

### Pedido do cliente:

Prezados,

Me chamo Roberto Sampaio e trabalho com projetos sociais. Recentemente implementamos um projeto, chamado Jovens Construtores em duas comunidades (Borel e Mangueira). Alguns jovens receberam o treinamento (tratamento) e outros só participaram de encontros presenciais (controle).

Já realizamos várias comparações entre os grupos (participaram e não participaram), encontramos alguns resultados interessantes. Entretanto, existe um grupo de variáveis que foram coletadas no final do treinamento com todos os indivíduos avaliados que não sabemos bem como trabalhar. Todas as variáveis são do tipo escala likert, de modo que para algumas delas 1 = "Concordo Fortemente" até 5="Discordo Fortemente", o dicionário da base contém uma explicação para cada variável.

Gostaríamos de realizar alguma análise estatística para extrair o que essas variáveis possam estar medindo, mas não sabemos como realizar tal procedimento. Vocês poderiam nos ajudar?

Gostaria que vocês me passassem um orçamento e o tempo para a realização do trabalho. Como o trabalho está em construção, novas demandas podem surgir à medida que vocês forem me indicando o que estão fazendo, ok?

Att.

Roberto Sampaio

---





## Introdução

Esse relatório apresenta os resultados obtidos a partir de análises feitas através das perguntas respondidas por participantes do projeto Jovens Construtores que visa dar um treinamento a pessoas que vivem em comunidade, visando uma melhora da parte relacional e autoestima dos indivíduos. Esse estudo possui os objetivos de:

- Extrair medidas explicadas através das questões respondidas através da escala Likert.
- Avaliar o impacto das variáveis sociodemográficas de forma conjunta nas medidas extraídas

## Metodologia

### TRATAMENTO DE DADOS

Nesta primeira etapa do projeto foi realizada uma inspeção na base de dados visando adequá-la às análises que seriam feitas posteriormente. Nesse processo foi identificada a presença de 3 observações com dados faltantes. Uma vez que essa quantidade representava apenas 0,9% da base de dados, optouse por apenas excluí-las das análises. Por fim, foi feita toda a decodificação das variáveis categóricas que seriam analisadas.

### BASE DE DADOS

Teve-se bastante atenção à interpretação de cada questão presente no banco de dados. A Tabela 1 apresenta, de forma geral, as 17 perguntas respondidas pelos indivíduos.

**Tabela 1: Questões analisadas via Análise Fatorial**

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

Note que as perguntas estão separadas em grupos, onde cada grupo específico possui uma escala diferente de resposta.

## ANÁLISE DE EXTRAÇÃO DE MEDIDAS

Para extrair o que mediam as perguntas presentes no questionário, foi realizada uma análise fatorial. A análise fatorial busca sintetizar as relações observadas entre um conjunto de perguntas interrelacionadas, de forma a avaliar possíveis fatores que poderiam ser explicados através das perguntas respondidas. Para a realização da análise fatorial, é realizado diversos testes para verificar a Tabela 1: Questões analisadas via Análise Fatorial 5 adequabilidade da técnica aos dados e para associar perguntas inter-relacionadas a um determinado fator de forma a gerar conclusões interpretáveis.

Para a avaliação dos possíveis fatores, foram retiradas 4 questões das 17 disponíveis, tendo realizado a análise então com apenas 13 das 17 questões. Essa eliminação foi realizada admitindo alguns critérios: as questões 15 e 16 possuíam um índice de adequabilidade à técnica de análise fatorial abaixo de 0.6, considerados miseráveis segundo Daniel Abud (2019) [1]. Além disso, após a remoção dessas duas perguntas, as questões 12 e 13 se tornaram não significantes na explicação dos fatores, afetando negativamente nos valores obtidos. Outro critério analisado foi a correlação entre as perguntas do questionário, onde se obteve que as questões 13 e 15 obtiveram pouca correlação entre as outras questões. Para confirmar que a utilização da análise fatorial era válida foi realizado um Teste de Esfericidade de Bartlett, o que possibilitou o prosseguimento da análise.

