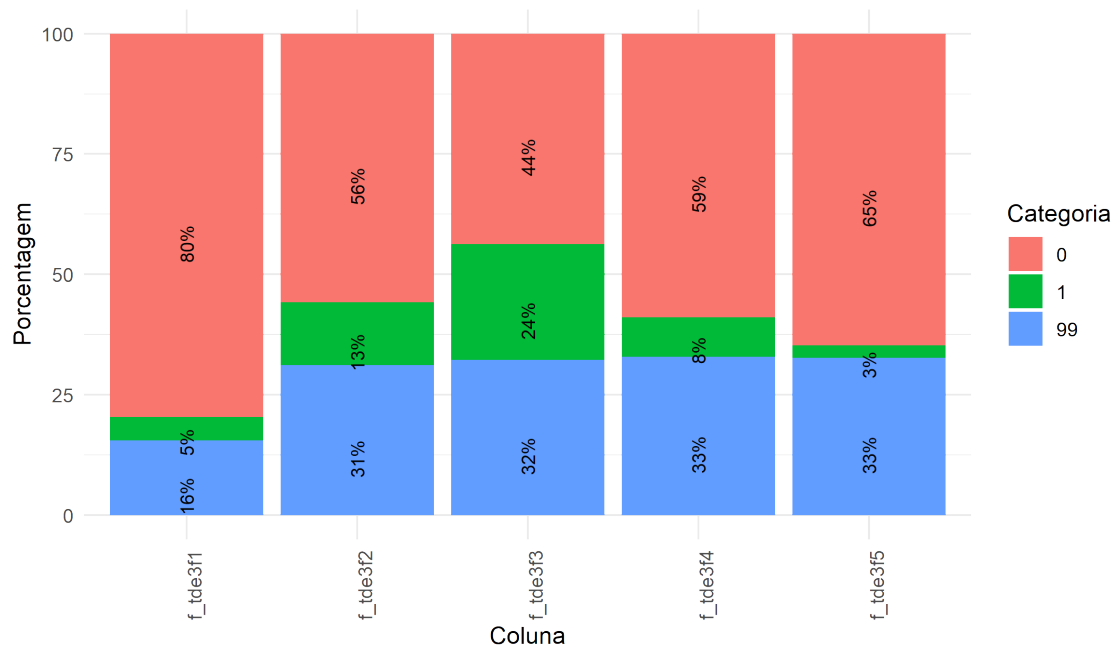


Figura A.9 : Percentual de categorias do subteste escrito de aritmética do TDE para o sexto ano

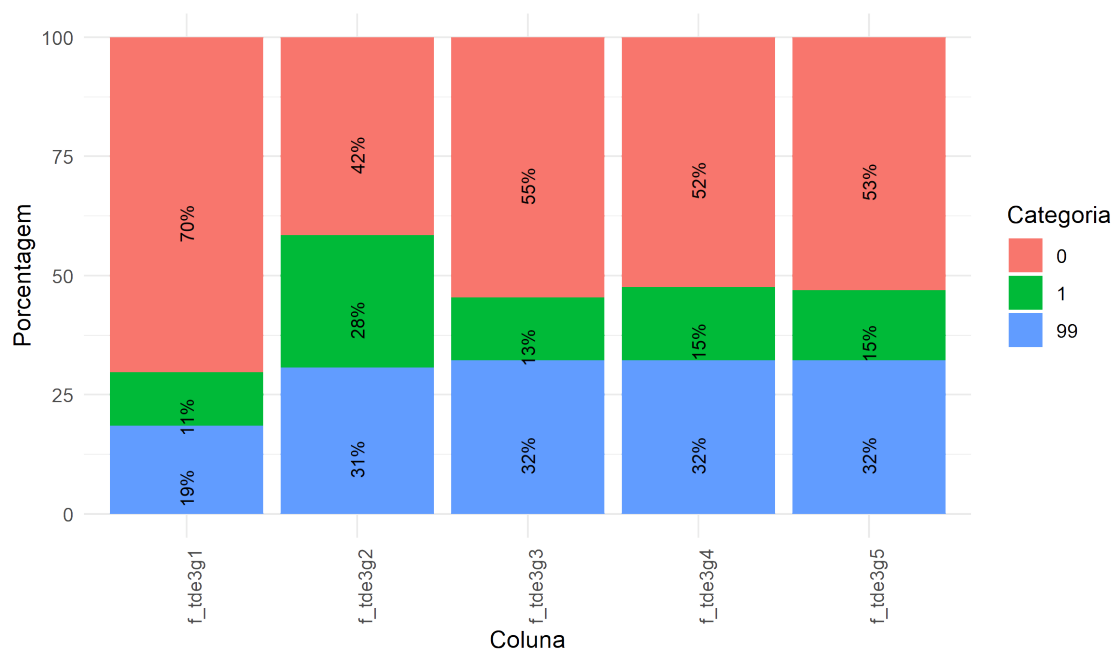


Figura A.10: Percentual de categorias do subteste escrito de aritmética do TDE para o sétimo ano

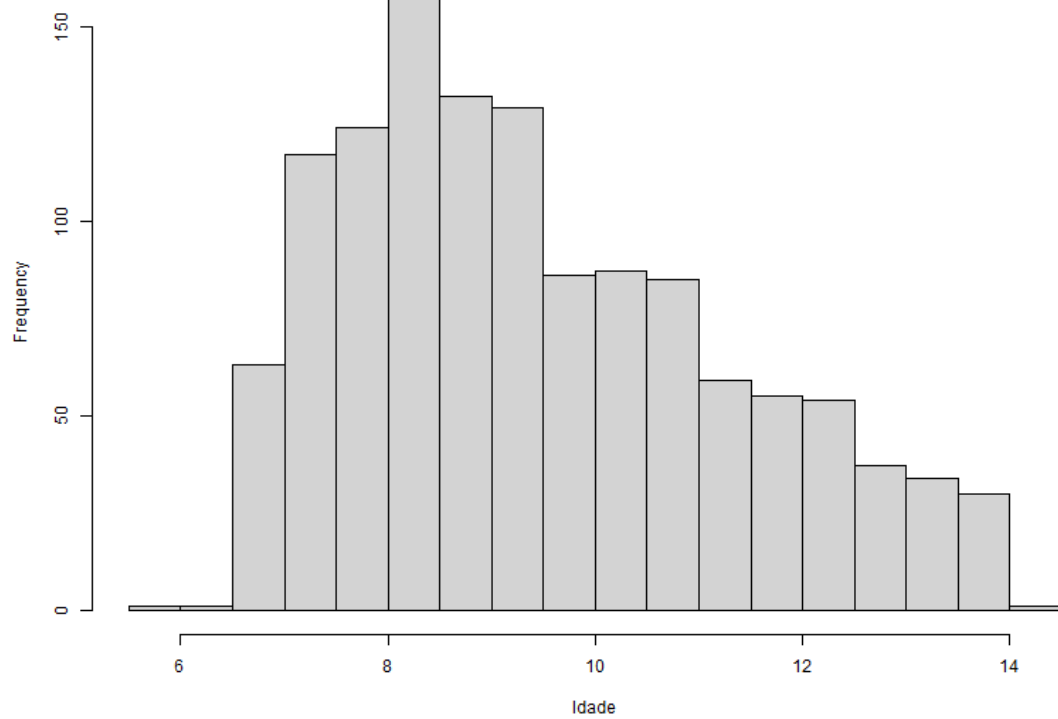


Figura A11: Histograma das idades das crianças que realizaram o subteste Solução de problemas escritos da seção de Aritmética do teste TDE para o primeiro ano.

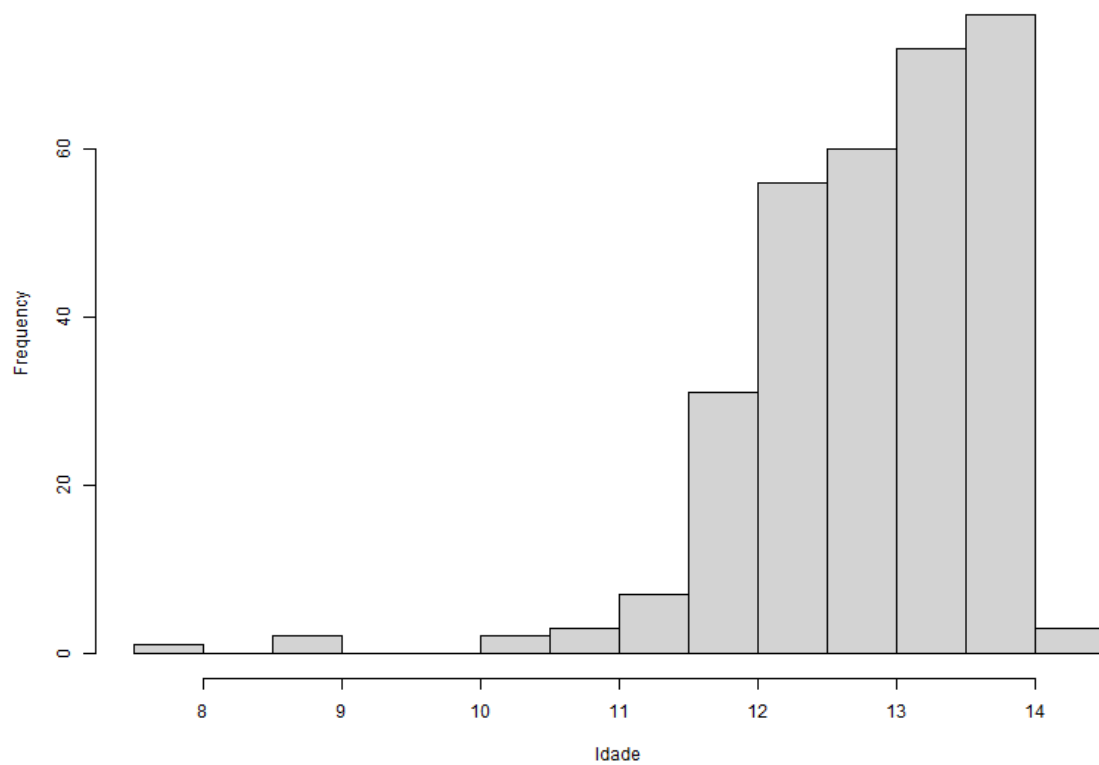


Figura A12: Histograma das idades das crianças que realizaram o subteste Solução de problemas escritos da seção de Aritmética do teste TDE para o sétimo ano.

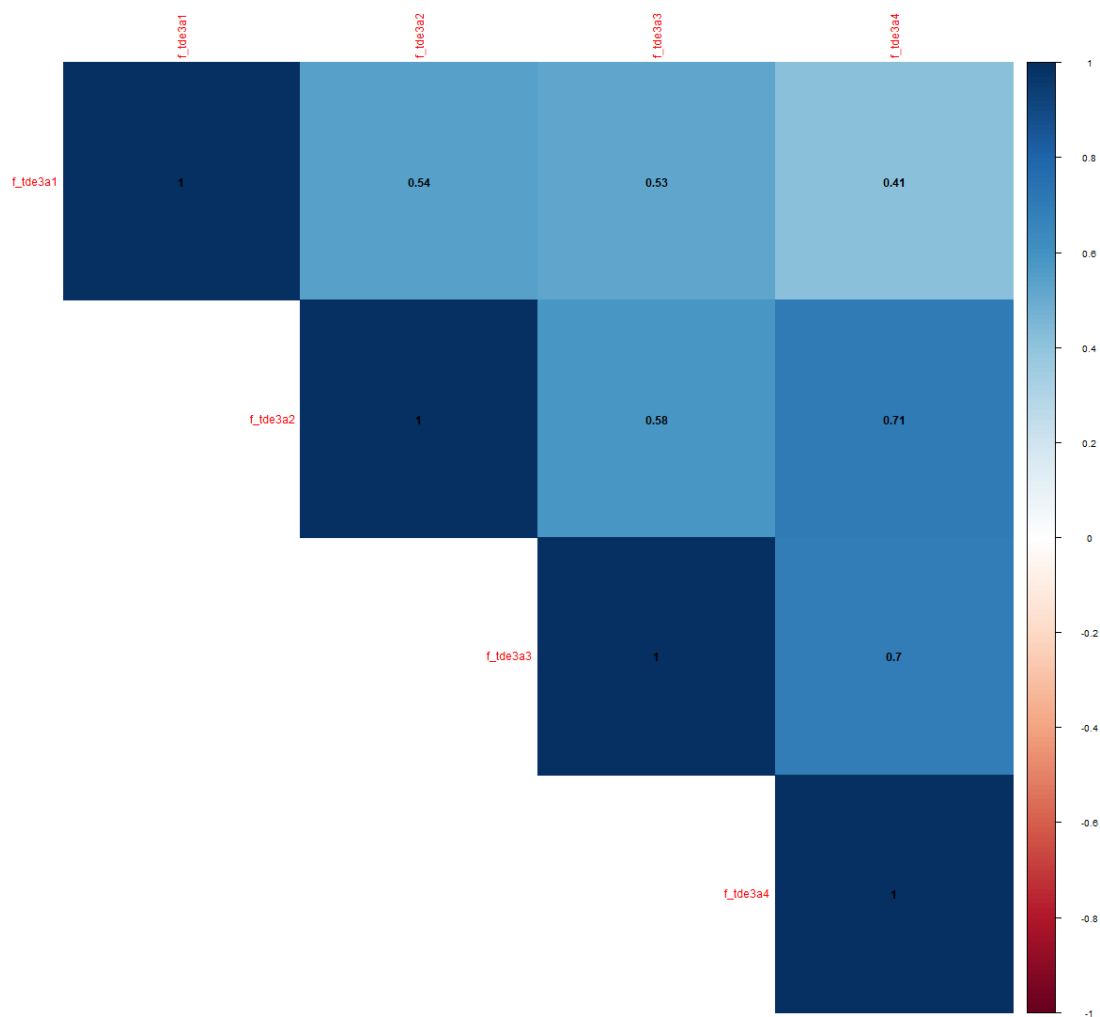


Figura A.13: Matriz de correlações entre as variáveis do subtteste Solução de problemas escritos da seção de Aritmética do teste TDE para o primeiro ano.

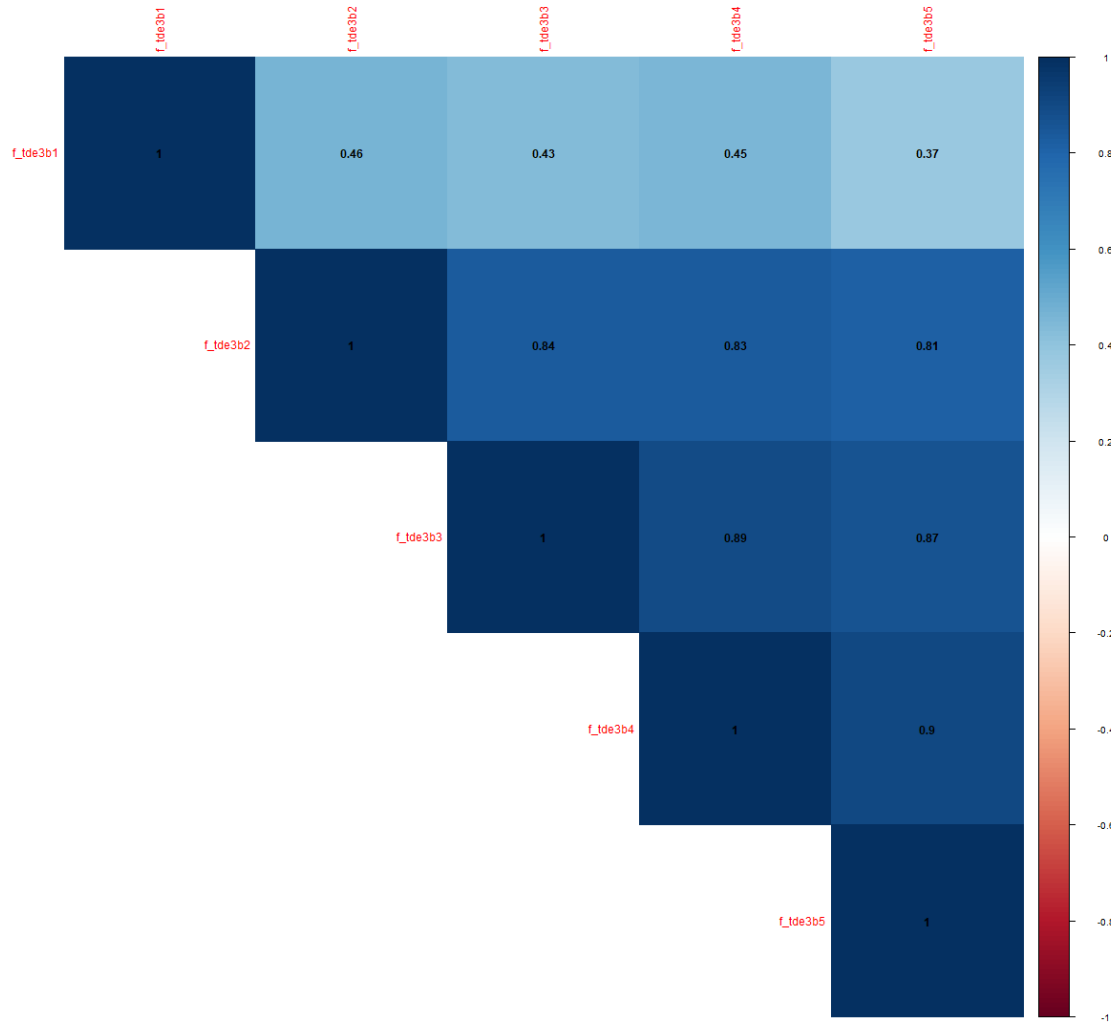


Figura A.14 Correlações do subteste escrito de aritmética do TDE para o segundo ano

