



Relatório de Consultoria

Áreas do Conhecimento:

Estatística, Modelos Lineares,
Análise Fatorial

Palavras-Chave:

Análise de dados, Jovens
Construtores, Social

Analizando dados do SAEB

Gabriel Mizuno^{a,b}, Larissa Estudante^{a,b},
Lyncoln Sousa^{a,b}, Marina Sudré^{a,b},
Rodolfo Hauret^{a,b}, Jony Arrais^{a,c}, Roberto
Sampaio^d

^aDepartamento de Estatística - UFF

^bConsultor

^cOrientação

^dConsulente

Este trabalho busca encontrar alguma característica não observável, em um conjunto de perguntas na escala Likert realizadas à indivíduos que participaram de um tratamento nas comunidades da Mangueira e do Borel.

Sumário

1	Introdução	4
2	Matérias e Métodos	4
3	Resultados	6
4	Conclusão	12
	Referência	13
5	Apêndice	14

Lista de Figuras

1	Gráfico de barras para Tratamento, Homem, e Mangueira respectivamente	6
2	Gráfico de barras para Branco, Mora com os pais, e Filhos respectivamente	7
3	Gráfico de barras da escolaridade.	7
4	Escala likert para perguntas Q1 a Q6	8
5	Escala likert para perguntas Q7 a Q12	8
6	Escala likert para perguntas Q13 a Q17	9
7	Diagrama dos fatores	10

Lista de Tabelas

1	Valores de adequação de KMO	5
2	Valores da medida de KMO	9
3	Exemplo de escore	10
4	Estimativas do modelo 1	11
5	Estimativas do modelo 2	11
6	Estimativas do modelo 3	12
7	Variáveis sociodemográficas	14
8	Perguntas do questionário	14

1 Introdução

Recentemente, o projeto social Jovens Construtores foi implementado em duas comunidades, Borel e Mangueira. Nesse projeto, alguns jovens receberam o treinamento (tratamento) e outros só participaram de encontros presenciais (controle). No final do treinamento, um questionário foi aplicado a todos os indivíduos do projeto, e as respostas de interesse foram medidas em Escala Likert, sobre 3 grupos de assuntos diversos. Também constam no questionário cinco variáveis binárias: se mora na mangueira, se é branco, se é homem, se mora com os pais e se tem filhos, onde 0 representa Não e 1 representa Sim e a variável escolaridade, cujas respostas são classificadas como 0 para indivíduo com ensino Fundamental completo, 1 para ensino Médio completo e 2 para ensino Superior completo.

Primeiramente, o pesquisador Roberto Sampaio solicitou uma análise com objetivo de verificar se o conjunto de variáveis que estão em Escala Likert medem alguma característica ou informação que não estava sendo observada diretamente. Para isto, foi realizada uma análise fatorial visando agrupar essas variáveis de modo que fossem formados conjuntos de variáveis altamente correlacionadas, conhecidos como fatores. Os fatores tem o objetivo resumir as diversas variáveis em um conjunto menor de dimensões com uma perda mínima de informação, podendo representar um determinado conceito ou significado que não seria possível verificar individualmente.

Após a criação dos fatores foi gerado um escore para cada um deles, isto é, uma pontuação para cada indivíduo observado, proporcional as respostas dadas para cada pergunta. No segundo momento, foi solicitada outra análise, com o objetivo de verificar se as variáveis sociodemográficas explicam esses escores. Para isso foram ajustados modelos estatísticos para quais as variáveis respostas foram os escores e as variáveis explicativas foram as variáveis sociodemográficas.

Este relatório está construído seguindo a seguinte estrutura, na seção de materiais e métodos são descritos de forma mais aprofundada as técnicas utilizadas para o tratamento dos dados e as ferramentas utilizadas para construção dos gráficos e dos modelos. Na seção de resultados serão relatados os resultados obtidos durante os procedimentos realizados na seção de materiais e métodos. Na seção de conclusão será discutido os resultados obtidos neste estudo.

2 Matérias e Métodos

A base de dados do Projeto Jovens Construtores apresenta 307 indivíduos e 25 variáveis. As variáveis apresentam informações sobre Tratamento, Local, se mora na Mangueira, se é branco, se é homem, se mora com os pais e se tem filhos. Além dessas variáveis, também consta na base de dados as respostas, em escala likert, das perguntas feitas aos indivíduos nos questionários.

A escala Likert é um tipo de escala para questionários, amplamente usadas para medir posturas e opiniões com um nível maior de nuance. Suas perguntas usam uma escala de 5 ou 7 pontos, muitas vezes chamada de escala de satisfação, que varia de uma atitude extrema a outra. Normalmente, a pergunta de pesquisa da escala Likert de satisfação inclui uma opção moderada ou neutra.

A questão 1 mede se o indivíduo está satisfeito consigo mesmo, a questão dois avalia se o indivíduo se sente tão apto a fazer coisas como a maioria das pessoas e as quinze questões restantes se encontram na 8 nos Anexos. Ainda na base de dados analisada, a escala Likert das perguntas de 1 a 6 variam de 1 a 5, onde a pontuação mais alta significa discordo totalmente e a mais baixa significa concordo totalmente. As questões de 7 a 12 também variam de 1 a 5, onde cinco significa muito satisfeito e 1 pouco satisfeito. Por fim, as questões de 13 a 17 variam de 1 a 7, em que a pontuação mais baixa significa discordo totalmente e a mais alta concordo totalmente.