Para se definir o número de fatores explicados pelas perguntas, foram utilizados os critérios de Kaiser, critério de variância explicada e o critério do gráfico Scree. Porém, também se deu a importância na clareza das interpretações dos fatores. Dessa forma, foi definido a construção de apenas 4 fatores.

Esses fatores podem ser construídos através de diversos métodos e aplicando metodologias diferentes de rotações. Normalmente o usual é o método de Máxima Verossimilhança, porém como os dados não seguiam uma distribuição Normal, esse método foi considerado inadequado para a realização da análise, sendo substituído pelo método de Mínimos Quadrados Generalizados, que obedeciam a todos os pressupostos para a realização da análise. Quanto a rotação, foi utilizada uma rotação oblíqua *promax* de forma a avaliar, além das relações entre as perguntas, as relações entre os fatores construídos, avaliando possíveis correlações. Existem diversas rotações oblíquas, porém a rotação *promax* nos resultou em associações melhor interpretáveis e com números de comunalidade superiores às outras rotações oblíquas. Os fatores construídos explicam 44% da variância dos dados, valor considerado baixo.

*## Determinação do número de fatores*

*# Calculo dos autovalores*

*round(eigen(correlacao\$rho)\$values,2)*

*# Numero de autovalores maiores que 1*

*sum(eigen(correlacao\$rho)\$values>1)*

*# Scree-plot*

*plot(eigen(correlacao\$rho)\$values, type = "b", ylab = "Autovalores")*

```
# Calculo da proporcao explicada por cada fator
```

```
proporcao_explicacao <- eigen(correlacao$rho)$values/sum(eigen(correlacao$rho)$values)
```

```
# Calculo da proporcao de explicacao acumulada
```

```
proporcao_acumulada <- cumsum(proporcao_explicacao)
```

```
# Imprime a proporcao acumulada com 2 casas decimais
```

```
round(proporcao_acumulada, 2)
```

```
## Extração das cargas fatoriais e Rotação dos fatores
```

```
## Extracao dos 4 fatores sem rotacao e usando o metodo ml
```

```
fa_com_rotacao <- fa(correlacao$rho, n_fac, rotate="promax", fm="gls")
```

```
# Mostra as cargas fatoriais
```

```
loadings(fa_com_rotacao)
```

```
print(fa_com_rotacao)
```

```
fa.diagram(fa_com_rotacao)
```

```
#comunalidades
```

```
round(fa_com_rotacao$communalities,3)
```

```
poly_model <- fa(data_q,nfactors=4,cor="poly",fm="gls",rotate="promax")
```

```
my.scores = poly_model$scores
```

## **AVALIAÇÃO DO IMPACTO DAS VARIÁVEIS SOCIODEMOGRÁFICAS**

Por fim, com o objetivo de verificar como as variáveis sociodemográficas impactam nos valores desses escores de forma conjunta, foram construídos modelos lineares múltiplos para cada um dos fatores. A construção e avaliação do impacto das variáveis foram analisados segundo DOBSON (2019)[2]. Foi utilizado o método de seleção de variáveis Stepwise, na criação de alguns dos modelos. Os testes de adequabilidade do modelo estão presentes no ANEXO B, caso queira verificar. Outro fator analisado nos modelos foi o possível problema de multicolinearidade, que não se fez presente em nenhum dos quatro modelos estimados. Como forma de complementar as análises, foram realizadas análises inferenciais das variáveis sociodemográficas, porém com um olhar univariado. Foram aplicados testes de comparação de médias (Teste T e ANOVA) para verificar se as médias dos scores calculados se diferenciam em diferentes grupos. Para verificação de pressupostos, foram aplicados testes de normalidade e de igualdade de variância, para a utilização da metodologia correta de análise. [3] Foi adotado um nível de significância de 5% para os testes aplicados.

## **SOFTWARES**

Todas as análises estatísticas foram realizadas pelo Software R. [3] Também foram usados o Microsoft Excel e o Microsoft Word para a criação de tabelas e elaboração do relatório.