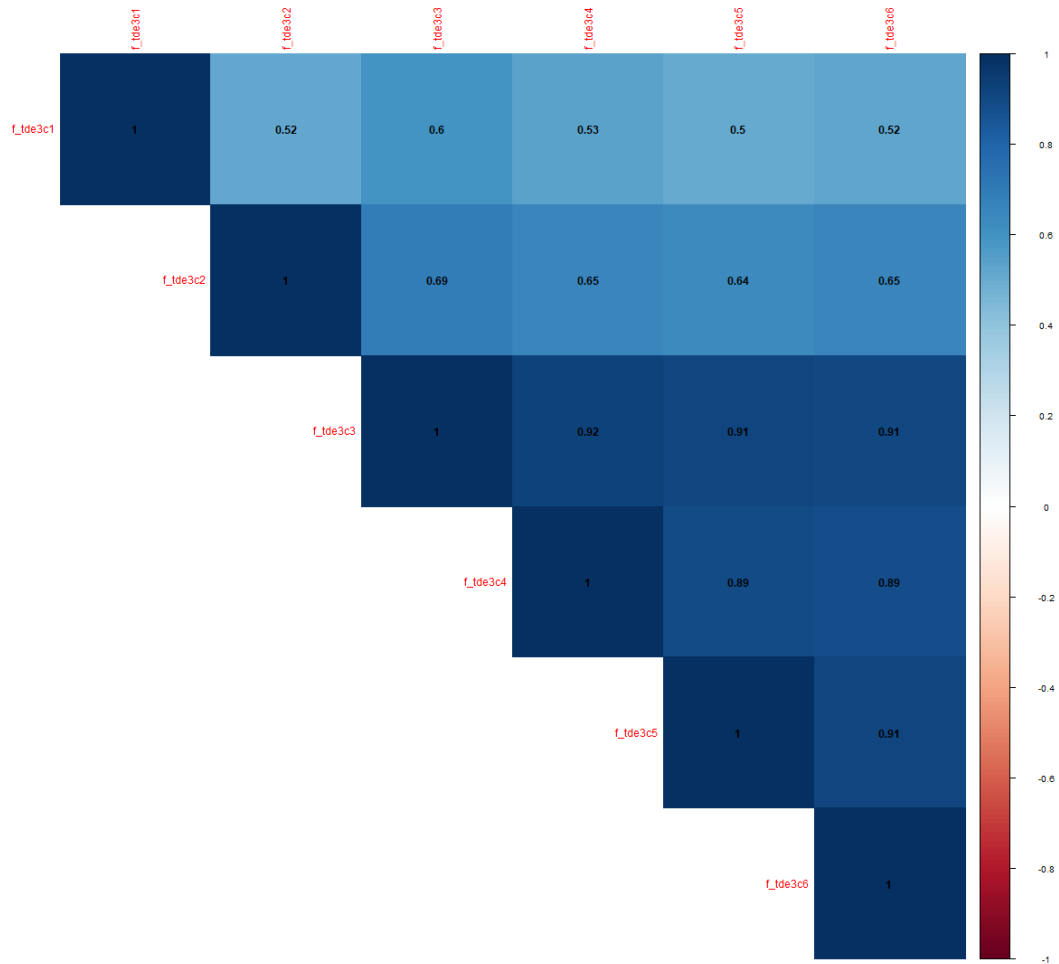


Figura A15 : Correlações do subteste escrito de aritmética do TDE para o terceiro ano

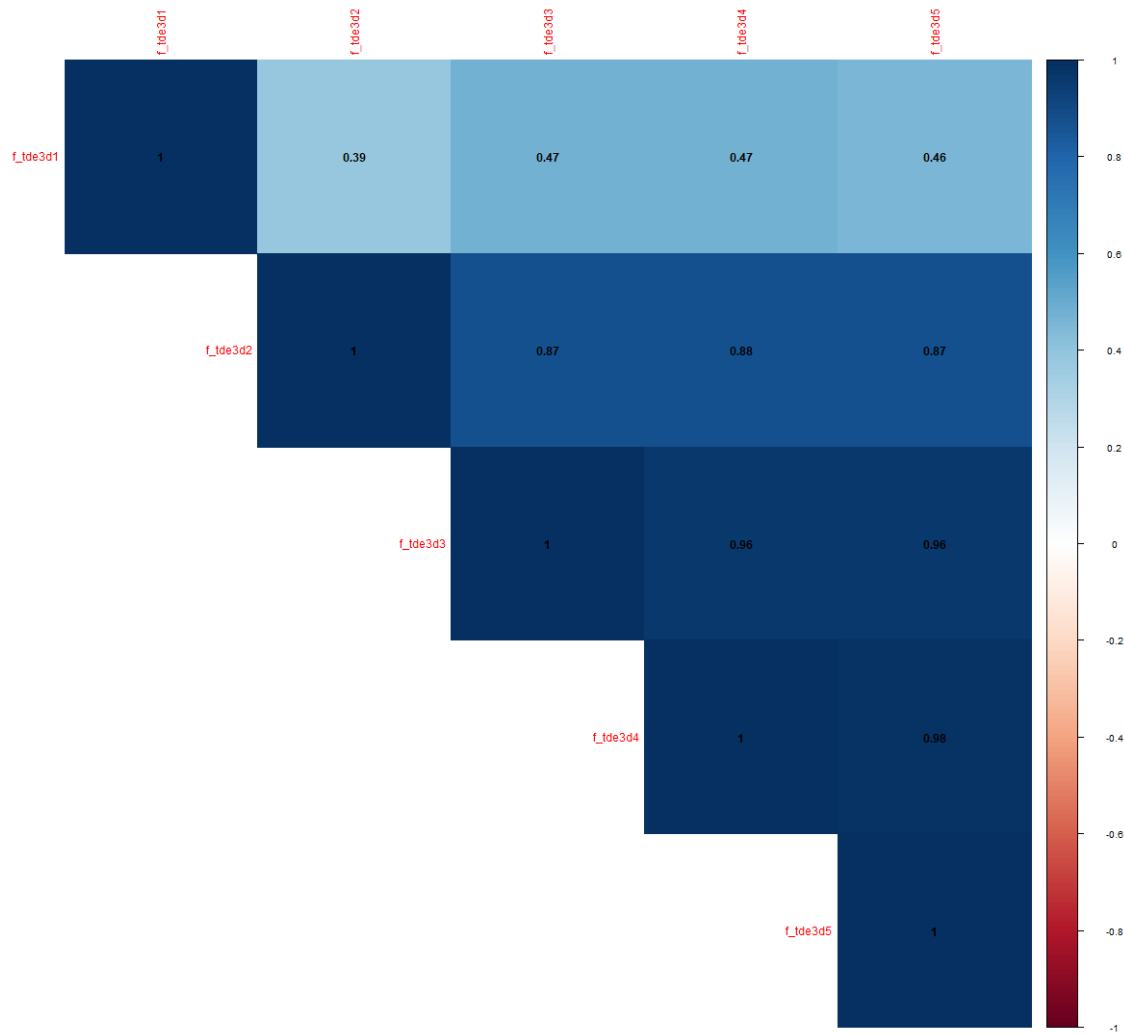


Figura A16: Correlações do substeste escrito de aritmética do TDE para o quarto ano

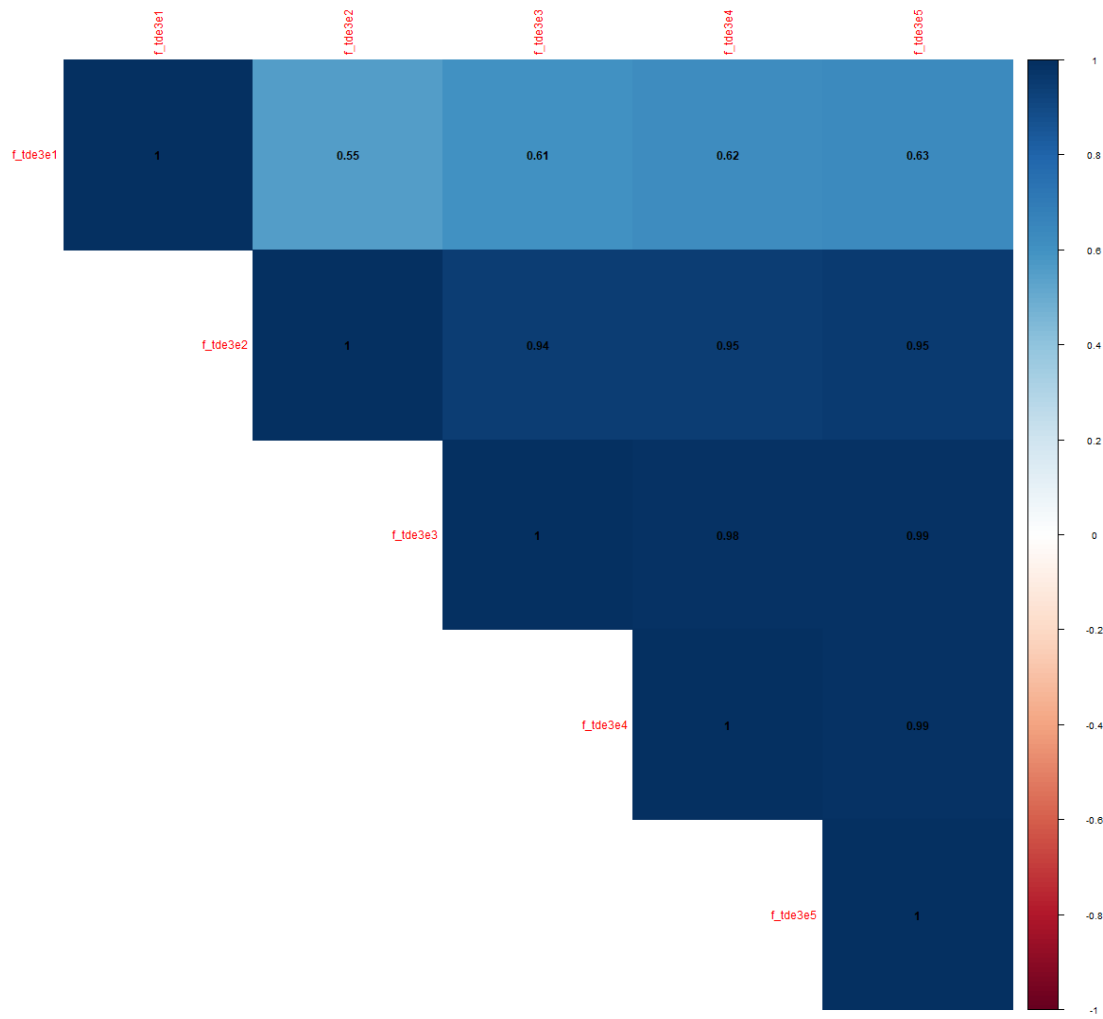


Figura A17: Correlações do subteste escrito de aritmética do TDE para o quinto ano

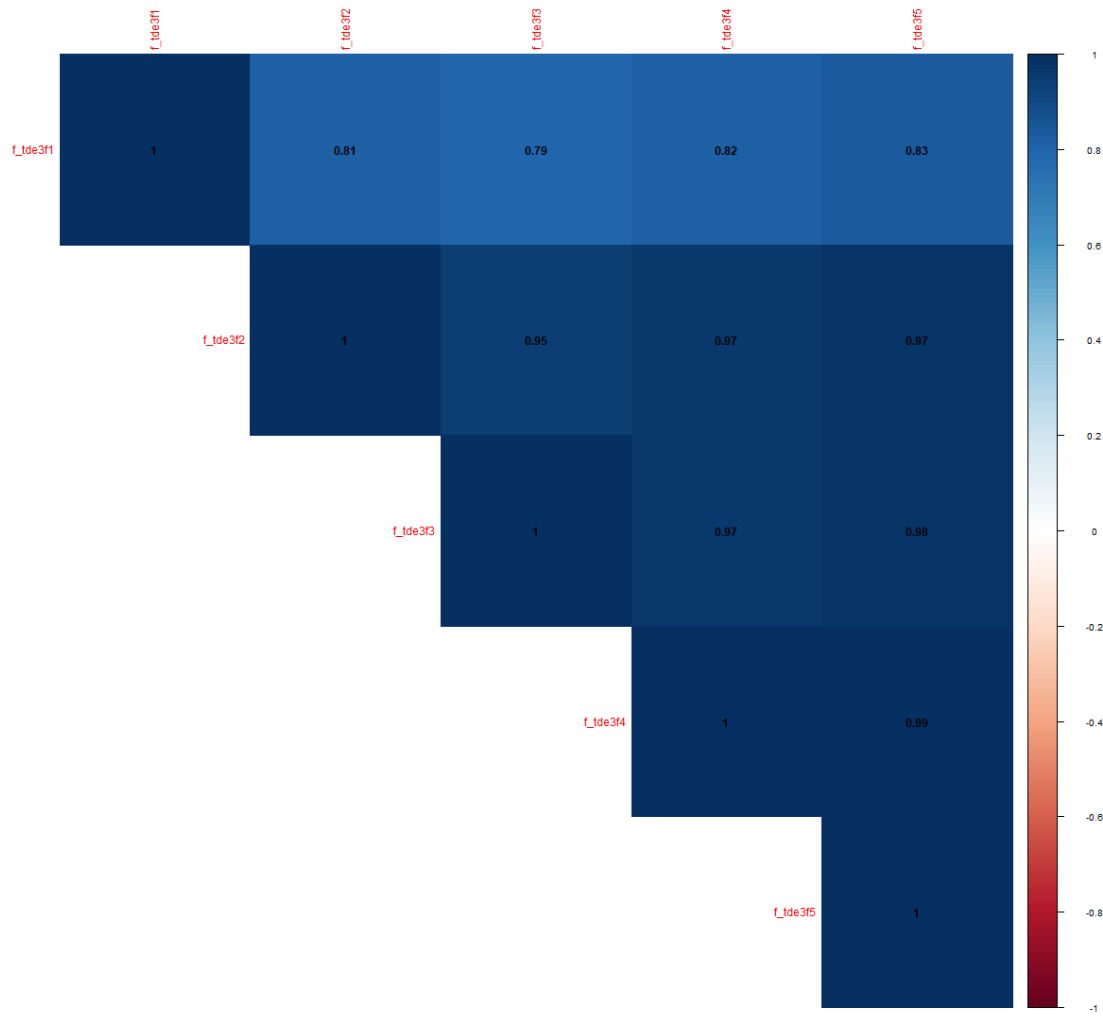


Figura A18: Correlações do subteste escrito de aritmética do TDE para o sexto ano