Para realizar uma análise descritiva foram construídos gráficos de barras através do pacote *ggplot2* e tabelas de frequências, além de gráficos de likert através do pacote *likert* [8].

Para verificar se o conjunto de variáveis que estão em Escala Likert medem alguma característica ou informação que não estava sendo observada diretamente, foi realizada uma análise fatorial. Análise Fatorial é um método estatístico que tem como objetivo explicar a variabilidade de variáveis observáveis, como nota de uma prova, respostas de um questionário e avaliação de determinado produto, com variáveis não observáveis diretamente (também chamadas de latentes). Inteligência, personalidade, motivação, nível socioeconômico, democracia e vulnerabilidade social são alguns exemplos de variáveis latentes. Espera-se que tais variáveis observáveis tenham correlação entre elas. As variáveis observadas estão ligadas com as variáveis latentes através do seguinte modelo

$$\begin{aligned}
Z_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1 \\
Z_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2 \\
&\vdots \\
Z_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p
\end{aligned}$$

Onde Z_i representa as variáveis observáveis, F_i as latentes, l_i são chamados de *loadings* e ϵ_i são os erros do modelo. Para mais detalhes ver em [4]

Para a realização deste método, a escala das perguntas 1 a 6 foi invertida, de forma que seguisse o mesmo padrão das demais perguntas. Tal método necessita de uma matriz de correlação entre as variáveis, diante da natureza das variáveis do questionário, utilizou-se a matriz de Correlação Policórica, pois as variáveis são ordinais com três ou mais categorias [1].

Em seguida, para verificar a adequabilidade da amostra, foi calculada a medida de Kaiser-Meyer-Olkin (doravante KMO) [7]. Essa medida, que varia entre 0 e 1, representa a proporção da variância das variáveis que pode ser explicada pelos fatores. Como mostrado na Tabela ??, quanto mais próximo esse valor estiver de 1, mais adequados os dados estão para se ajustar uma Análise Fatorial. Por fim, foi feito o teste de *Bartlett* para verificar se existia correlação entre as perguntas do questionário, para mais detalhes ver em [4].

Tabela 1: Valores de adequação de KMO

KMO	Adequabilidade da amostra
<0,5	Inaceitável
[0,5-0,7]	Medíocre
[0,7-0,8]	Bom
[0,8-0,9]	Ótimo
[0,9>]	Excelente

Verificados os pressupostos, foi utilizado o critério de *Kaiser* para encontrar o número ideal de fatores. Este critério diz que o número de fatores a serem construídos é igual ao número de autovetores maiores que 1 na matriz de correlação[4]. Através desse método definiu-se cinco como sendo o número ideal de fatores.

Para construir os fatores, foi utilizada a função *fa* do pacote *psych* e [6], como argumentos, foram passados o número de fatores definidos pelo critério de *Kaiser*, o tipo de correlação (policórica), o método *Generalized least squares* e a rotação do tipo *varimax*, que, após verificação, constatou-se que seria a melhor rotação para o problema. Finalmente, foi feita a associação das perguntas ao fator que tem maior correlação.

Para avaliar a relação entres os fatores obtidos e as variáveis sociodemográfica, foram construídos três modelos de Regressão Linear Múltipla, um para cada fator, em que a variável resposta são os escores dos fatores. Para aplicar este método foi necessário realizar uma transformação nas variáveis de numérica para categóricas.

Definidos os fatores a etapa final é estimar o escore individual destes para cada indivíduo. Os escores são uma espécie de média ponderada das variáveis observadas em cada uma das unidades amostrais, onde os pesos são dados justamente pelas cargas fatoriais. Para esta análise os escores foram obtidos pelo método de Thurstone (Regressão simples).

O modelo de regressão linear múltipla é definido da seguinte maneira. Seja X a matriz de observações referentes às variáveis explicativas. Considere Y o vetor com os escores. De acordo com 2.1, podemos expressar a dependência entre Y e X através da equação

$$Y = X\beta + \epsilon \quad (2.1)$$

Onde $\beta = [\beta_0, \beta_1, \dots, \beta_{p-1}]$ são os coeficientes estimado por 2.2 e ϵ um vetor que segue algumas hipóteses que serão verificadas mais a frente (para mais detalhes ver em Neter et al.).

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.2)$$

Ajustado o modelo, é necessário determinar se os coeficientes são estatisticamente significantes, isto é, verificar se há uma relação linear entre Y e X_i . Para realizar esta verificação é utilizado o teste t . Sob a hipótese nula, assume-se que o estimador possui distribuição t com $n - p$ graus de liberdade.

$$\begin{cases} H_0 : \beta_i = 0, \\ H_1 : \beta_i \neq 0 \end{cases}$$

A validação de um modelo linear é dada quando ele cumpre as seguintes hipóteses, independência dos ϵ , homocedasticidade, e normalidade dos resíduos. A primeira hipótese é verificada através de um gráfico de dispersão entre os resíduos e a variável resposta estimada (\hat{y}). A segunda hipótese é verificada através do Teste de White (para mais detalhes ver em Gujarati and Dawn), utilizando-se a função `white_lm` do pacote [2]. Para verificar a normalidade dos resíduos é utilizado o Teste de Shapiro-Wilk. Nas hipóteses que foram aplicados testes de verificação optou-se por utilizar um nível de significância (α) de 5%.