#### **Pacotes utilizados ao longo do trabalho:**

```
library(readxl)
library(tidyverse)
library(psych)
#library(MASS)
#require(DAAG)
require(RColorBrewer)
library(tidyr)
library(ggthemes)
library(rsq)
library(flextable)
library(stringr)
require(corrplot)
```

## **Apresentação dos Scores**

### **Análise Fatorial**

Após a seleção das variáveis, após critérios citados na seção anterior, a análise inicial consistiu na avaliação da estrutura de correlação entre as 13 questões selecionadas por meio da construção de uma matriz de correlação policórica. O valor desta correlação varia entre -1 e 1, com valores mais próximos dos extremos significando relações de correlação mais fortes e valores mais próximos de 0 que o par de variáveis não é correlacionado. A matriz de correlação é apresentada a seguir na Figura 4.

```
## Calcula a correlacao
correlacao <- polychoric(data_q)

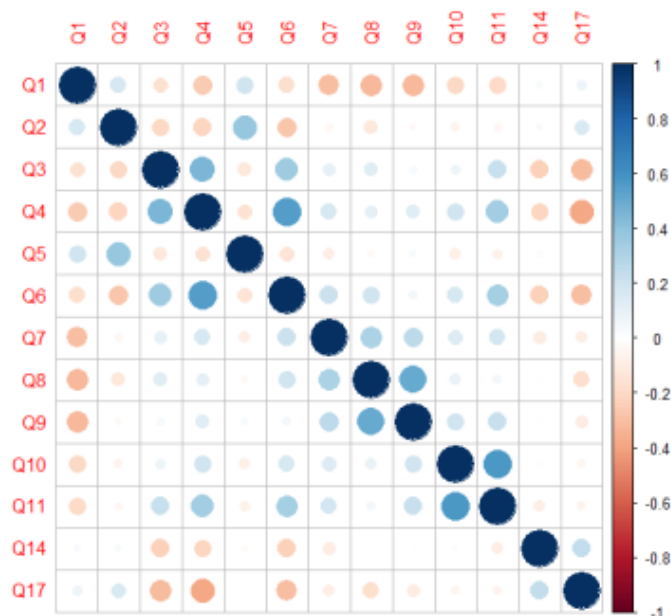
## Imprime as correlacoes com 2 casas decimais
round(correlacao$rho,2)

## Gráfico com essas correlações
corPlot(correlacao$rho, numbers=TRUE, upper=FALSE, diag=FALSE)

## Passo 4 – Teste de Bartlett (BTS)
# Teste bartlett
cortest.bartlett(correlacao$rho,n=nrow(data_q))

## Passo 5 – Teste de Kaiser-Meyer-Olkin (KMO)
KMO(correlacao$rho)
```





**Figura 4: Matriz de correlação policórica para as questões significativas na análise fatorial**

Nota-se que algumas perguntas são correlacionadas, tanto positivamente quanto negativamente entre as outras, enquanto há também a não correlação entre outras perguntas. Através disso, começa a se construir os cenários e hipóteses de grupos de questões associadas à fatores que serão construídos. Na realização da análise fatorial foram utilizadas 13 das 17 questões respondidas pelos indivíduos participantes do projeto Jovens Construtores. A Tabela 1 abaixo apresenta as questões utilizadas na análise e suas respectivas escalas.

Diante das perguntas selecionadas e apresentadas, tem-se a realização da análise fatorial com o objetivo de construir fatores associadas às perguntas inter-relacionadas. Através disso, foram construídas 4 medidas que serão apresentadas de forma detalhada neste relatório. Para confirmar que a utilização da análise fatorial era válida foi realizado um Teste de Esfericidade de Bartlett e, a um nível de significância de 5%, a hipótese nula foi rejeitada, o que possibilitou o prosseguimento da análise. Além disso, calculouse o calor de adequabilidade da base de dados à técnica e todas as questões selecionadas obtiveram o índice de KMO superiores a 0,6. O Anexo A contém mais detalhes dos testes aplicados na análise fatorial. Após a estimação dos fatores através do método dos mínimos quadrados generalizados e método de rotação promax, foram construídos 4 fatores. A Figura 5 apresenta a árvore de associação de cada fator construído, atribuindo pesos para cada uma das questões associadas.