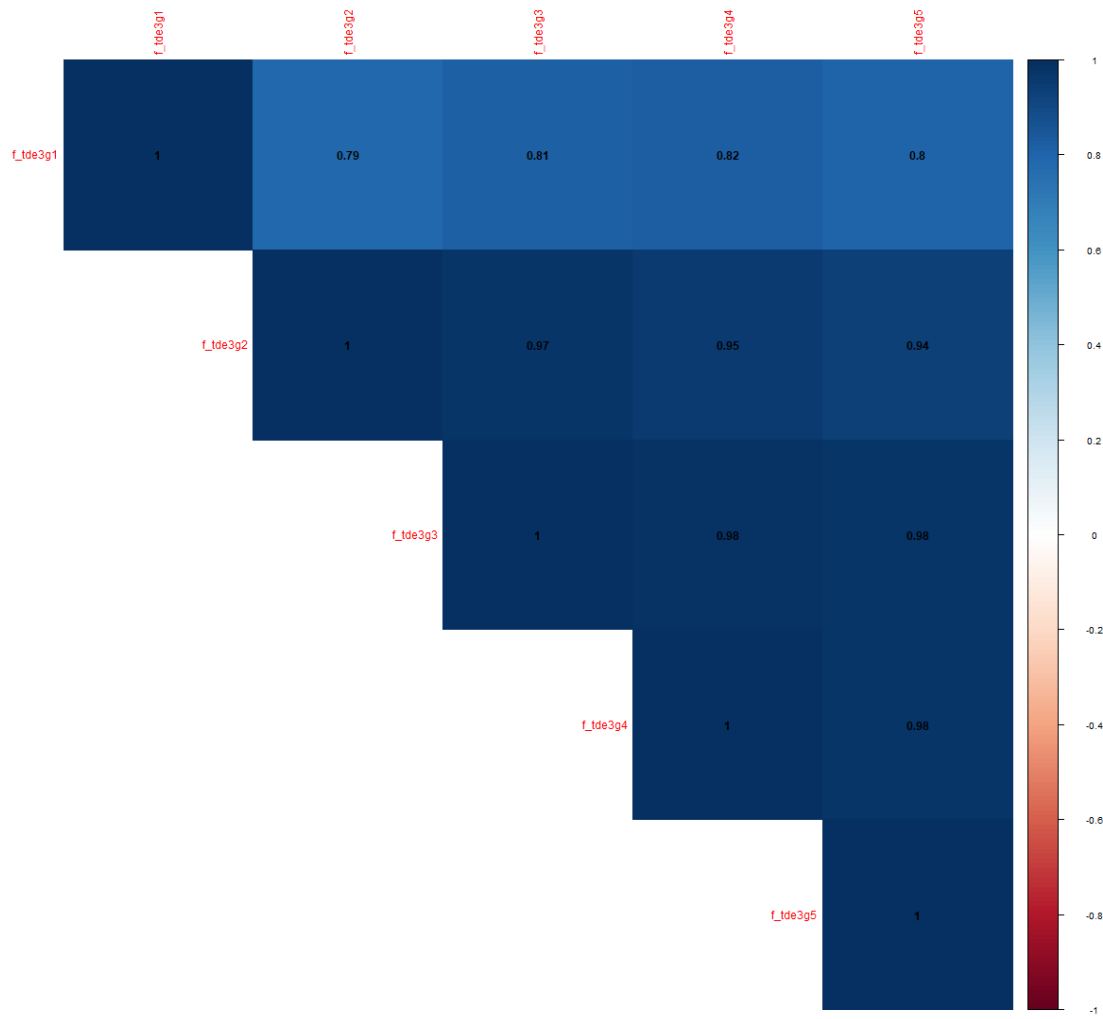


Figura A19: Correlações do subteste escrito de aritmética do TDE para o sétimo ano

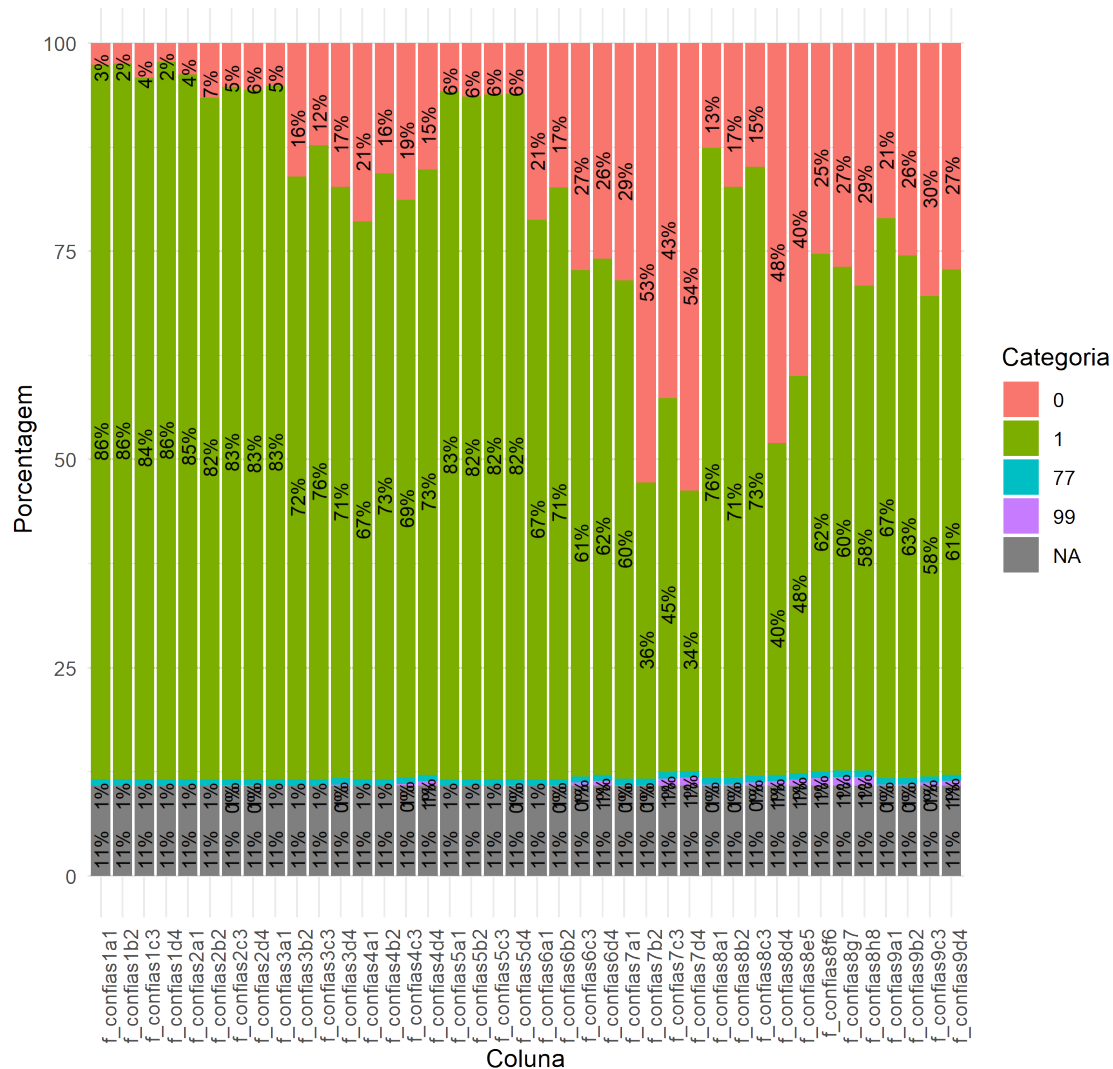


Figura A.20: Distribuição de frequências das variáveis do Nível da sílaba do teste Confias

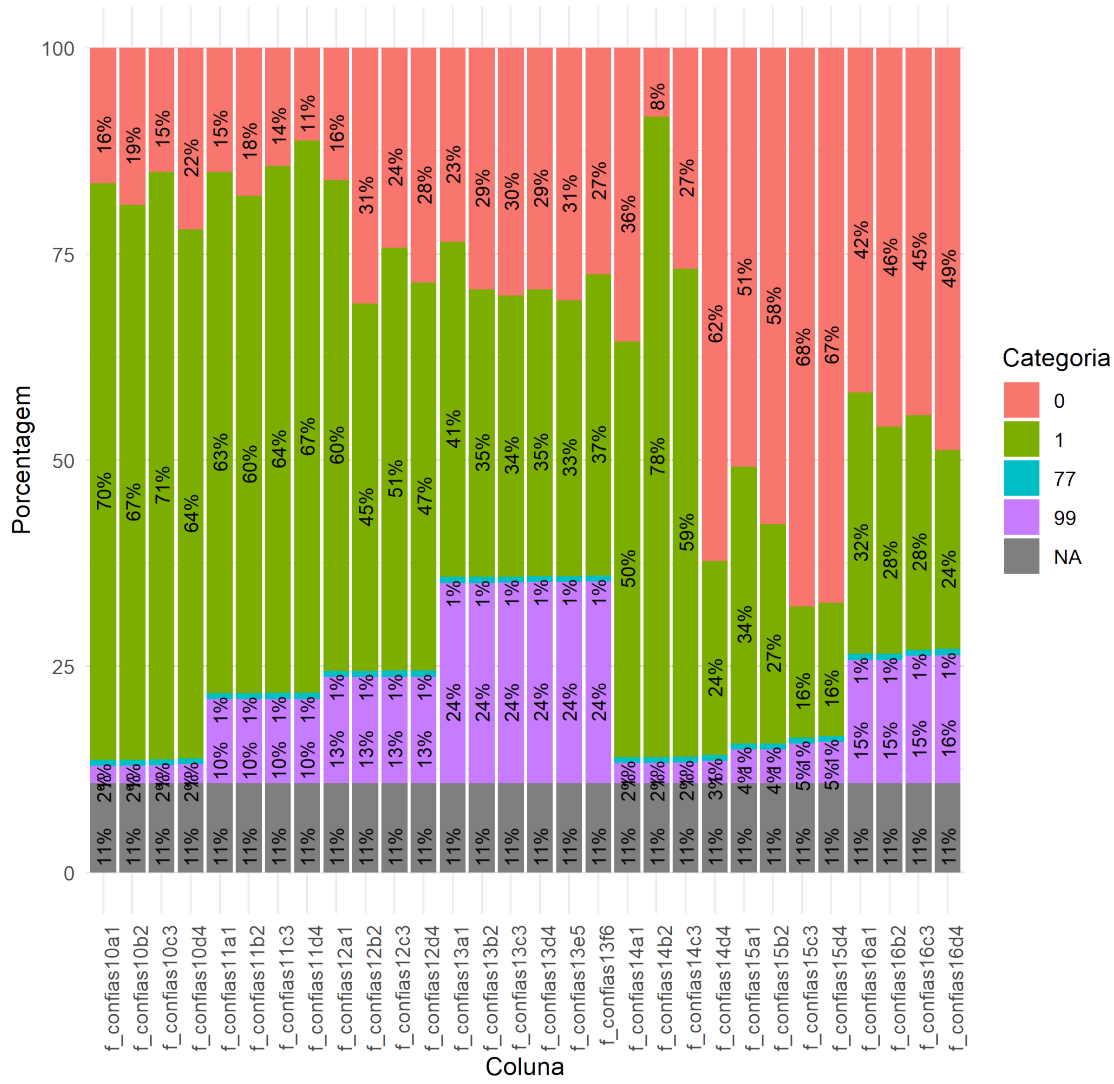


Figura A.22: Distribuição de frequências das variáveis do Nível do fonema do teste Confias

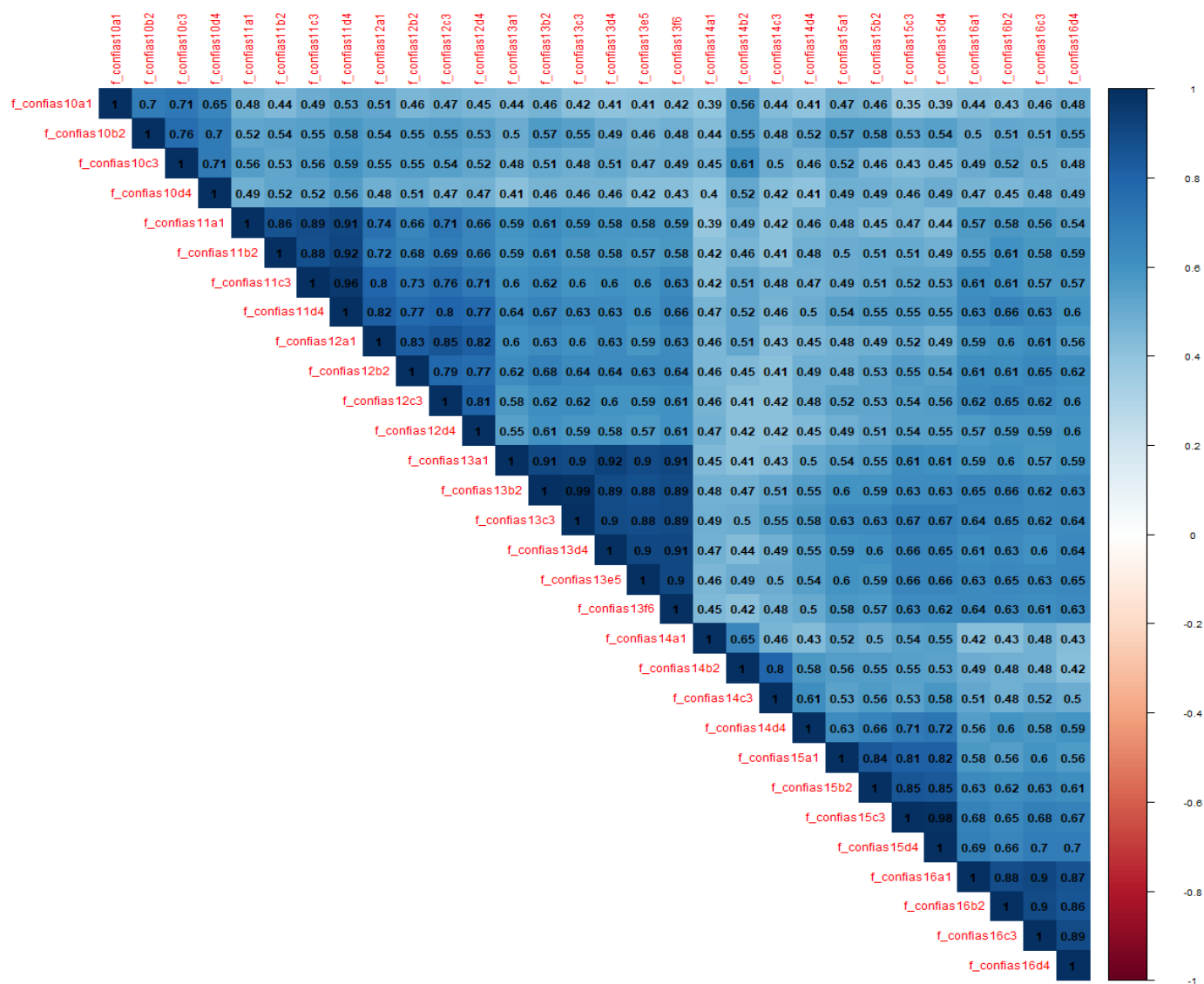


Figura A.23: Matriz de correlações entre as variáveis do Nível do fonema do teste Confias.