Após o ajuste do modelo é necessário verificar a ausência do efeito multicolinearidade. Isto ocorre quando duas variáveis possuem correlação entre elas, tal efeito prejudica as estimativas dos parâmetros modelo e suas variâncias, para verificar tal problema utiliza-se uma medida chamada VIF que verifica o quanto a variância de cada parâmetro está "inflacionada", caso obtém-se uma variável com VIF maior ou igual do que 10 opta-se por excluir tal variável da análise.

3 Resultados

Os resultados obtidos serão apresentados e explicados com a seguinte configuração, primeiro serão apresentadas algumas análises exploratória dos dados e, em seguida, serão apresentados os resultados da análise fatorial, associação das perguntas ao fator que tem maior correlação e dos modelos de regressão.

Na Figura 1, pode-se observar que não existe uma discrepância entre a frequência de indivíduos por Tratamento, 54,6% receberam o tratamento e 45,4% só participaram de encontros presenciais. Em relação ao sexo, também é observado um equilíbrio, pois 51% dos indivíduos analisados são homens e 49% mulheres. Observando a localidade, nota-se que 156 são da Mangueira 148 do Borel.

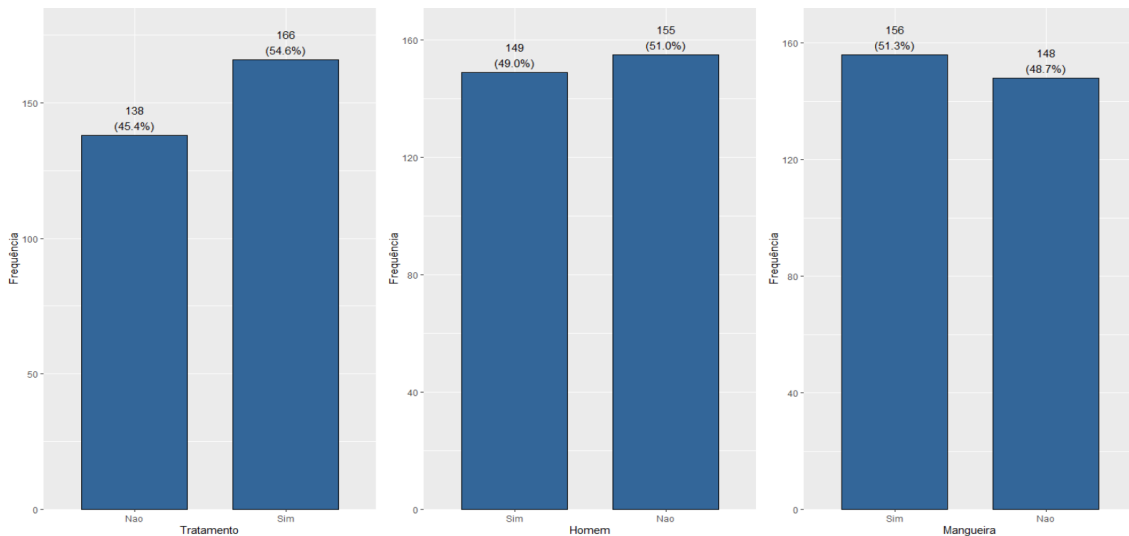


Figura 1: Gráfico de barras para Tratamento, Homem, e Mangueira respectivamente

Ao contrário do que foi observado acima, a Figura 2 abaixo mostra uma discrepância dentro das variáveis raça, mora com os pais e se tem filhos. No primeiro gráfico de barras da Figura 2, nota-se que existe uma grande frequência de observações para não-brancos, um total de 278 das 304 observações. Em relação a moradia, percebe-se que a maioria dos indivíduos analisados moram com os pais (70,4%), enquanto somente 29,6% não. Para a variável filhos, nota-se que existe uma maior concentração de pessoas sem filhos entre os indivíduos.