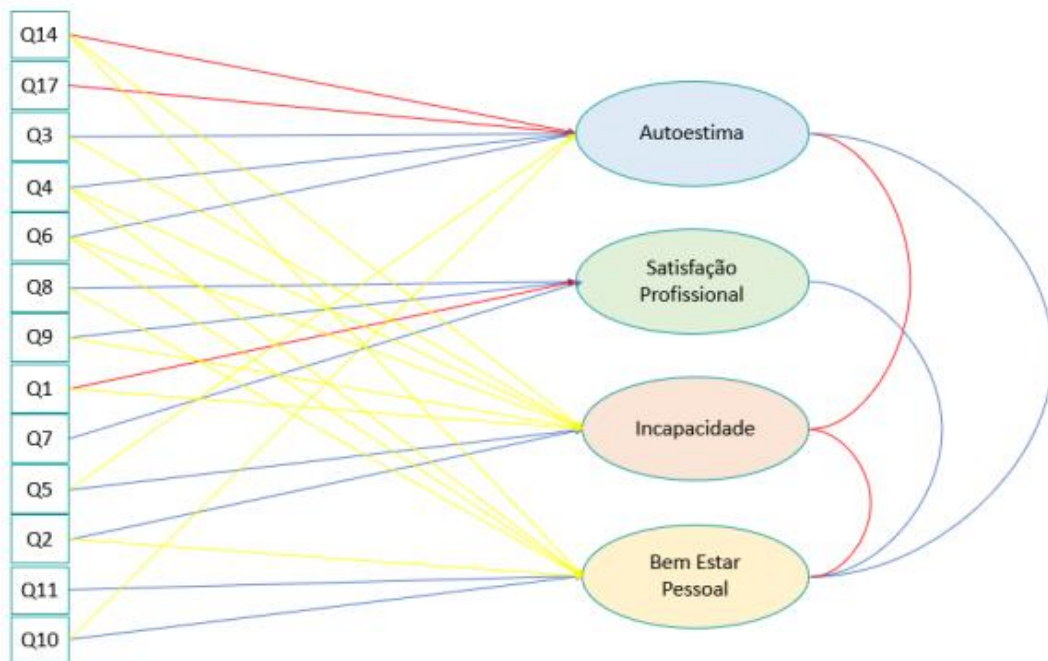


Figura 5: Árvore de Fatores

Nota-se que na Figura 5 apresentada temos associações em azul, vermelho e amarelo. As associações em azul e vermelho são as consideradas fortes. As retas azuis são as associações cujas cargas fatoriais são altas, sendo maior ou iguais a 0,5, influenciando assim, no cálculo do score de maneira positiva. As retas vermelhas são as associações cujas cargas fatoriais são menores ou iguais a -0,5, indicando influência negativa para o cálculo do fator em questão. As retas em amarelo representam as associações mais fracas, mas que podem impactar no cálculo dos scores de maneira positiva ou negativa, porém, por sua natureza, possuem menos importância.

Outra característica dos fatores é que pode haver uma associação entre eles. As associações apresentadas na região direita da Figura 5 apresentam as correlações entre esses fatores. Dessa forma, tem-se que a autoestima, quando alta, afeta negativamente a incapacidade, enquanto que afeta positivamente o bem estar pessoal. Com isso, podemos concluir que, quando uma relação forte é apresentada pela cor azul, a relação é positiva, caso contrário, a relação é considerada negativa. O tamanho dessas associações será detalhado nas sessões dos respectivos fatores extraídos.

## Apresentação dos Scores

### Distribuição dos Scores

Os scores apresentaram médias centradas ao redor de 0 e suas medianas não se apresentaram muito diferentes, principalmente com relação aos escores de autoestima e satisfação profissional (os escores de Bem-estar Pessoal e Incapacidade apresenta valores um pouco



diferentes). O score de Incapacidade possui o menor mínimo (- 3,045) e o menor máximo (1,610), enquanto o score de Bem Estar Pessoal possui os maiores para ambas as medidas (- 1,835, mínimo e 2,903, máximo). Pode-se dizer que os scores de cada fator são distribuídos de maneira semelhante, mas com os extremos divergindo consideravelmente, como pode ser visto no gráfico da Figura 6 e nos valores da Tabela 2.