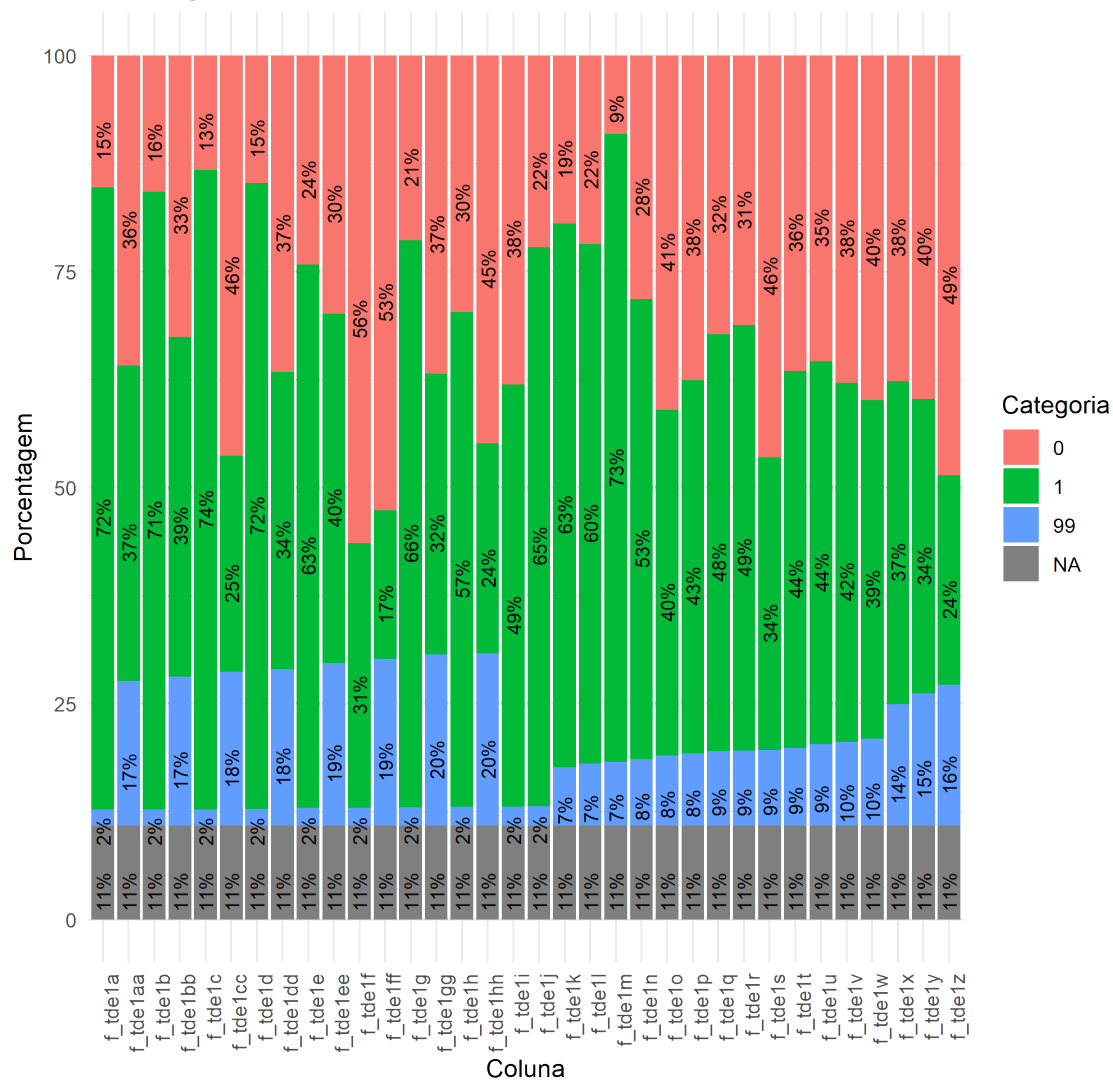


Figura A.24: Distribuição de frequências das variáveis o teste TDE escrita

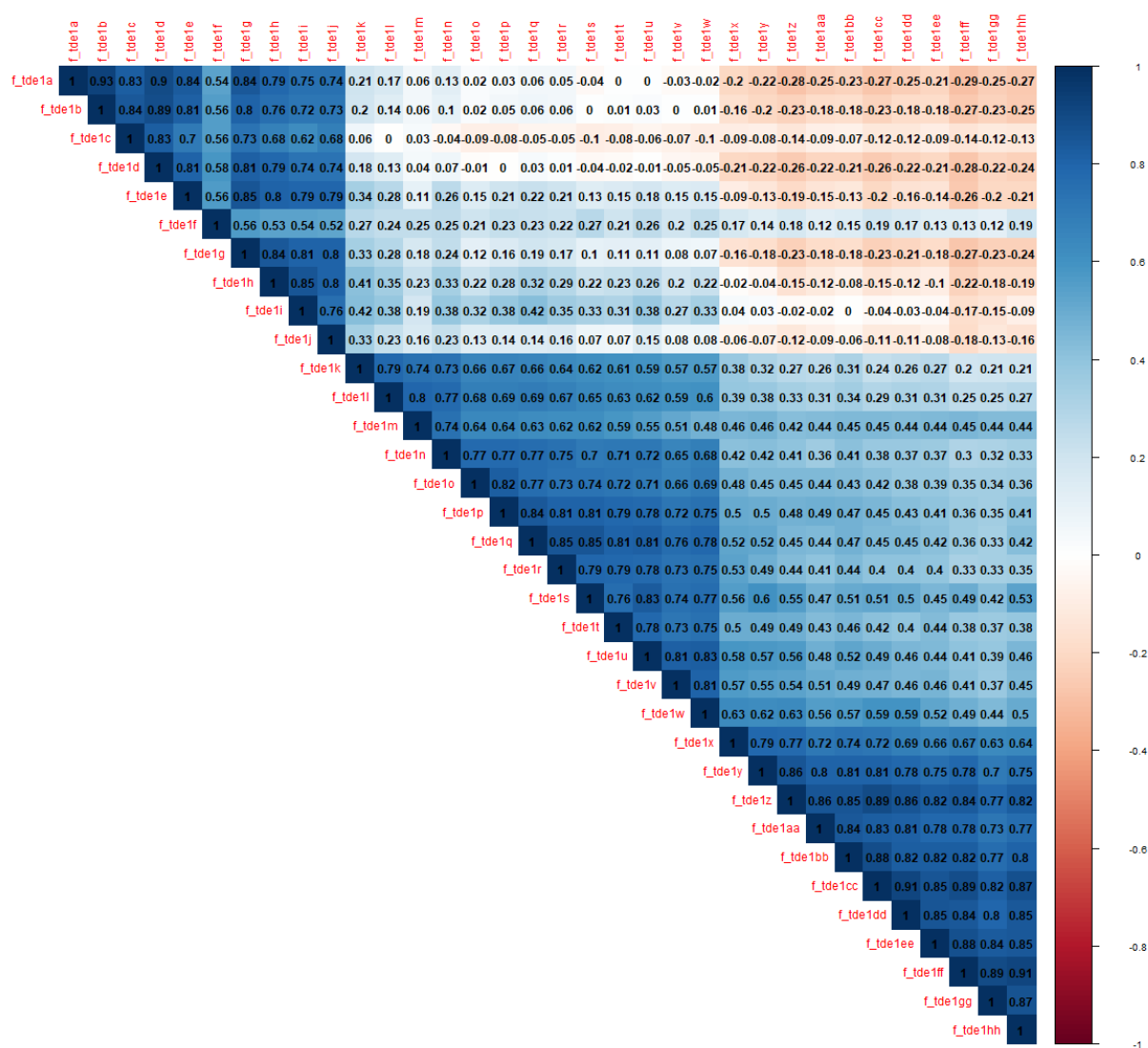


Figura A25: Matriz de correlações entre as variáveis do TDE escrita

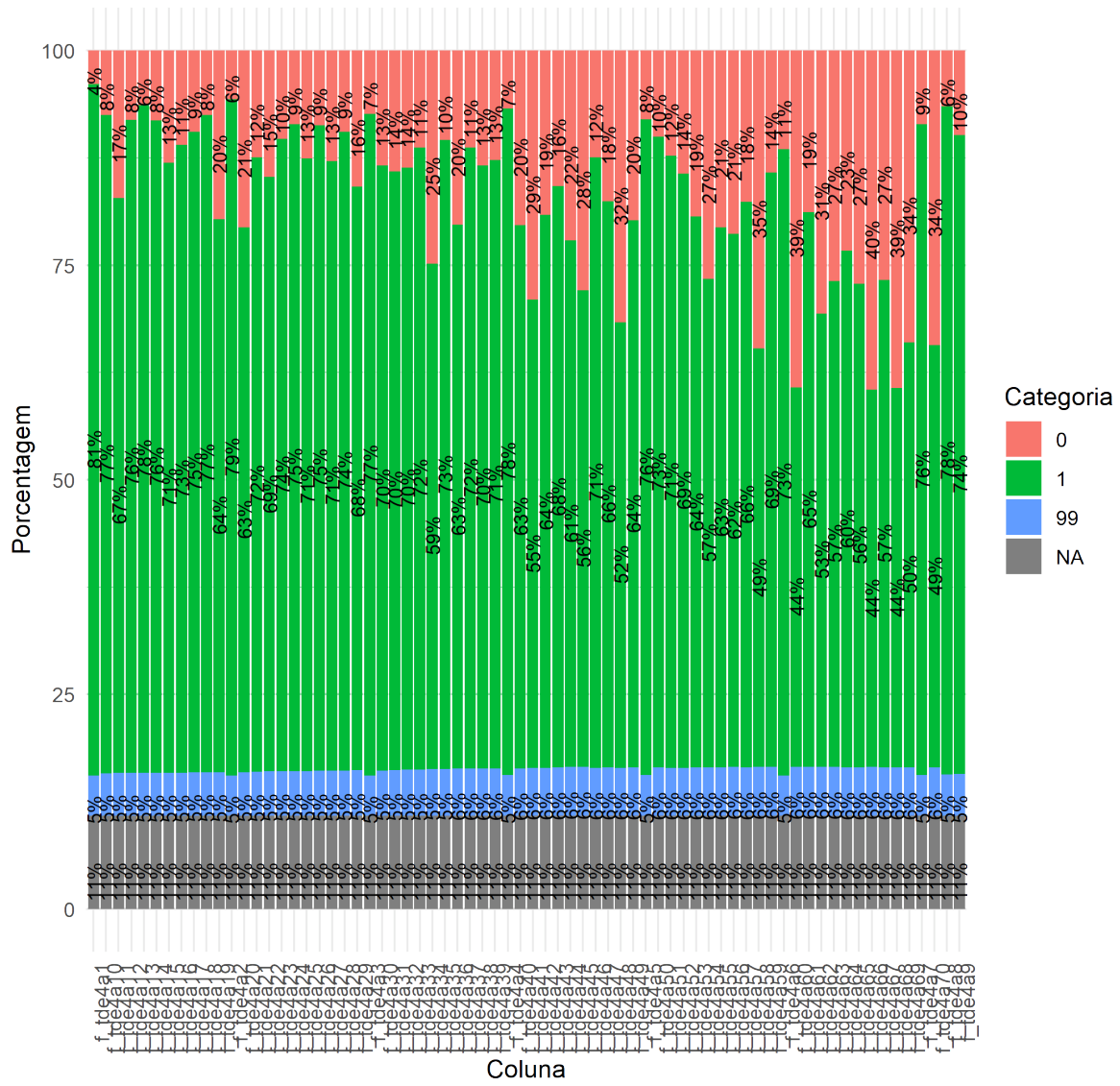


Figura A26: Percentual de categorias no teste TDE leitura

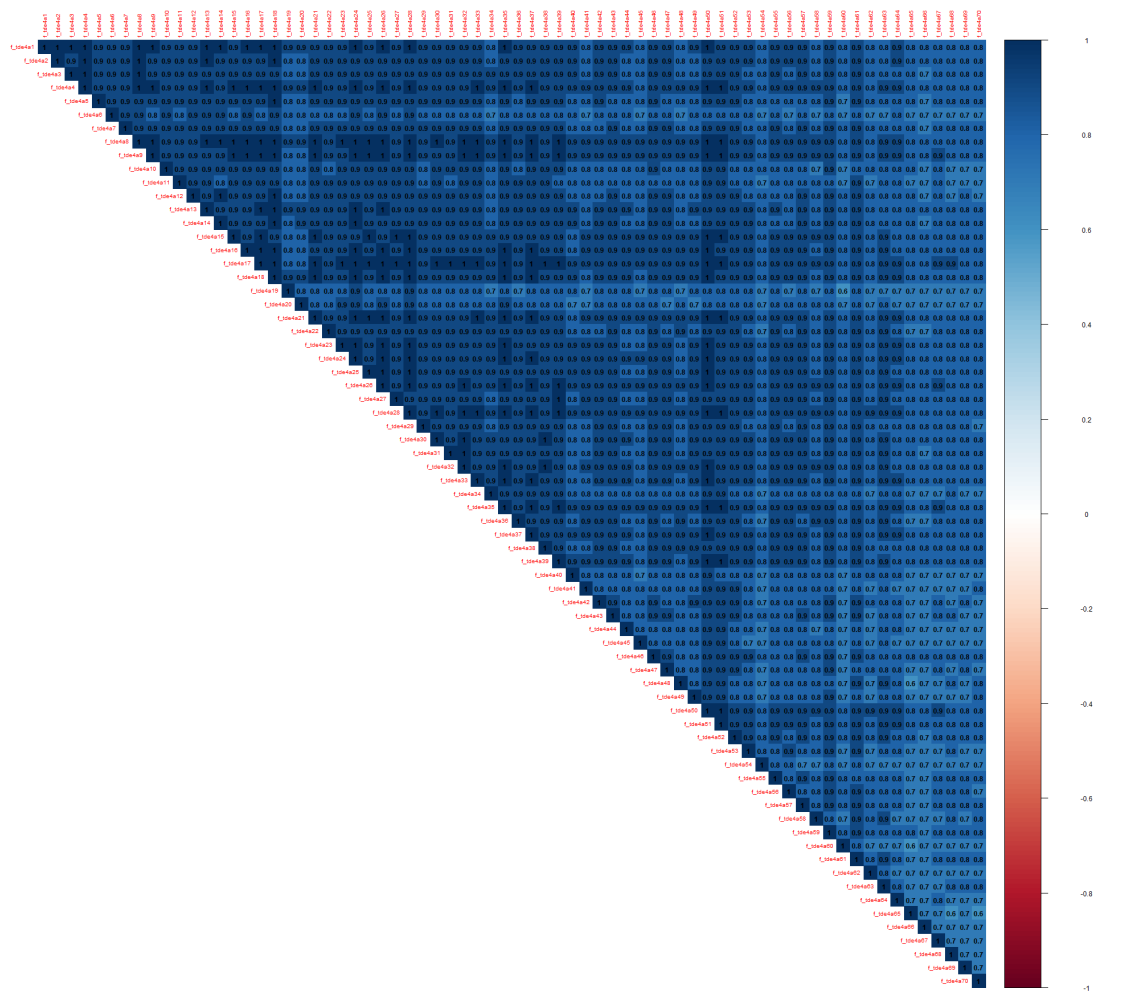


Figura A27: Correlações das variáveis do teste TDE Leitura

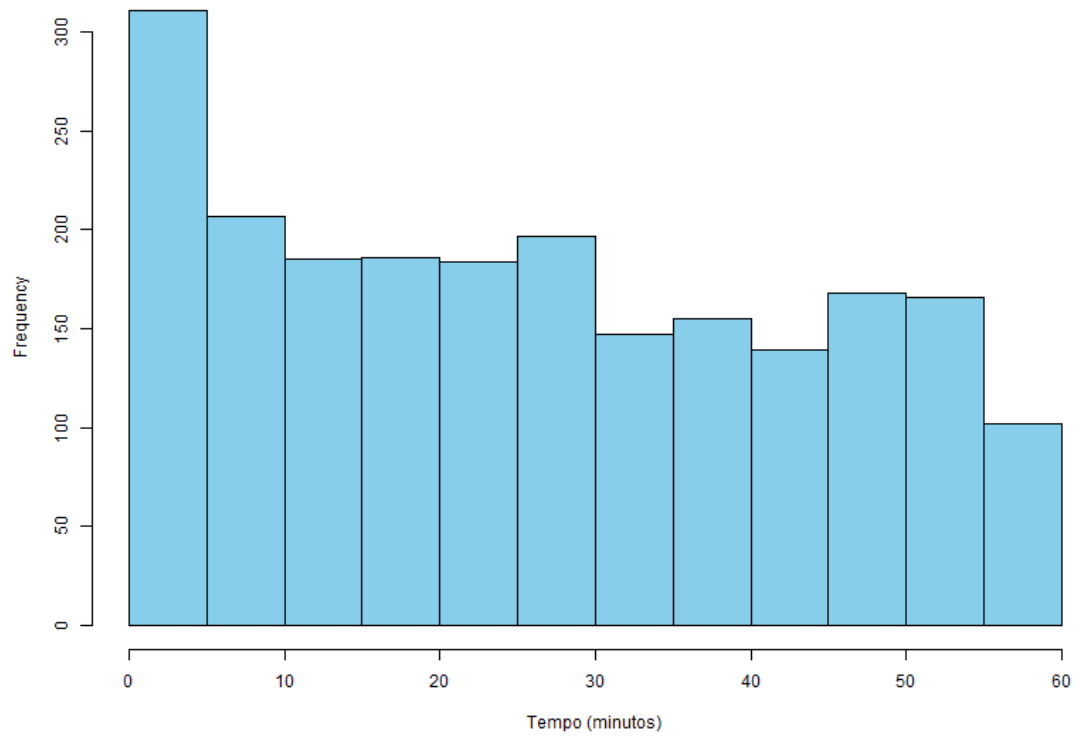


Figura A28: Histograma do tempo de ditado TDE Leitura

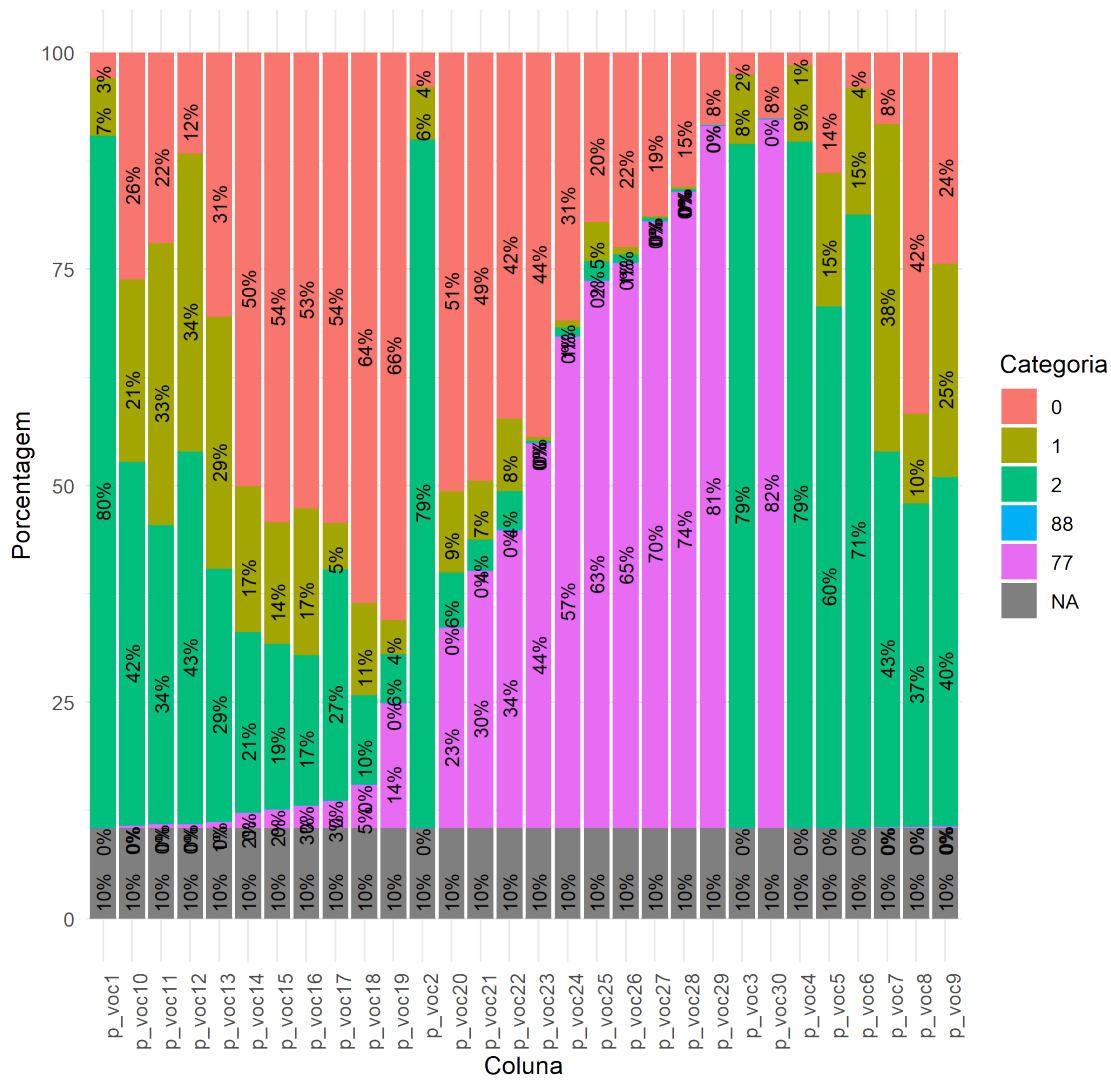