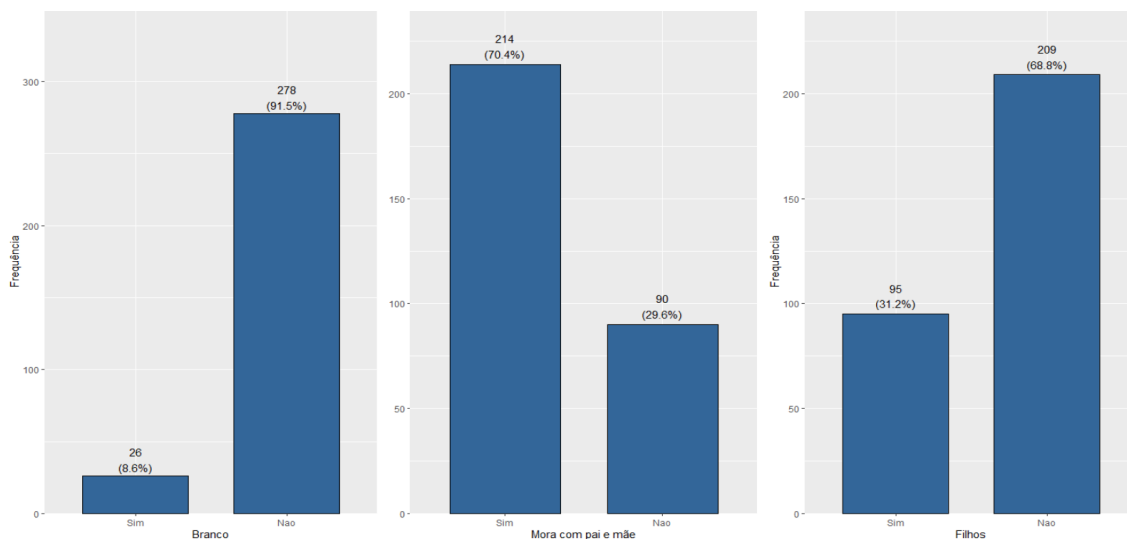


Figura 2: Gráfico de barras para Branco, Mora com os pais, e Filhos respectivamente

Por fim, analisando o gráfico da Figura 3 observa-se que existe uma maior concentração de indivíduos que completaram o ensino médio, seguido por quem completou o fundamental e por fim o superior completo.

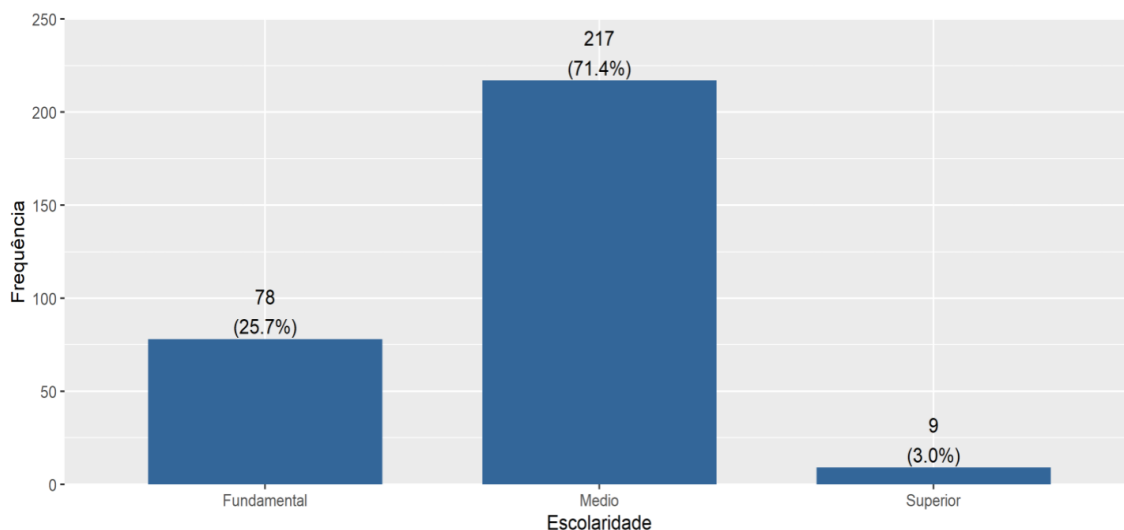


Figura 3: Gráfico de barras da escolaridade.

Para observar qual foi o comportamento das respostas dos indivíduos para as perguntas, foram utilizados gráficos de barras apropriados para a escala likert, onde é possível observar a frequência relativa de cada uma das respostas. Também é possível observar a porcentagem de respostas positivas, neutras, e negativas. Confira a tabela 8 presente no apêndice para ter um dicionário das perguntas por variáveis Q1 a Q17.

Observando o gráfico da figura 4 referente as perguntas Q1 até a Q6, nota-se que para as perguntas Q1, Q2 e Q5 houve um número maior de discordância em relação as perguntas Q3, Q4 e Q6.