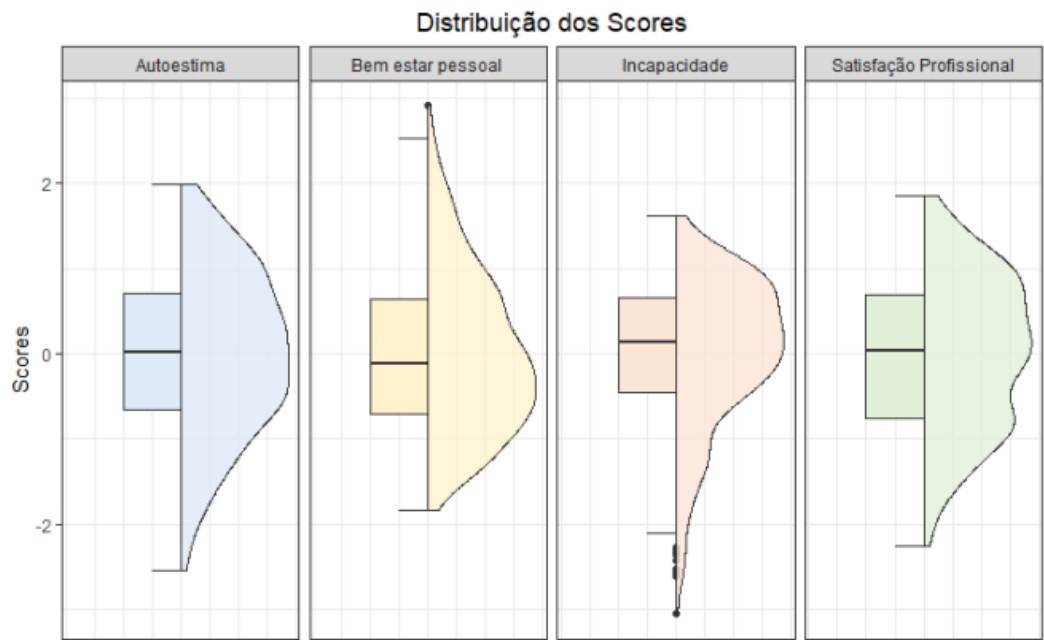


Figura 6: Distribuição dos Valores estimados dos Scores dos Fatores

Tabela 2: Medidas Descritivas dos Scores dos Fatores

Scores	Medidas Descritivas				
	Mínimo	Mediana	Média	Máximo	Desvio Padrão
Autoestima	-2,535	0,023	-0,001	1,979	0,967
Satisfação Profissional	-2,326	0,032	-0,001	1,788	0,904
Bem Estar Pessoal	-1,835	-0,111	0,000	2,903	0,938
Incapacidade	-3,045	0,142	0,010	1,610	0,869