Figura A.29: Percentual de categorias no teste de Vocabulário

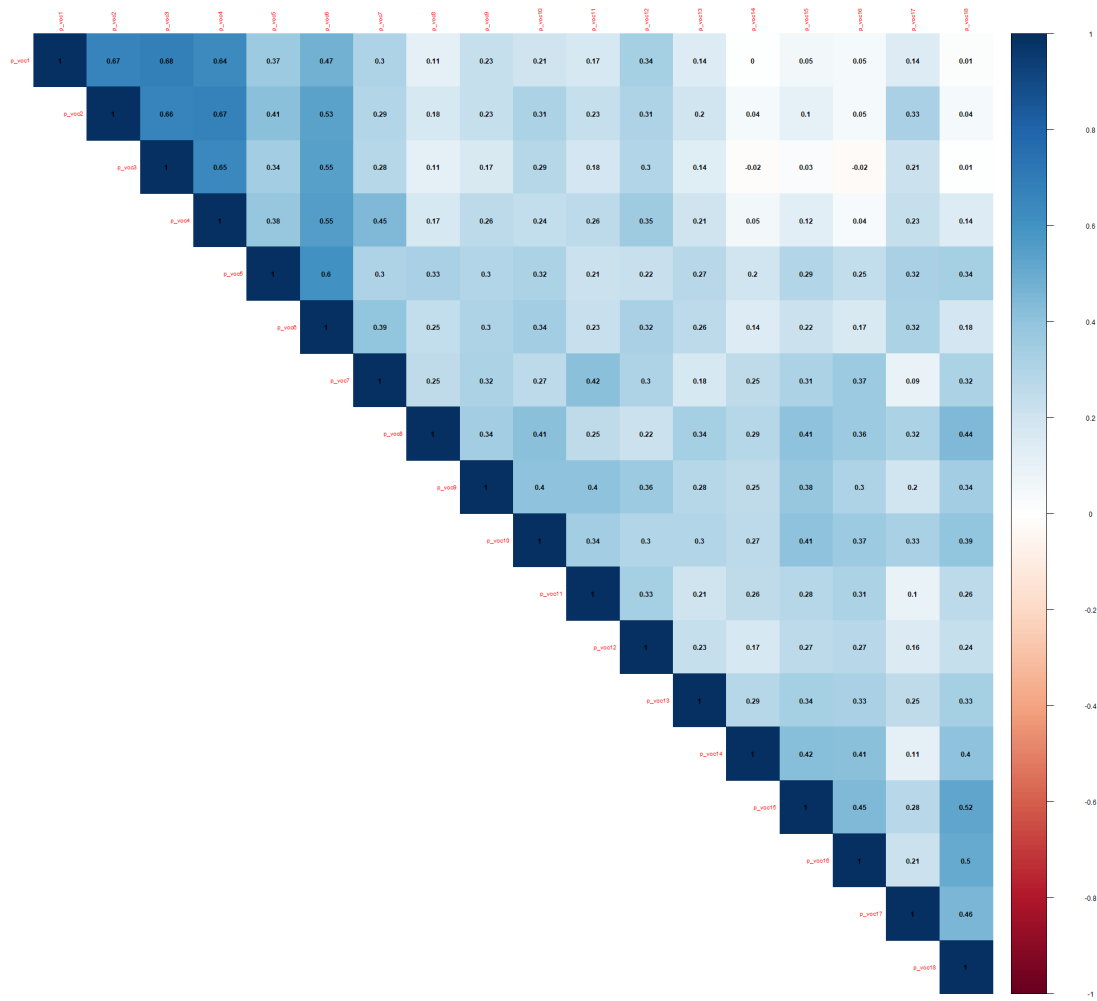


Figura A.30: Correlações das variáveis do teste de Vocabulário

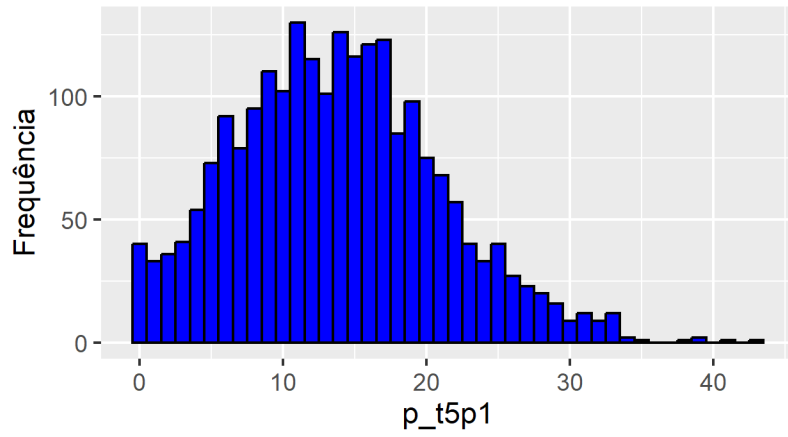


Figura A.31: Histograma da variável p_t5p1 do teste Fluência no desenho

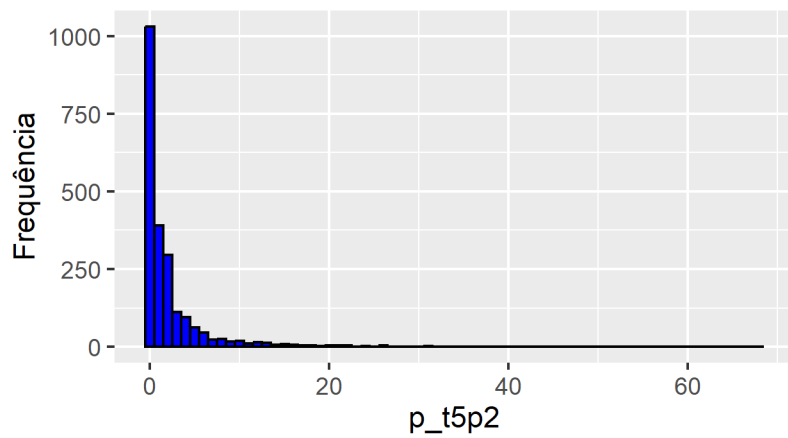


Figura A.32: Histograma Fluência desenho p_t5p2

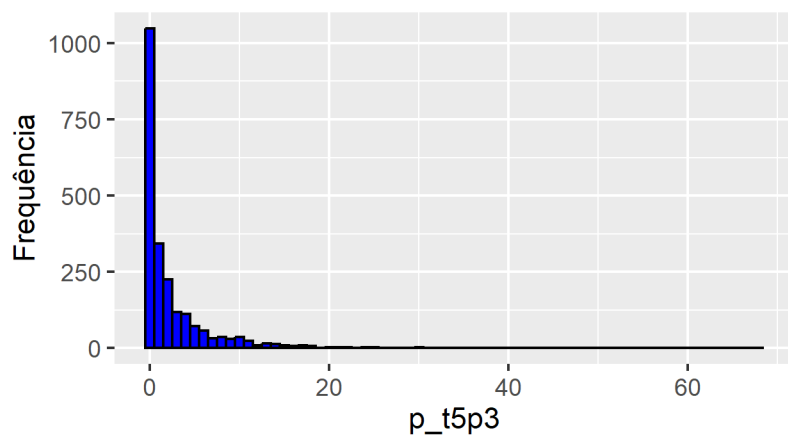


Figura A.33: Histograma Fluência desenho p_t5p3

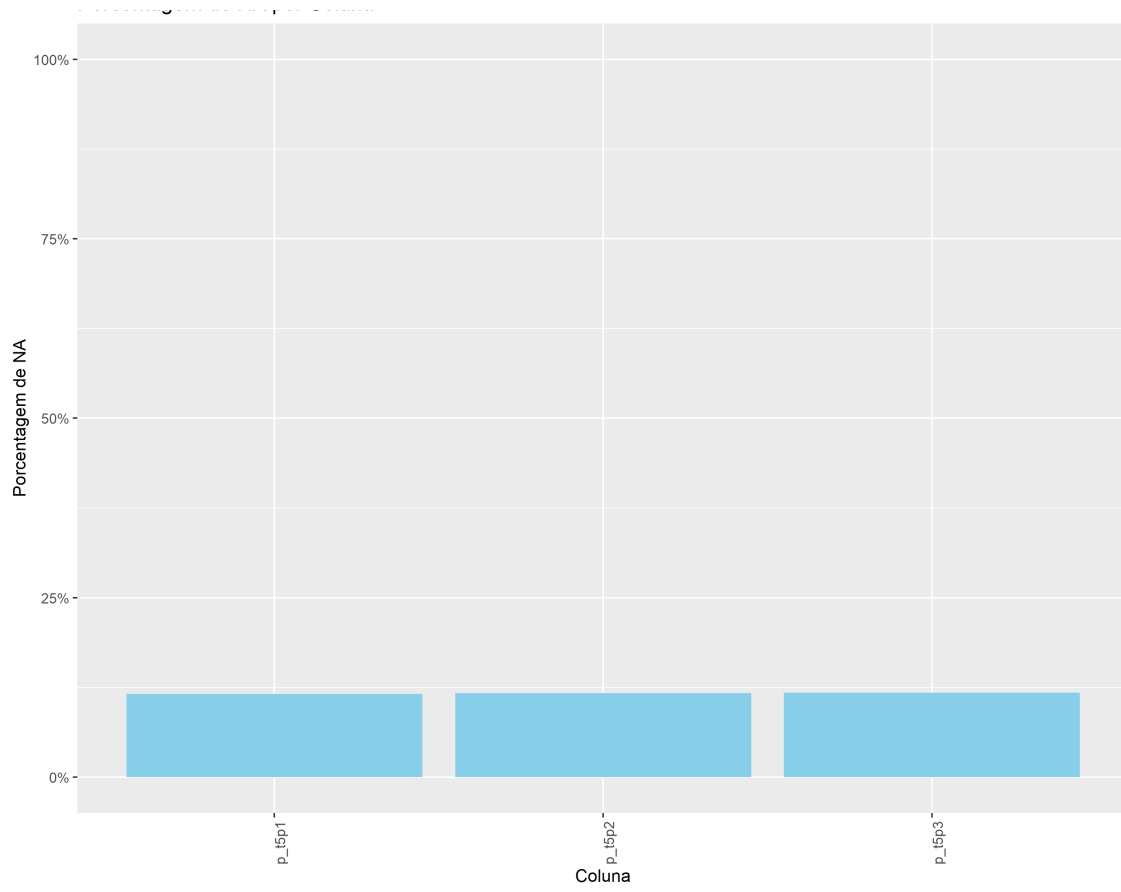


Figura A.34: Porcentagem de NA por Coluna Fluência desenho