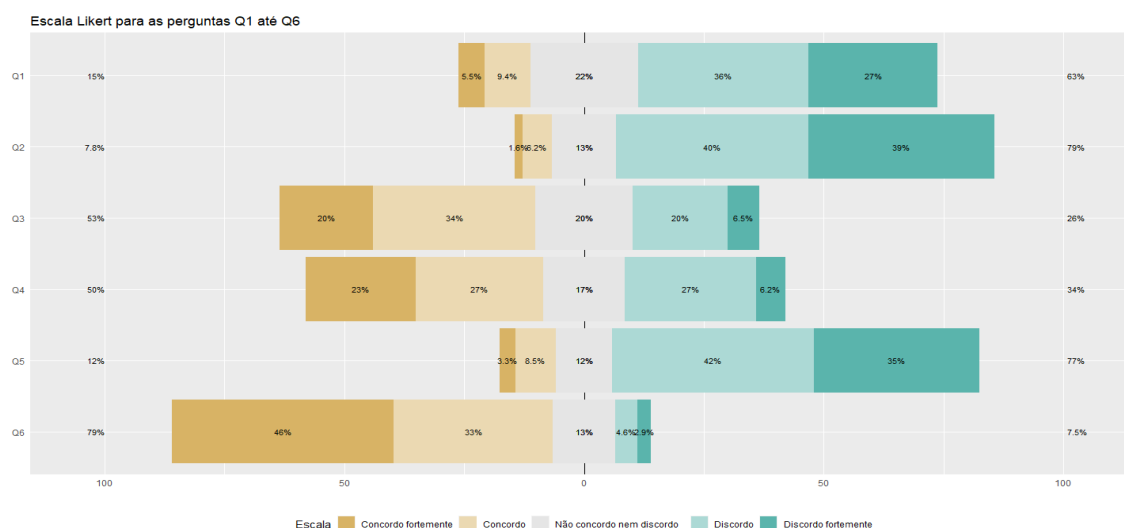


Figura 4: Escala likert para perguntas Q1 a Q6

Note que pela figura 5, as perguntas Q7, Q8, Q9, Q12 apresentaram uma quantidade grande de respostas neutras, cerca de 25%, porém a Q12 se destaca pelo sua grande quantidade de respostas discordantes, juntamente com Q11 e Q10.

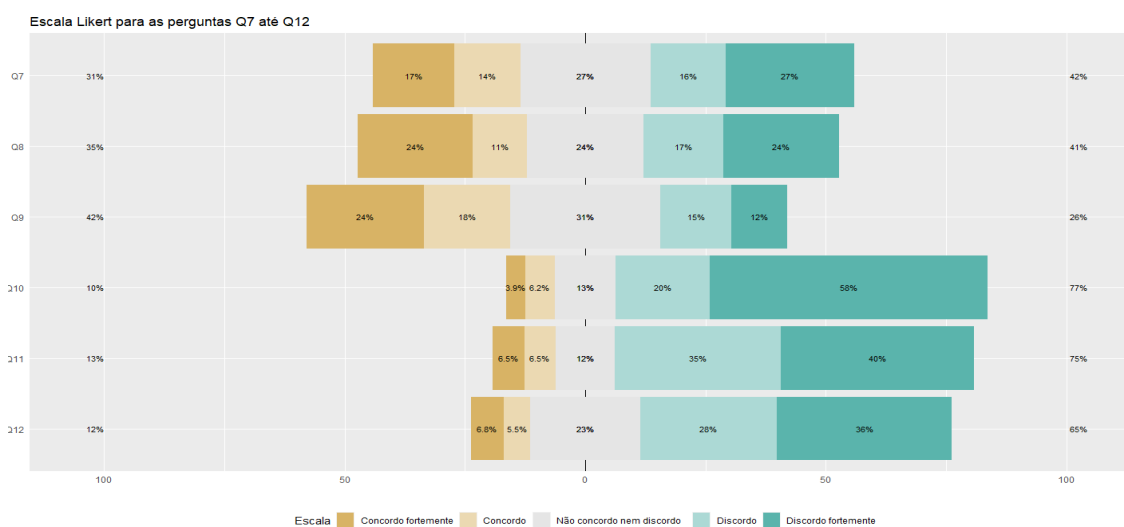


Figura 5: Escala likert para perguntas Q7 a Q12

Já na figura 6 é possível notar que a pergunta Q13 a grande maioria dos indivíduos, cerca de 74% responderam como "discordo fortemente". Para as perguntas Q14, Q15 e Q16 houve um número significativo de respostas neutras, cerca de 20%. Já a Q17 apresentou um pouco mais de concordância em relação a discordância.