## Apresentação dos Scores

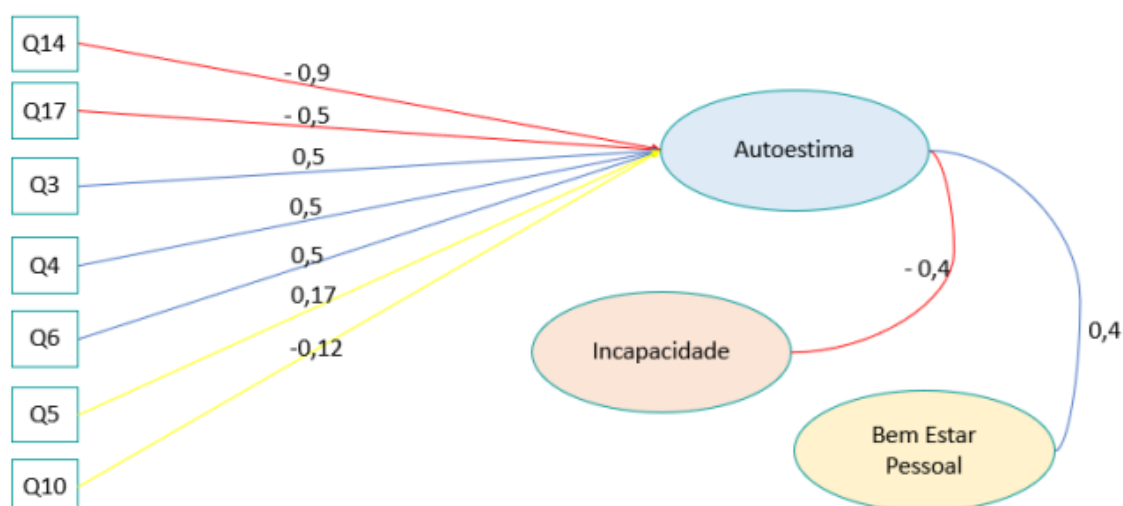
### Autoestima

Defina-se como autoestima a qualidade de quem se valoriza, se contenta com seu modo de ser e demonstra, conseqüentemente, confiança em seus atos e julgamentos. Através disso, temos as seguintes perguntas associadas ao cálculo dessa característica apresentado na Tabela 3.

**Tabela 3: Perguntas associadas à autoestima do indivíduo**

Perguntas	Peso	Interpretação em Relação a Escala
Q3 - Eu sinto que eu não tenho muito do que me orgulhar.	0,493	V
Q4 - Eu me sinto inútil às vezes.	0,489	V
Q6 - No geral, eu estou inclinado a me sentir como um fracassado.	0,456	V
Q14 - O que uma pessoa conquista na vida é principalmente uma questão de destino ou sorte.	-0,918	X
Q17 - Tenho pouco controle sobre as coisas que acontecem na minha vida.	-0,534	X

Baseado nas escalas das perguntas Q3, Q4 e Q6, que associam a maior nota em discordar da afirmação, temos, para notas altas, um efeito positivo na pontuação de autoestima do indivíduo. Enquanto, em contraponto, as questões Q14 e Q17 possuem escalas opostas, associando a maior nota em concordar com a afirmação, tornando assim o efeito negativo na pontuação da autoestima. As influências de cada questão na autoestima estão apresentadas na Figura 7:



**Figura 7: Influências no score de Autoestima**

## Apresentação dos Scores

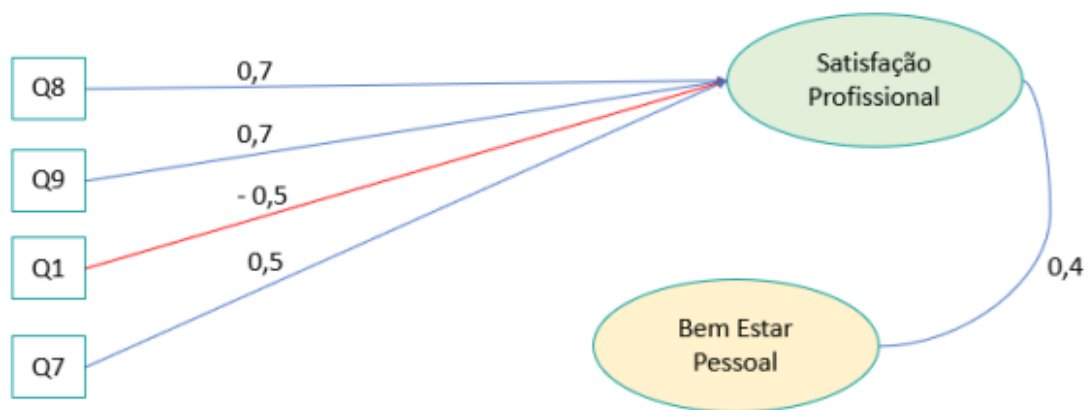
### Satisfação Profissional

A satisfação profissional pode ser definida como a satisfação do indivíduo em relação à sua carreira profissional seja quanto a sua área de estudo, remuneração, local de trabalho e a equipe/colegas ao seu redor. Através disso, temos as seguintes perguntas associadas ao cálculo dessa característica apresentado na Tabela 4.