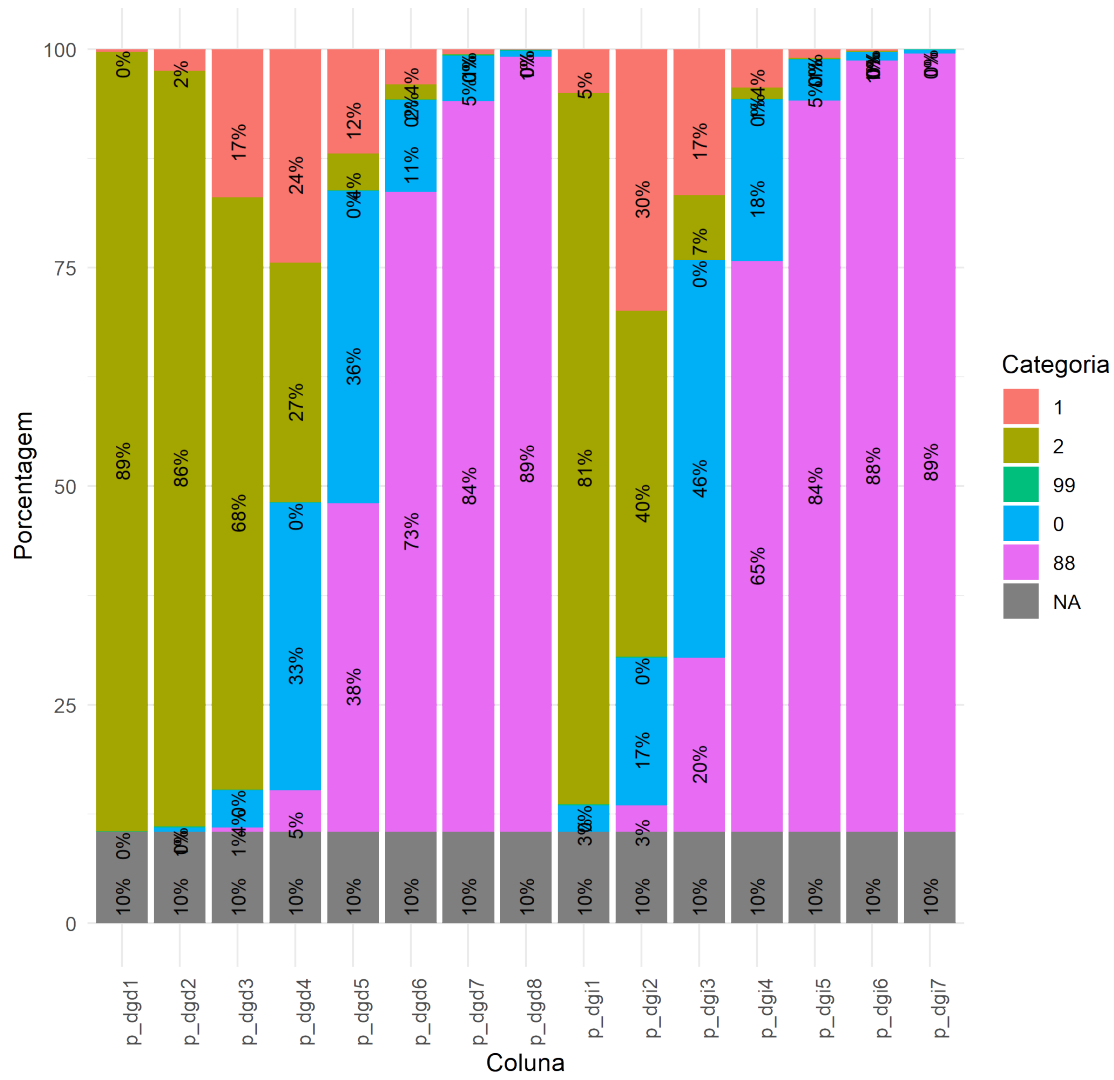


Figura A.35: Percentual de categorias no teste Dígitos – Memória Operacional

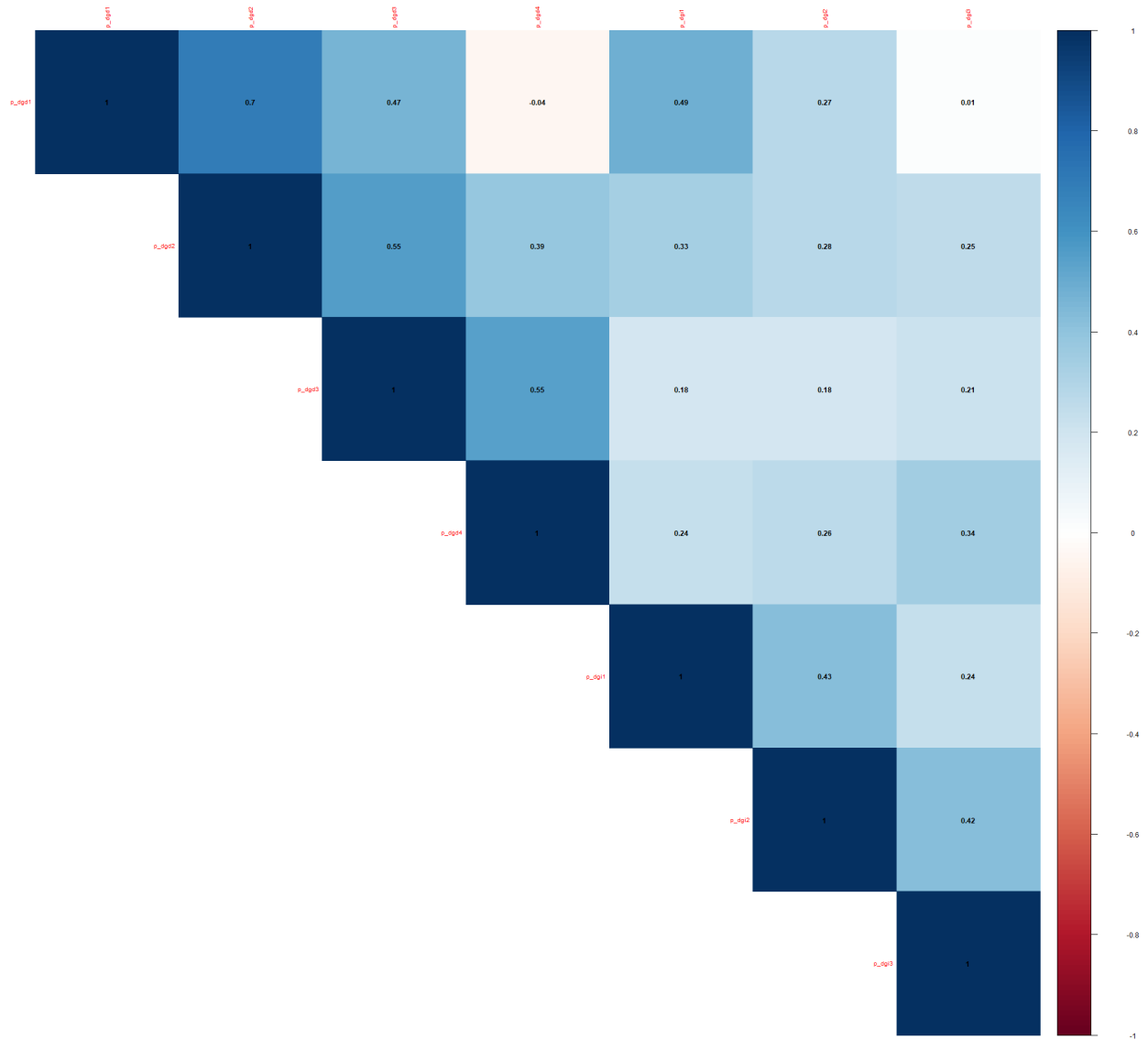


Figura A.36: Correlações das variáveis do teste Dígitos – Memória Operacional

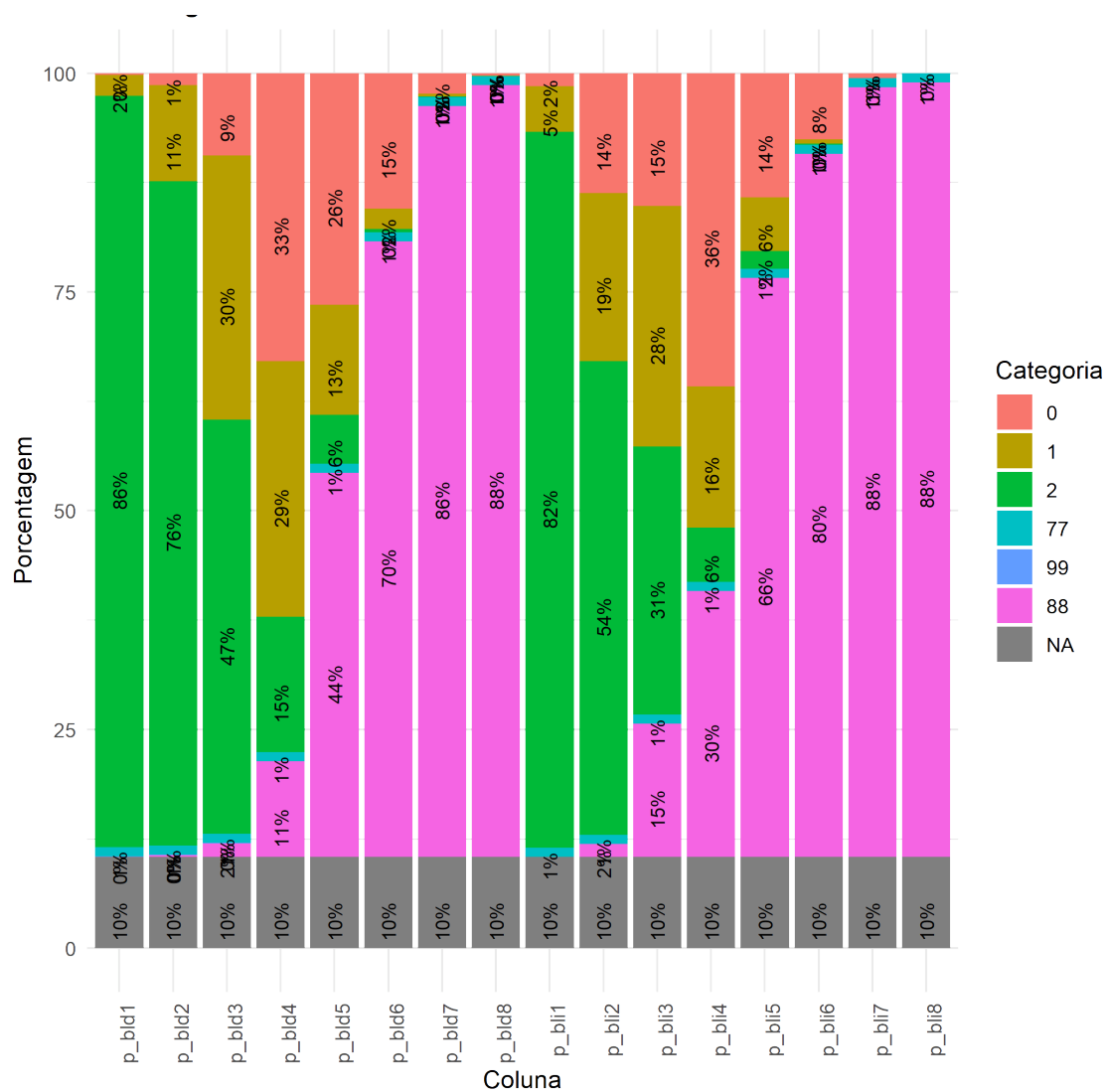


Figura A.37: Percentual de categorias no teste blocos de Corsi

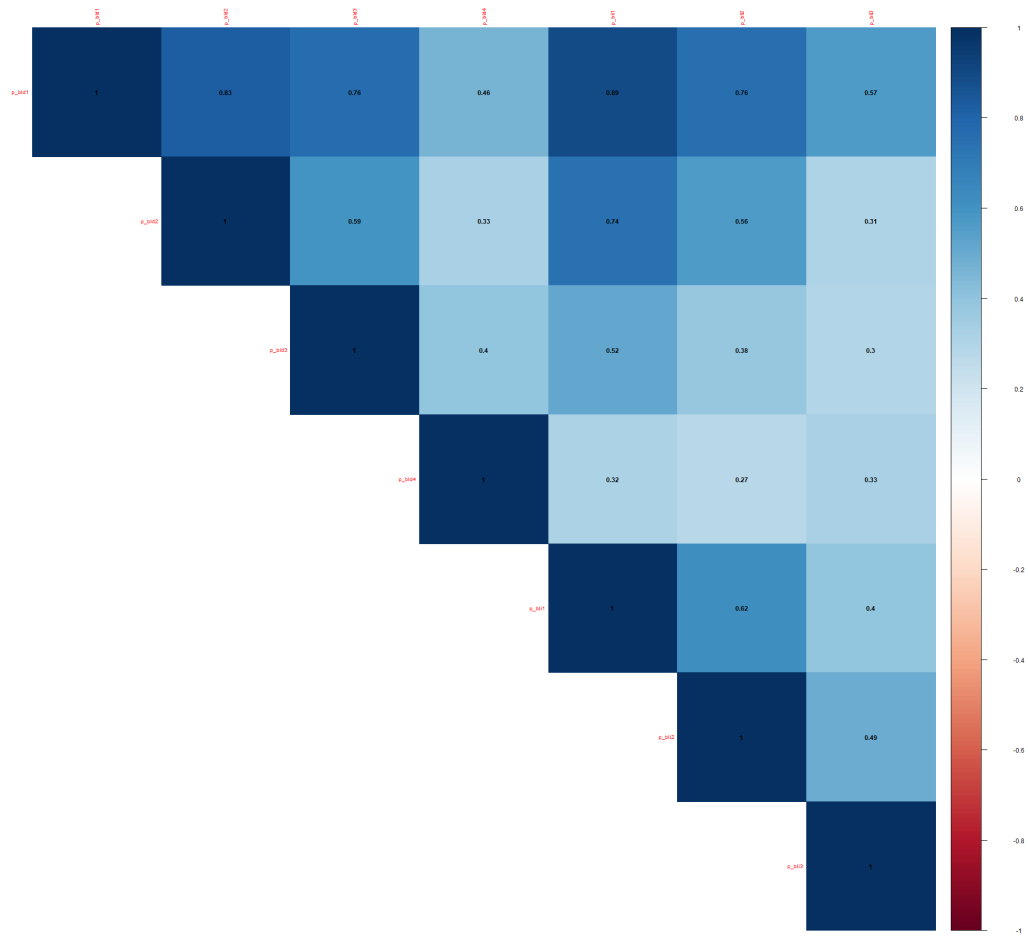


Figura A.38: Correlações das variáveis do teste blocos de Corsi

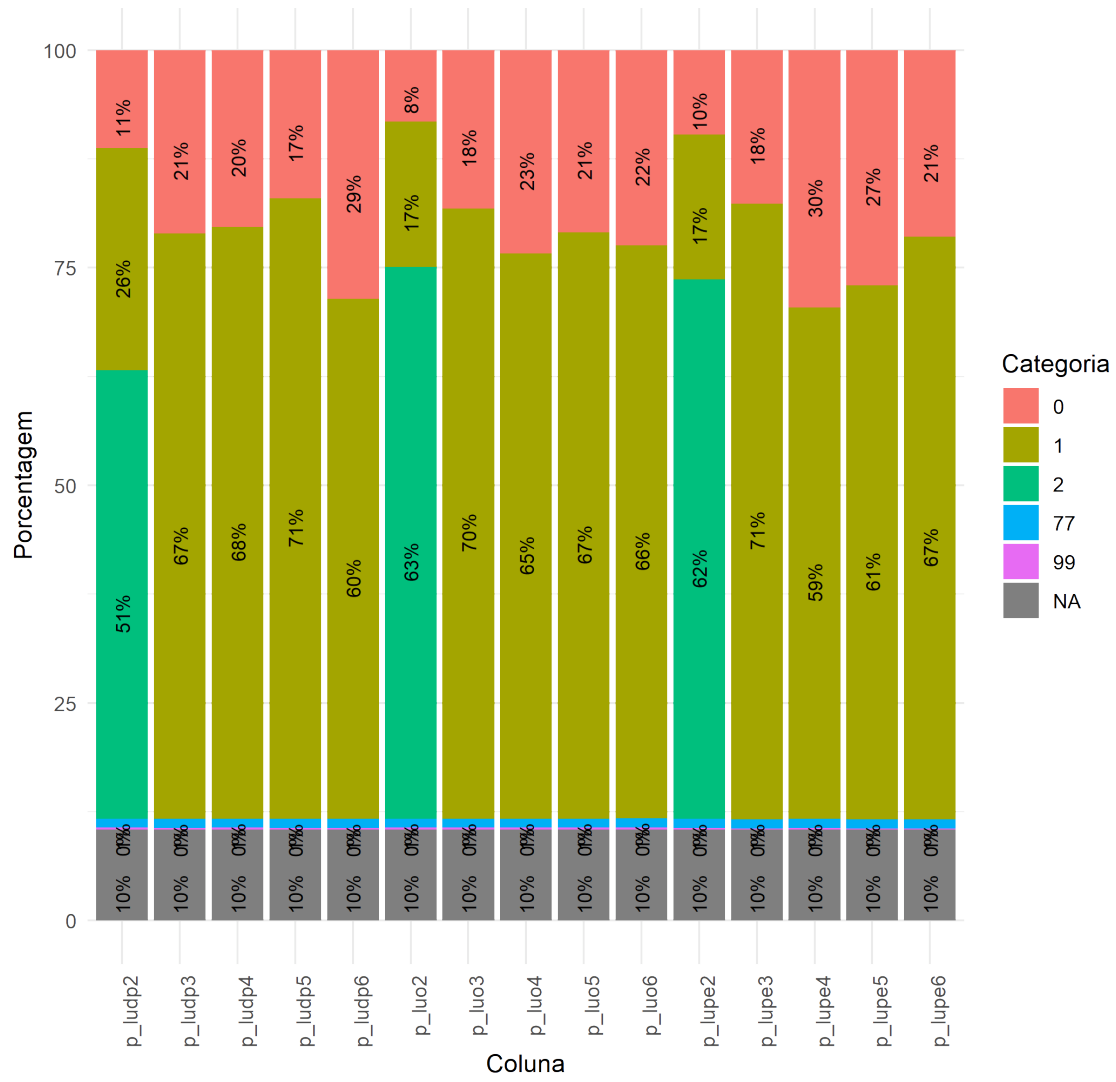


Figura A.39: Percentual de categorias no teste Luria

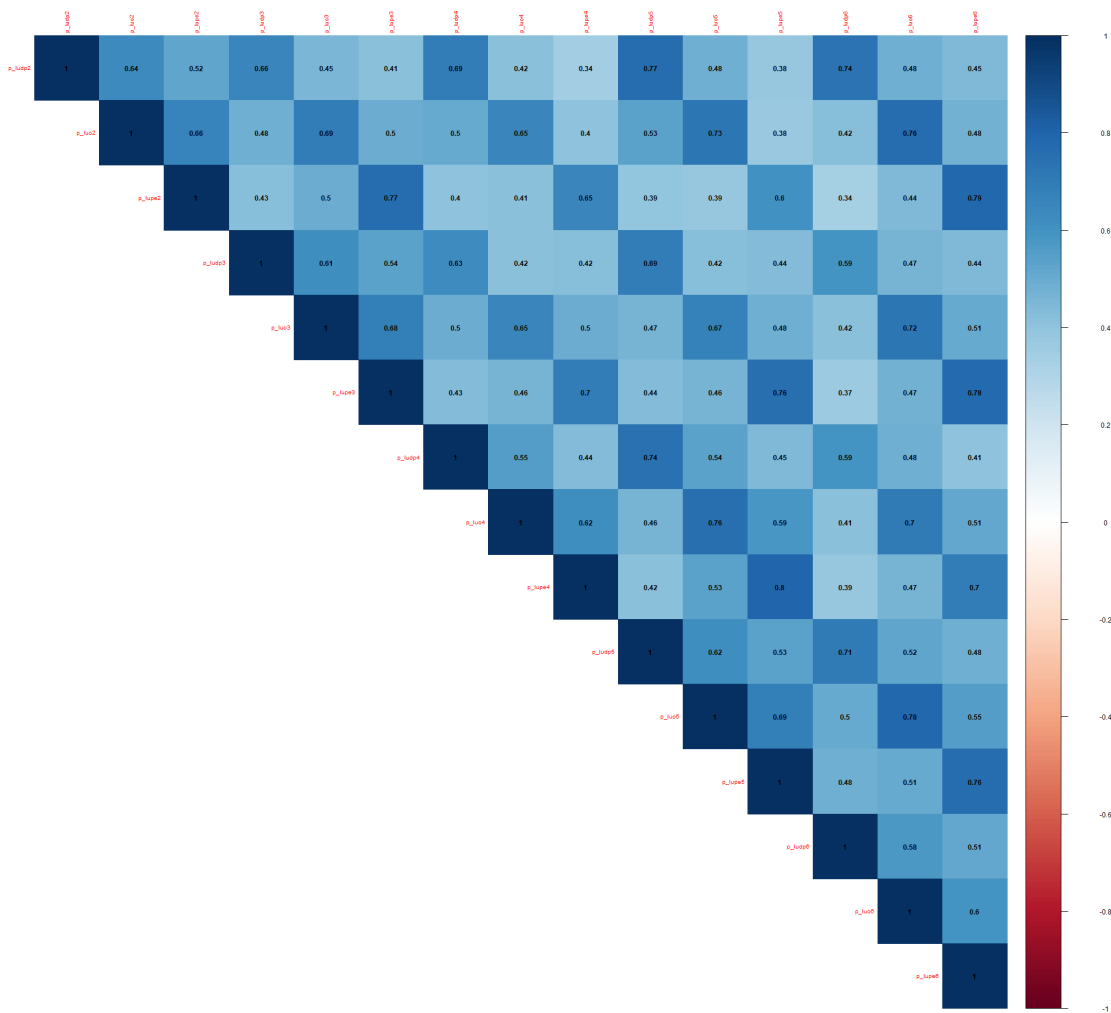