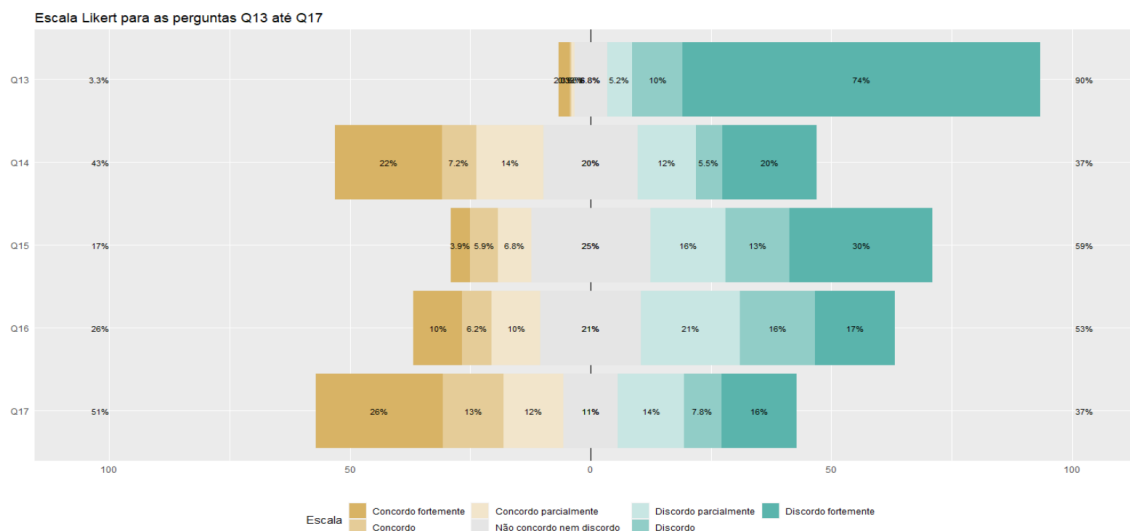


Figura 6: Escala likert para perguntas Q13 a Q17

A fim de verificar se o conjunto de variáveis que estão em Escala Likert medem alguma característica ou informação que não estava sendo observada diretamente, foi realizada uma análise fatorial. Para verificar a adequabilidade da amostra, foi calculada a medida de KMO. Observando a Tabela 2 e comparando os resultados com a Tabela 1 percebe-se que nenhuma pergunta apresenta KMO menor que 0.5, logo todas as perguntas são utilizadas na construção dos fatores.

Tabela 2: Valores da medida de KMO

Variável	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Overall MSA
Medida	0.78	0.69	0.79	0.79	0.64	0.79	0.84	0.66	0.67	0.65	0.66	0.73	0.74	0.67	0.60	0.59	0.77	0.72

Em seguida, para verificar se existia correlação entre as perguntas do questionário, foi realizado o teste de *Bartlett*, o teste retornou um p-valor menor que 0.05, logo rejeita-se a hipótese nula, isto é, a matriz de correlação não é a matriz identidade, portanto existe correlação entre as perguntas do questionário.

Para decidir o número ideal de fatores a serem construídos, foi utilizado o critério de Kaiser, que indicou 5 como o número ideal de fatores. Entretanto, após construir os fatores, foi observado que reduzir o número de fatores para 3 facilitava a explicação dos mesmos. Foram construídos então 3 fatores que estão representados na Figura 7. O primeiro fator agrupou perguntas sobre como o indivíduo se analisa, este fator foi chamado de fator psicológico. O segundo fator agrupou perguntas quanto a situação do indivíduo, este fator foi chamado de fator pessoal. Por fim, o terceiro fator agrupou perguntas quanto á ideologias que o indivíduo possui, este fator foi chamado de fator ideológico.

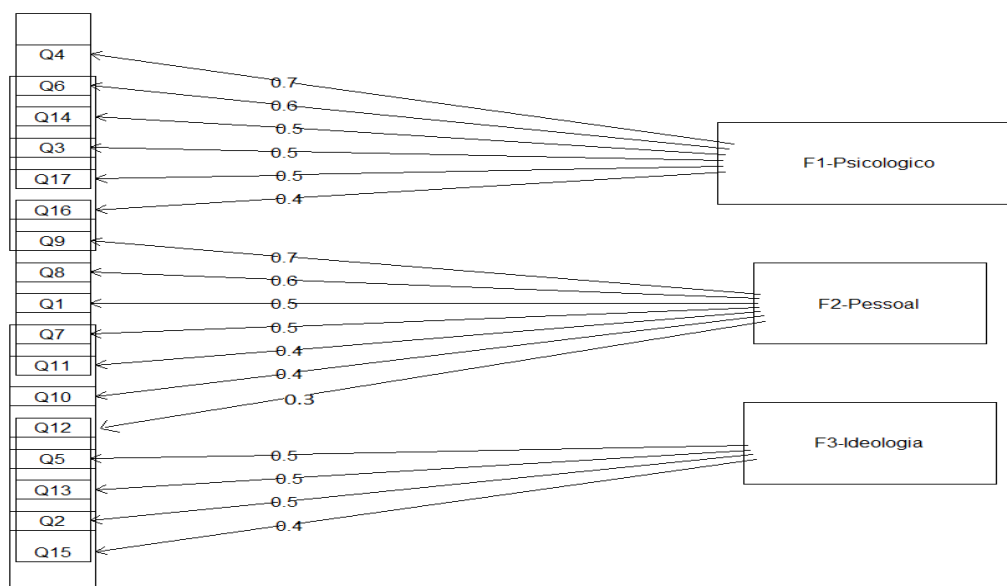


Figura 7: Diagrama dos fatores

Utilizando os fatores criados, foram calculados escores. Os escores são pontuações obtidas através do somatório da multiplicação das respostas dadas, pela matriz de pesos que é obtida através dos fatores. Os escores do fator psicológico variam de -2.36 a 2.47 , em que indivíduos que responderam mais perguntas com menores valores na escala likert (1, 2, 3) apresentaram valores mais próximos de -2.36 , enquanto que indivíduos que responderam mais perguntas com maiores valores na escala likert (4, 5, 6 ou 7) apresentaram valores mais próximos de 2.47 . Através desse escore, interpreta-se que indivíduos que apresentam menores escores aparentam ter um psicológico melhor.

Os escores do fator Pessoal variam de -2.59 a 2.06 , em que indivíduos que responderam mais perguntas com menores valores na escala likert (1, 2, 3) apresentaram valores mais próximos de -2.59 , enquanto que indivíduos que responderam mais perguntas com maiores valores na escala likert (4, 5, 6 ou 7) apresentaram valores mais próximos de 2.06 . Através desse escore, interpreta-se que indivíduos que apresentam menores escores aparentam estar menos satisfeitos com a sua situação pessoal.