**Tabela 4: Perguntas Associadas à satisfação profissional do indivíduo**

Perguntas	Peso	Interpretação em Relação a Escala
Q8 - Satisfação quanto ao seu trabalho.	0,716	V
Q9 - Satisfação quanto sua situação econômica.	0,715	V
Q1 - No geral, eu estou satisfeito comigo mesmo.	-0,501	X
Q7 - Satisfação quanto seus estudos.	0,47	V

Tendo base nas escalas das perguntas Q7, Q8 e Q9, que associam a maior nota em estar satisfeito temos, para notas altas, um efeito positivo na pontuação de satisfação profissional do indivíduo. Enquanto, em contraponto, a questão Q1 possui uma escala oposta, associando a maior nota em discordar com a afirmação, tornando assim o efeito negativo na pontuação da satisfação profissional. As influências de cada questão na satisfação profissional estão apresentadas na Figura 8:



**Figura 8: Influências no score de Satisfação Profissional**

## Apresentação dos Scores

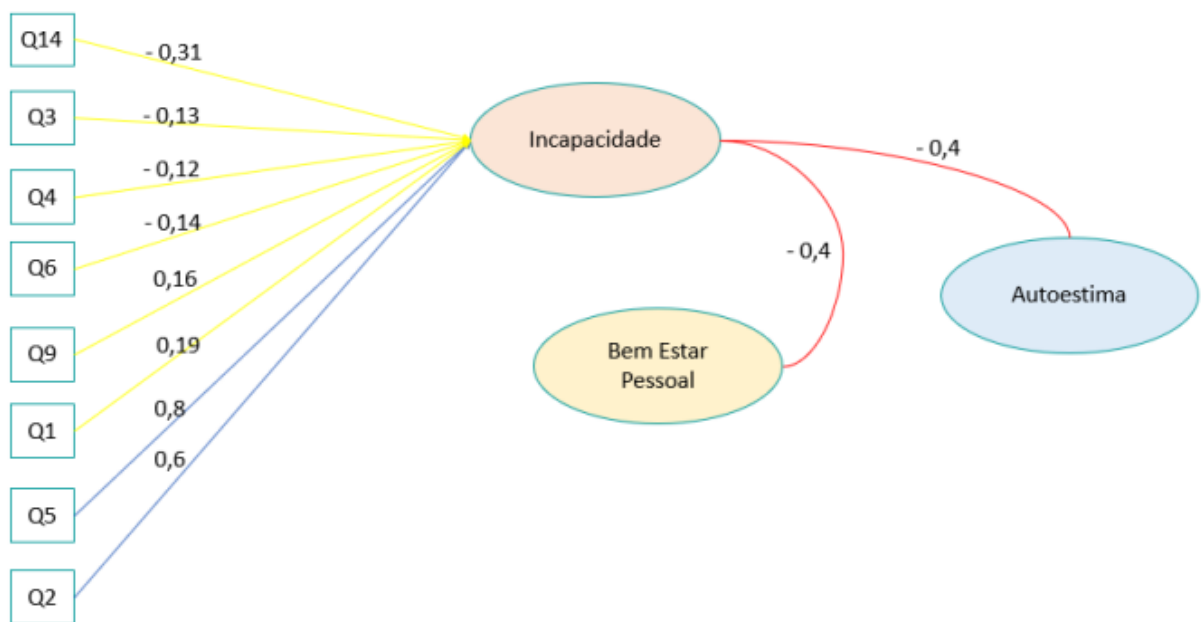
### Incapacidade

Defina-se incapacidade como restrição resultante de uma deficiência, da habilidade para desempenhar uma atividade considerada normal para o ser humano. Surge como consequência direta ou é resposta do indivíduo a uma deficiência psicológica, física, sensorial ou outra. Através disso, tem-se as seguintes perguntas associadas ao cálculo dessa característica apresentado na Tabela 5.