Figura A.40: Correlações das variáveis do teste de Luria

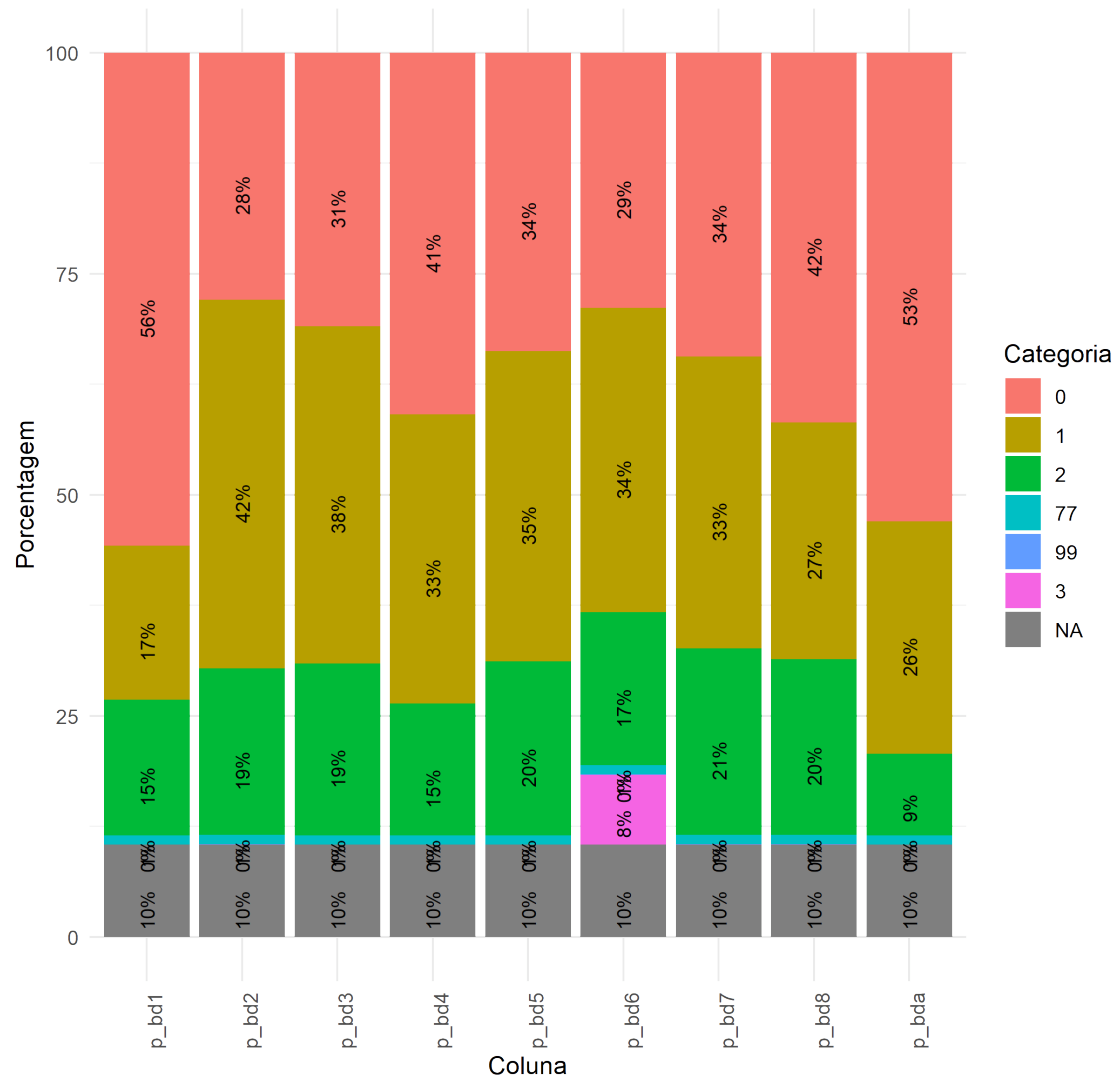


Figura A.41: Percentual de categorias no teste Bender

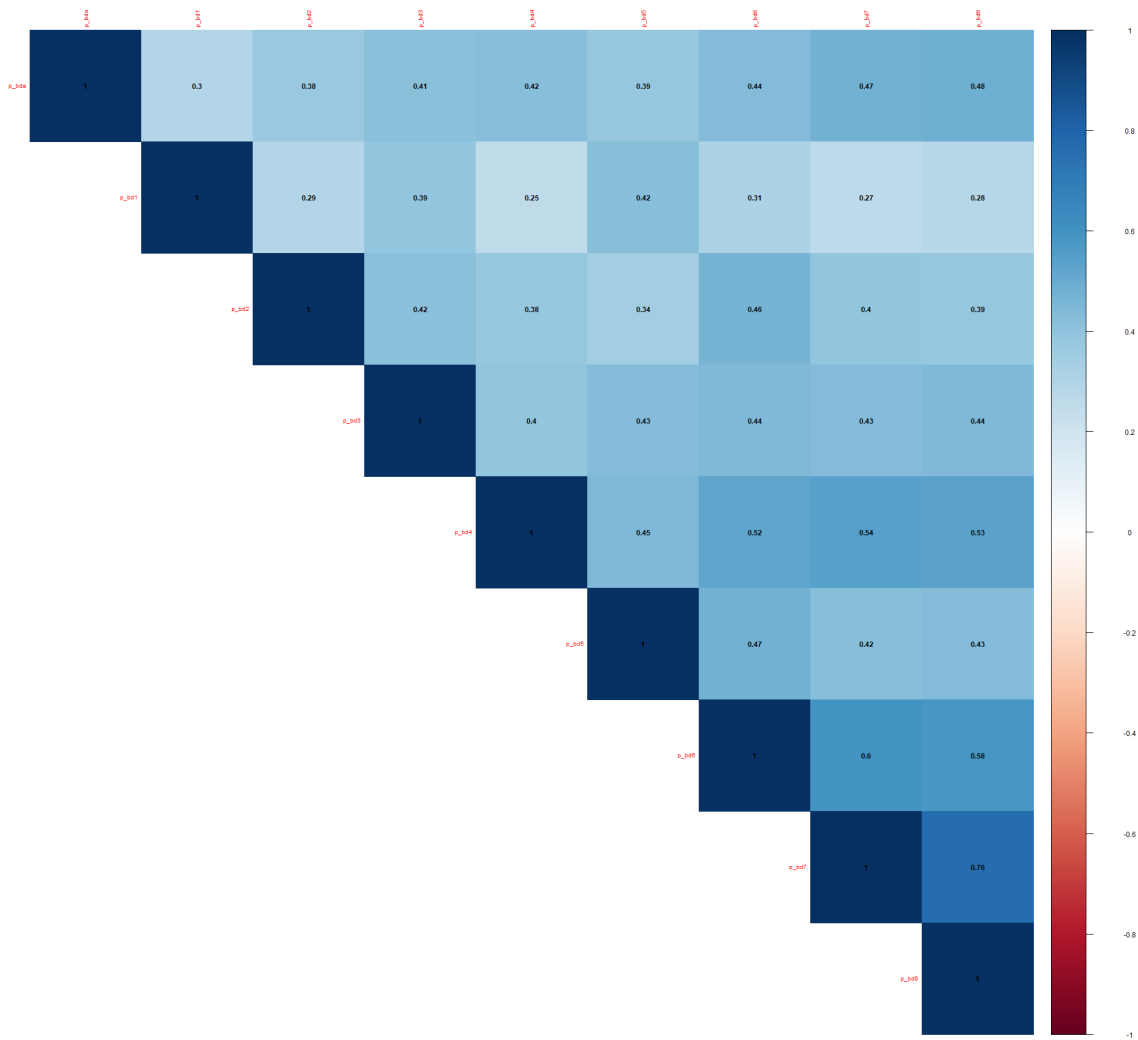


Figura A.42: Correlações das variáveis do teste Bender

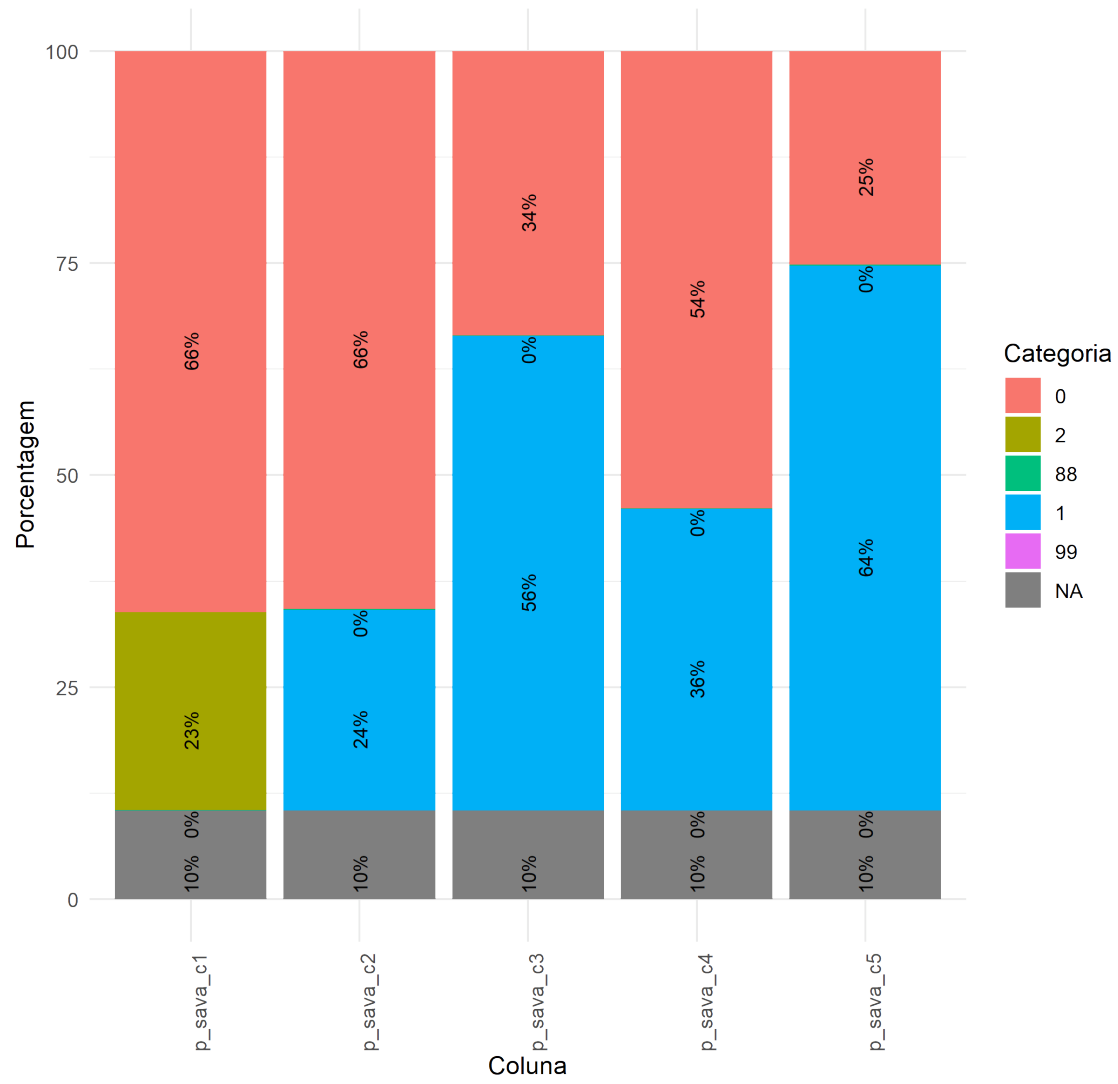


Figura A.43: Percentual de categorias no teste figura de Rey

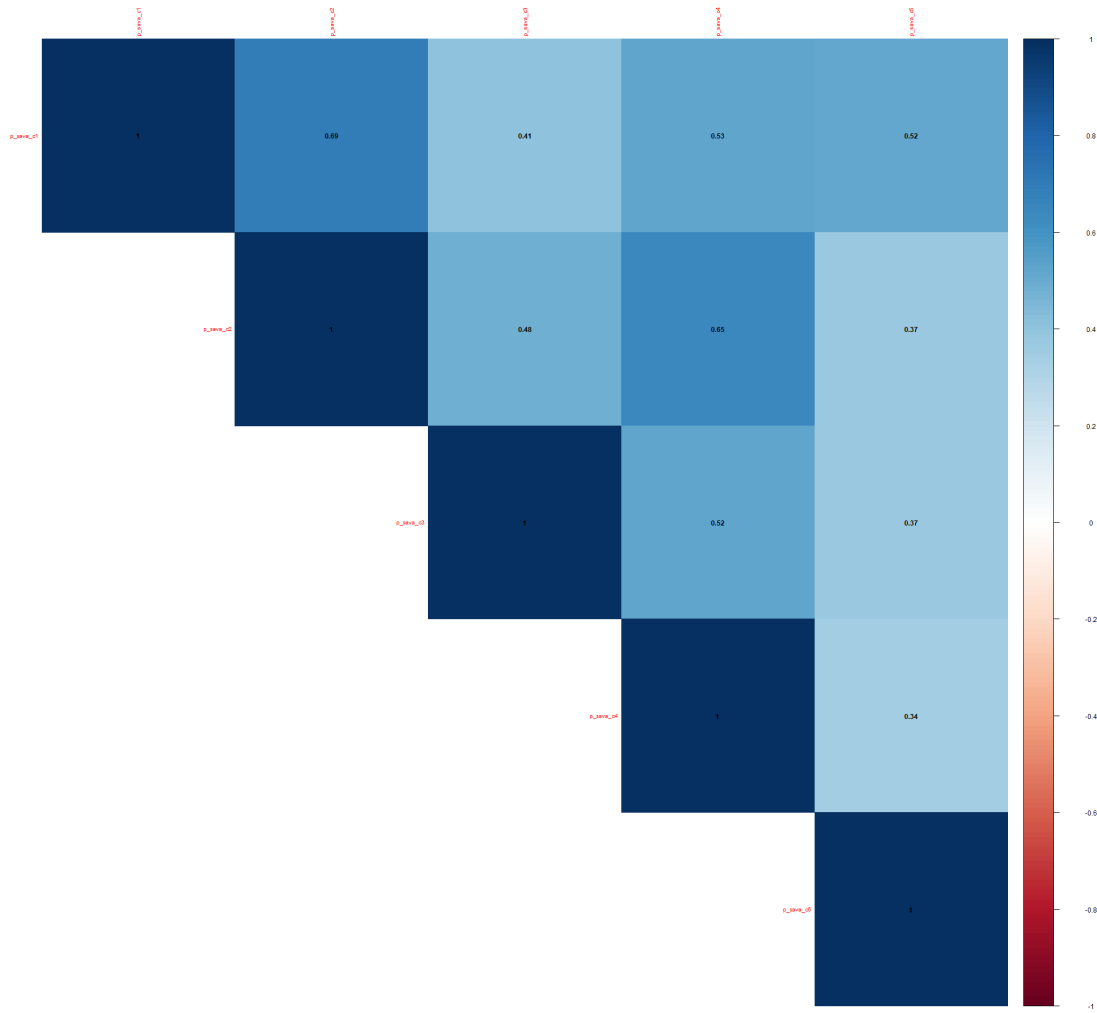


Figura A.44: Correlações das variáveis do teste figura de Rey

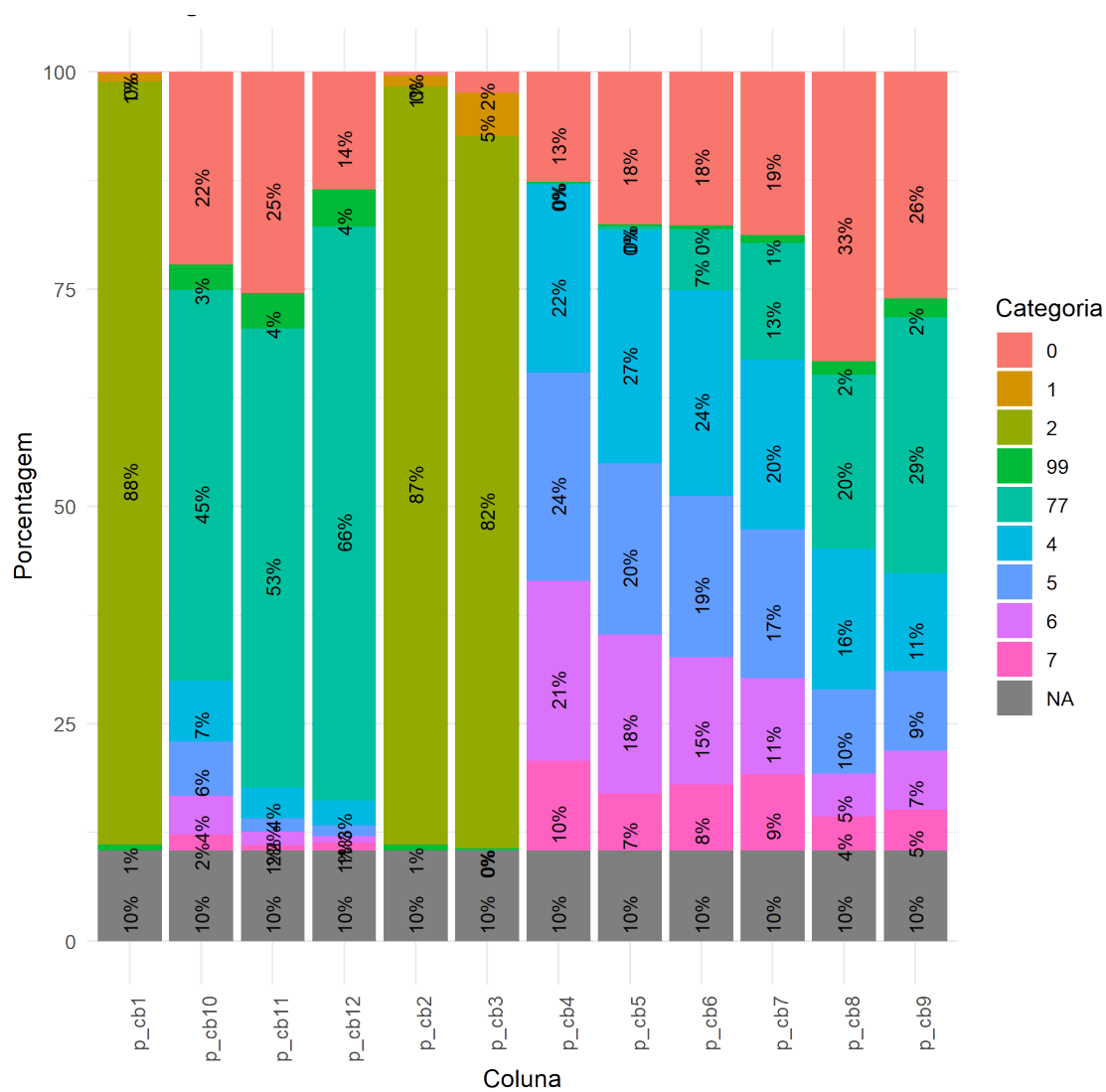


Figura A.45: Percentual de categorias no teste Cubos

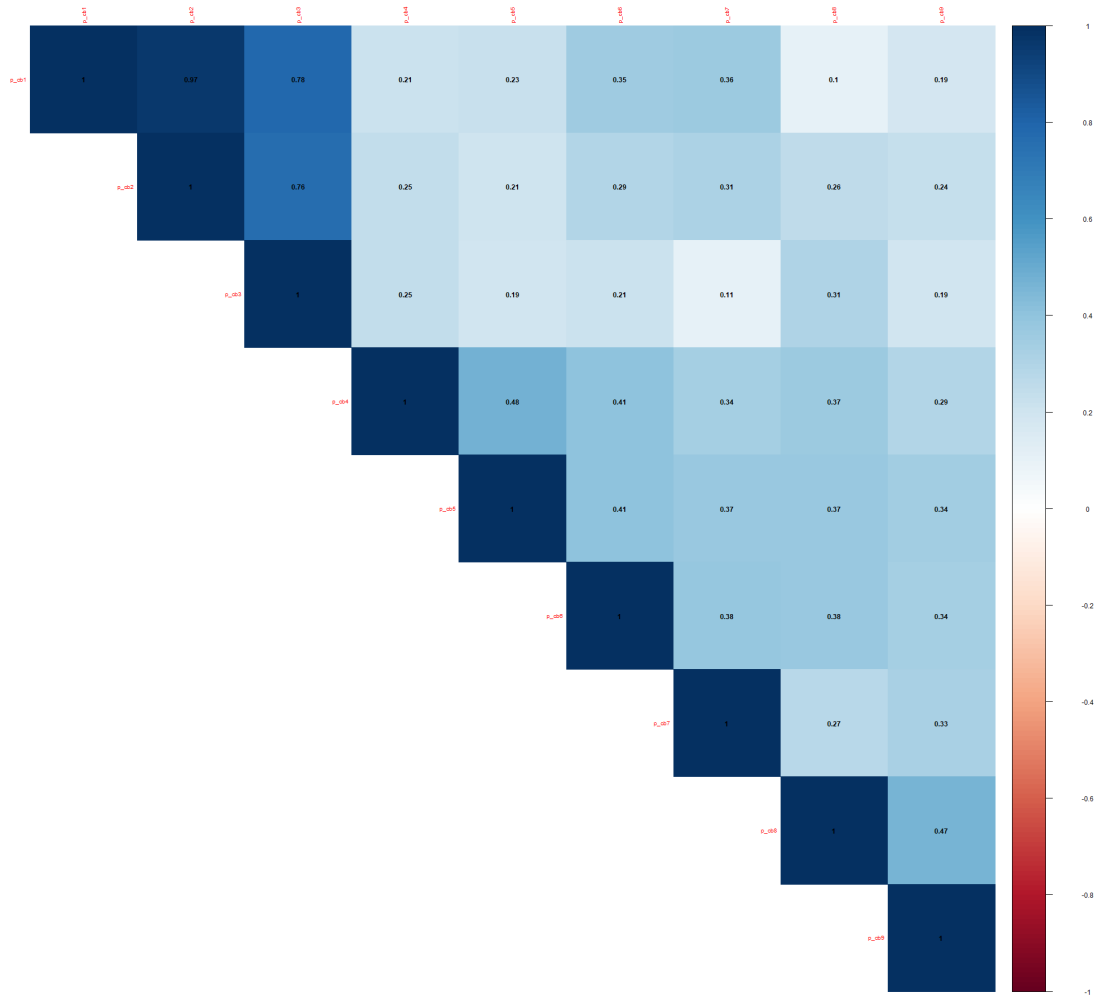