Os escores do fator Ideológico variam de -3.30 a 2.02 , em que indivíduos que responderam mais perguntas com menores valores na escala likert (1, 2, 3) apresentaram valores mais próximos de -3.30 , enquanto que indivíduos que responderam mais perguntas com maiores valores na escala likert (4, 5, 6 ou 7) apresentaram valores mais próximos de 2.02 . Através desse escore, interpreta-se que indivíduos que apresentam menores escores aparentam ser mais negativos quanto a si mesmos e a situação em que se encontram.

Observando a Tabela 3, ela apresenta os escores do indivíduo 46. Interpretando os valores obtidos, o indivíduo 46 seria um indivíduo com um psicológico bom, é um pouco insatisfeito com a situação pessoal e aparenta ser neutro quanto a si mesmo e a situação em que se encontra.

Tabela 3: Exemplo de escore

Indivíduo	Escore fator 1	Escore fator 2	Escore fator 3
46	-2.36	-1.02	-0.21

Obtidos os escores, houve um questionamento do consultante para verificar uma possível relação entre os escores e as variáveis sociodemográficas presentes na tabela 7 presente no apêndice. Para verificar tal relação utiliza-se de modelos de regressão linear múltipla [5]. Foram ajustados três modelos. A tabela 4 apresenta as estimativas do modelo 1, referente aos escores obtidos do fator 1. A tabela 5 apresenta as estimativas do modelo 2, referente aos escores obtidos do fator 2. A tabela 6 apresenta as estimativas do modelo 3, referente aos escores obtidos do fator 3.

Primeiramente, os modelos foram ajustados utilizando todas as variáveis sociodemográficas, em seguida verificou-se a significância das variáveis para a construção do modelo, isto é, se elas são relevantes para o ajuste

do modelo. Para isso, utiliza-se o teste t de significância individual, se a variável apresentar um p-valor maior que 0.05 então ela é pouco relevante e se remove esta variável do modelo, caso contrário, ela é mantida no ajuste. Foi realizada também uma verificação de multicolinearidade através do VIF, foi verificado que nenhuma variável apresentava multicolinearidade.

Através do teste t, ajustou-se o modelo 1, presente na Tabela 4, que indica que apenas a variável escolaridade apresenta alguma relação com os escores gerados pelo primeiro fator. Através do teste de Shapiro-Wilk, foi obtido um p-valor de 0.12 para os resíduos, que é maior que 0.05, indicando que os resíduos do modelo apresentaram normalidade. Aplicando o teste de homocedasticidade de White, foi obtido um p-valor de 0.77, logo, os resíduos apresentam variância igual. Por fim, foi realizado o teste de distâncias de Cook para a verificação de observações influentes.

Tabela 4: Estimativas do modelo 1

Preditores	Psicológico	
	Estimativas	p-valor
(Intercepto)	0.32	0.002
escolaridade: Fundamental	Referência	
escolaridade: Médio	-0.43	<0.001
escolaridade: Completo	-0.65	0.04
Observações	304	
R ² / R ² ajustado	0.045 / 0.039	

Para o segundo modelo indicado na Tabela 5, apenas a variável mangueira apresentou uma relação com os escores do fator 2. Já a variável branco, apesar de violar o teste t, apresentando um p-valor maior que 0.05 não é muito distante deste valor, podendo então ser considerada relevante. Através do teste de Shapiro-Wilk, foi obtido um p-valor de 0.076 para os resíduos, que é maior que 0.05, indicando que os resíduos do modelo apresentaram normalidade. Aplicando o teste de homocedasticidade de White, foi obtido um p-valor de 0.73, logo, os resíduos apresentam variância igual. Por fim, foi realizado o teste de distâncias de Cook para a verificação de observações influentes.

Tabela 5: Estimativas do modelo 2

Preditores	Pessoal	
	Estimativas	p-valor
(Intercepto)	-0.18	0.326
mangueira: Sim	Referência	
mangueira: Não	-0.29	0.006
branco: Sim	Referência	
branco: Não	0.35	0.056
Observações	304	
R ² / R ² ajustado	0.039 / 0.032	

O modelo 3, presente na Tabela 6, não apresentou nenhuma variável significativa, além disso, através do teste de Shapiro-Wilk, foi obtido um p-valor menor que 0.001 para os resíduos, logo os resíduos do modelo não apresentaram distribuição normal, violando uma das hipóteses do modelo de regressão linear.

Tabela 6: Estimativas do modelo 3

Preditores	Ideológico	
	Estimativas	p-valor
(Intercepto)	0.03	0.910
mangueira: Sim	Referência	
mangueira: Não	-0.17	0.111
branco: Sim	Referência	
branco: Não	-0.08	0.693
homem: Sim	Referência	
homem: Não	0.14	0.210
mora_pai_mãe: Sim	Referência	
mora_pai_mãe: Não	-0.03	0.802
filhos: Sim	Referência	
filhos: Não	-0.13	0.286
escolaridade: Fundamental	Referência	
escolaridade: Médio	0.22	0.077
escolaridade: Completo	0.03	0.931
Observações	304	
R^2 / R^2 ajustado	0.032 / 0.009	

4 Conclusão

A partir da base de dados fornecida e nas reuniões realizadas com o consultante foi possível entender o problema, o que é fundamental para saber quais ferramentas estatísticas seriam aplicadas para gerar uma resposta.

Foi necessário realizar um pequeno tratamento na base de dados para que as respostas dos indivíduos seguissem o mesmo sentido. Foi realizada uma análise descritiva nos dados, em relação as variáveis sociodemográfica foi evidenciado que a base de dados é bem dividida em relação aos indivíduos que realizaram tratamento, cerca de 55% realizaram enquanto 45% não. Em relação ao sexo dos respondentes, houve um equilíbrio em torno dos 50% para ambos. Também houve equilíbrio para o número os que residem ou não na mangueira. Há uma grande disparidade em relação a raça dos indivíduos, cerca de 92% deles são não brancos. Em geral, a maioria (70%) mora com os pais, não tem filhos (69%) e tem ensino médio completo (71%).

Na construção dos fatores, foi decidido utilizar um número menor de fatores do que o indicado pelo método de Kaiser, pois isso facilitava a interpretação dos mesmos. Com isso foram construídos 3 fatores, o primeiro agrupou perguntas sobre como o indivíduo se analisa, e foi chamado de fator psicológico. O segundo agrupou perguntas quanto a situação do indivíduo, este foi chamado de fator pessoal. Por fim, o terceiro fator agrupou perguntas quanto á ideologias que o indivíduo possui, este fator foi chamado de ideológico.

Por fim, foram construídos modelos de regressão linear múltipla, analisando as tabelas 4, 5 e 6, pode se observar que o fator de explicação dos modelos R^2 e R^2 ajustado possuem uma baixo poder explicativo da sua variabilidade. O modelo 1 indicou que a escolaridade possui relação com os escores do fator 1, e as variáveis branco e mangueira possuem relação com os escores do fator 2. Além disso, o modelo 3 da tabela 6 não apresentou nenhuma variável significativa e violou as hipóteses básicas de 2.1.

No geral os fatores criados não explicavam muito a variação dos dados, porém ficaram agrupados de forma compreensível e concisa. Ao tentar observar a relação dos escores com as variáveis sociodemográficas não foi observado muitas relações entre eles.

Referências

1. B.C. BISTAFFA. Incorporação de indicadores categóricos ordinais em modelos de equações estruturais. 2010.
2. Thomas J. Farrar. *skedastic: Heteroskedasticity Diagnostics for Linear Regression Models*. Bellville, South Africa, 2020. URL <https://github.com/tjfarrar/skedastic>. R Package Version 1.0.0.
3. Damodar N. Gujarati and C. Porter Dawn. *Econometria Básica*. AMGH, 2011.
4. Sueli A. Mingoti. *Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada*. Editora UFMG, 2017.
5. J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied Linear Statistical Models*. Mc Graw Hill International, 2005. ISBN 978-1-259-06474-6.
6. William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2020. URL <https://CRAN.R-project.org/package=psych>. R package version 2.0.9.
7. Dean W. Wichern Richard A. Johnson. *Applied Multivariate Statistical Analysis*. Pearson, 2018.
8. Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.

5 Apêndice

Tabela 7: Variáveis sociodemográficas

Variável	Significado
mangueira	1 se o indivíduo é da mangueira, 0 caso contrário
branco	1 se o indivíduo é branco, 0 caso contrário
homem	1 se o indivíduo é homem, 0 caso contrário
mora_pai_mae	1 se o indivíduo mora com os pais, 0 caso contrário
filhos	1 se o indivíduo também, 0 caso contrário
escolaridade	0 se a escolaridade do indivíduo for fundamental, 1 for médio e 2 se for superior

Tabela 8: Perguntas do questionário

Variável	Significado
Q1	Eu sou tão apto a fazer coisas como a maioria das pessoas
Q2	Eu sou tão apto a fazer coisas como a maioria das pessoas
Q3	Eu sinto que eu não tenho muito do que me orgulhar
Q4	Eu me sinto inútil às vezes
Q5	Eu sinto que tenho no mínimo tanto valor quanto qualquer outra pessoa
Q6	No geral, eu estou inclinado a me sentir como um fracassado
Q7	Seus estudos
Q8	Seu trabalho
Q9	Sua situação econômica
Q10	Sua saúde
Q11	Sua imagem/aparência física
Q12	Seus amigos
Q13	O caminho que a minha vida segue depende de mim
Q14	O que uma pessoa conquista na vida é principalmente uma questão de destino ou sorte
Q15	Se uma pessoa é social ou politicamente ativa, pode conseguir modificar as condições sociais
Q16	As oportunidades que eu tenho na vida dependem de circunstâncias sociais
Q17	Tenho pouco controle sobre as coisas que acontecem na minha vida

Teste de Bartlett

$$\begin{cases} H_0 : & \text{A matriz de correlações é a matriz identidade} \\ H_1 : & \text{A matriz de correlações não é a matriz identidade} \end{cases}$$

	Tratamento	Branco	Mangueira	Homem	Mora_pai_mae	Filhos
Nao	138	278	156	149	214	95
Sim	166	26	148	155	90	209

	Escolaridade
Fundamental	78
Médio	217
Superior	9