Figura A.46: Correlações das variáveis do teste Cubos

Figura A.47: Diagrama fator geral fornecido pelos pesquisadores

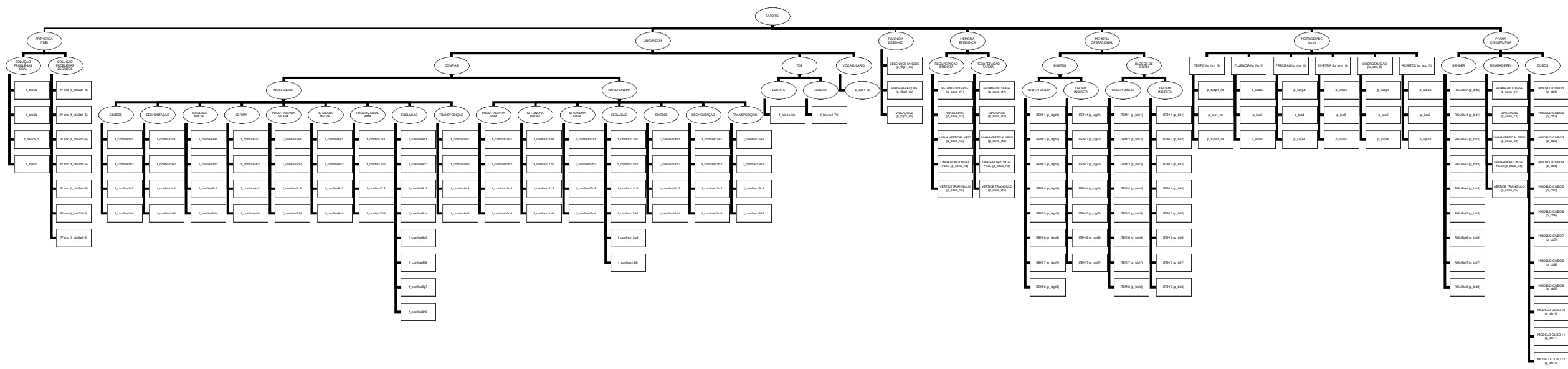
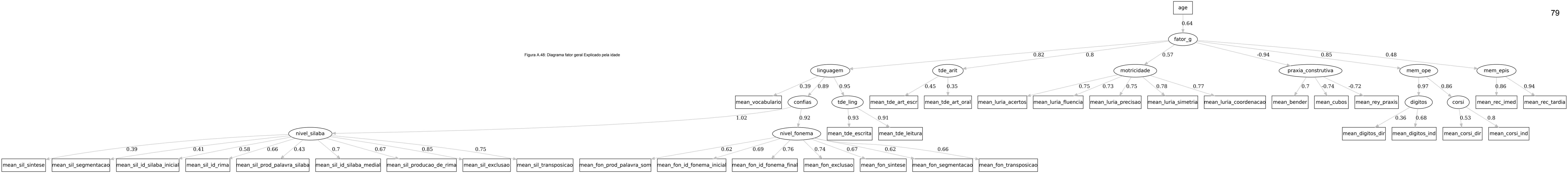


Figura A.48: Diagrama fator geral Explicado pela idade



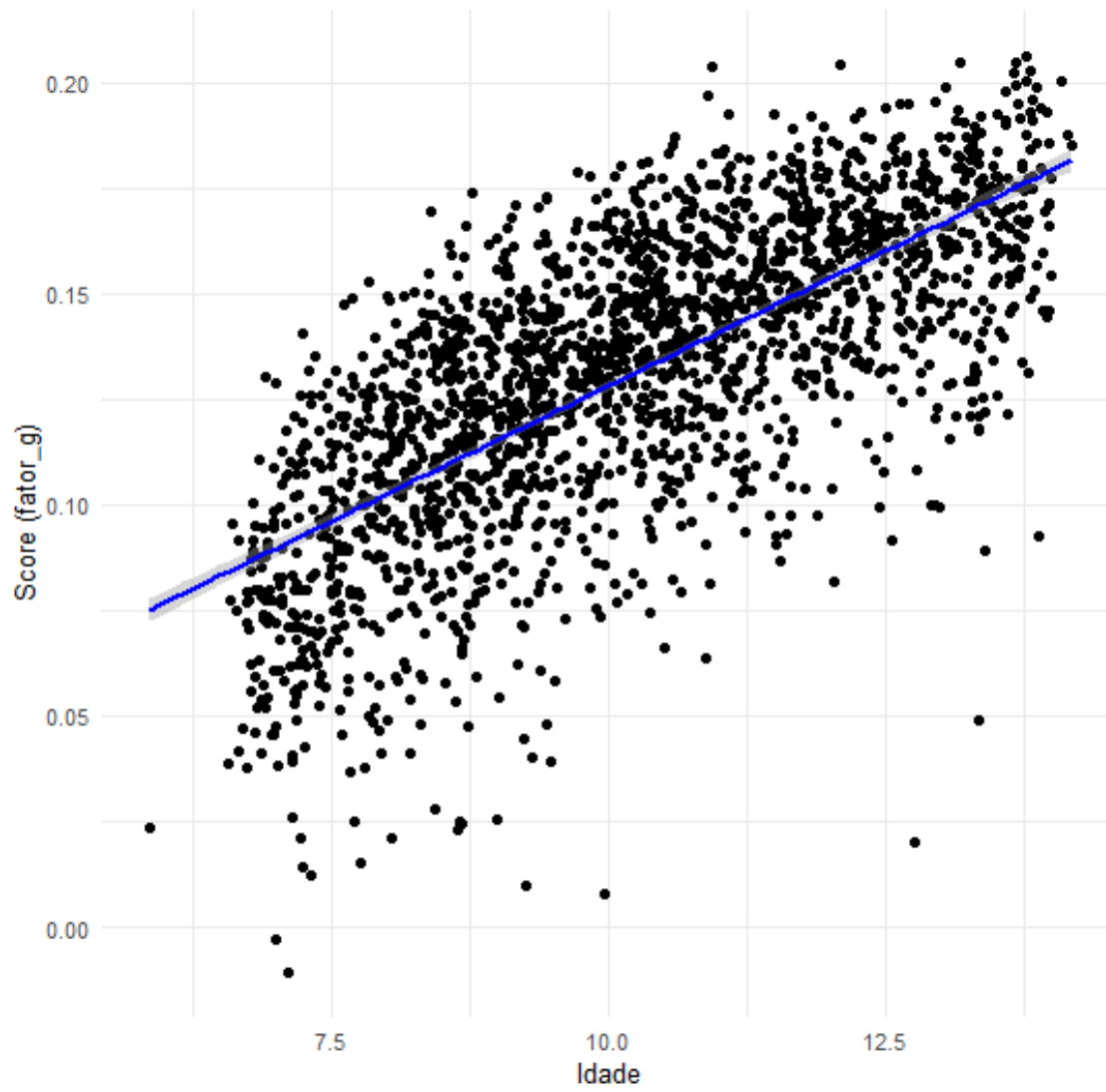
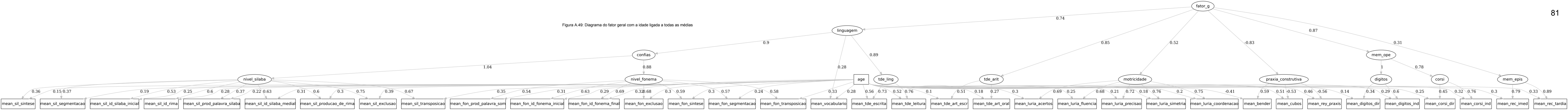


Figura A.49 Gráfico de dispersão Fator G e Idade

Figura A.49: Diagrama do fator geral com a idade ligada a todas as médias



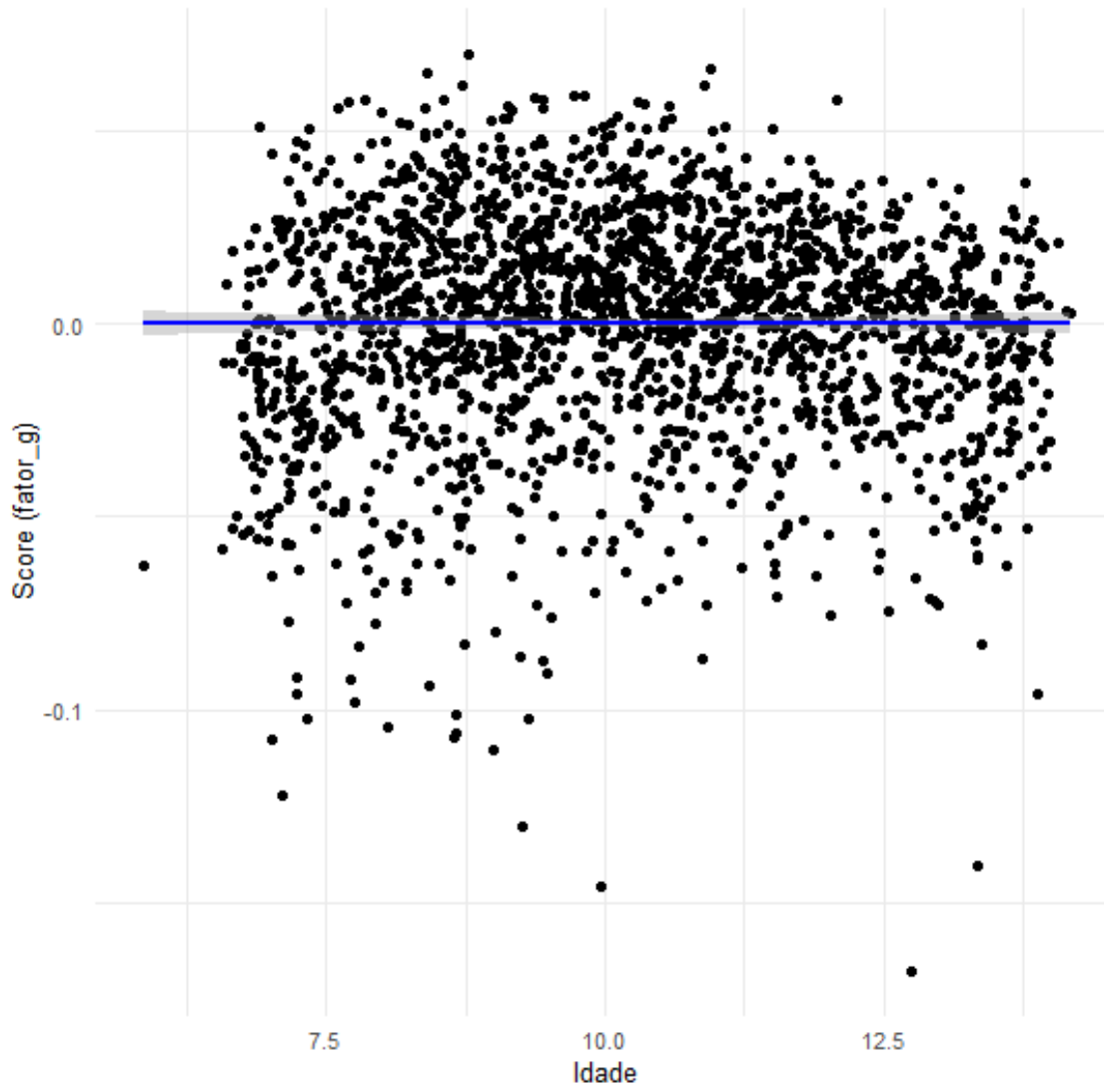


Figura A.51 Gráfico de dispersão Fator G e Idade