**Tabela 5: Perguntas associadas à Incapacidade do indivíduo**

Perguntas	Peso	Interpretação em Relação a Escala
Q2 - Eu sou tão apto a fazer coisas como a maioria das pessoas.	0,602	V
Q5 - Eu sinto que tenho no mínimo tanto valor quanto qualquer outra pessoa.	0,846	V

Baseado nas escalas das perguntas Q2 e Q5, que associam a maior nota em discordar da afirmação, temos, para notas altas, um efeito positivo na pontuação de incapacidade do indivíduo. As influências de cada questão nessa característica estão apresentadas na Figura 9:



**Figura 9: Influências na Incapacidade do Indivíduo**

## Apresentação dos Scores

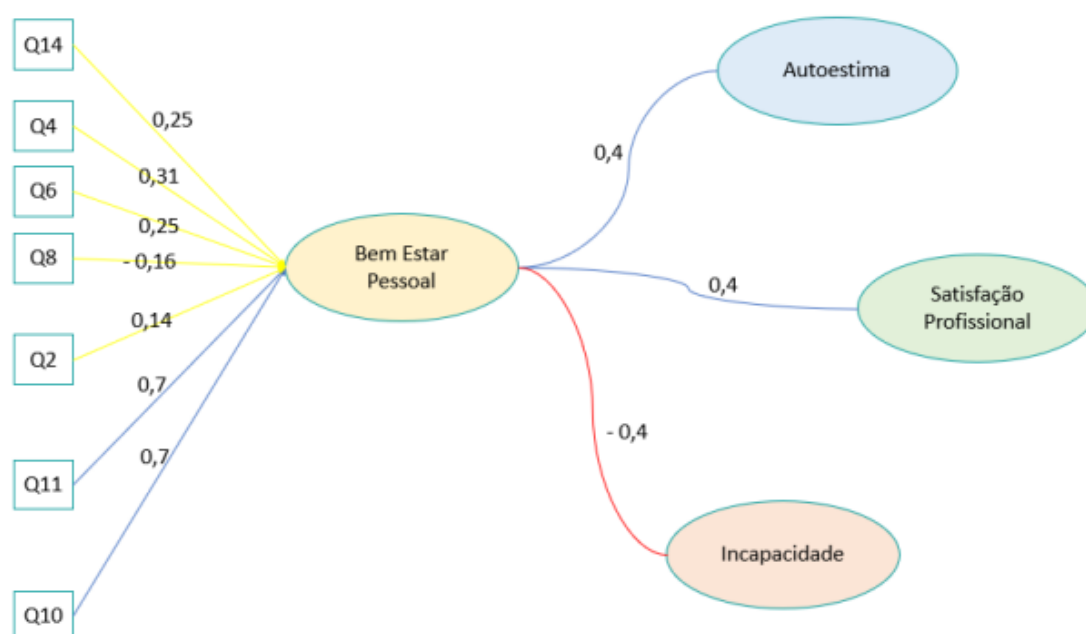
### Bem-estar Pessoal

O bem-estar pode ser definido como algo que influencia diretamente a felicidade e satisfação pessoal. Como existe a relação entre o bem-estar pessoal e à saúde ou à aparência física do indivíduo, tem-se as seguintes perguntas associadas ao cálculo dessa característica apresentadas na Tabela 6.

**Tabela 6: Perguntas associadas ao Bem Estar do indivíduo**

Perguntas	Peso	Interpretação em Relação a Escala
Q10 - Satisfação quanto a sua saúde	0,685	V
Q11 - Satisfação quanto a sua imagem/aparência física	0,727	V

Baseado nas escalas das perguntas Q10 e Q11, que associam a maior nota em estar satisfeito, temos, para notas altas, um efeito positivo na pontuação de bem-estar do indivíduo. As influências de cada questão nesta característica estão apresentadas na Figura 10:



**Figura 10: Influências na Bem Estar Pessoal**

## Impacto das Variáveis Sociodemográficas

### Análise Exploratória

Com o objetivo de averiguar de forma geral como as características dos indivíduos, será realizada uma breve análise exploratória. A distribuição de indivíduos para cada variável Sociodemográfica está apresentada na Figura 11.



**Figura 11: Distribuição dos indivíduos em relação às variáveis sociodemográficas**

Nota-se que, dentre os indivíduos participantes, 91% são não brancos e apenas 9% são brancos. Vale ressaltar também que grande maioria possui Ensino Médio Completo ou Incompleto, mora com os pais e não possui filhos. Enquanto isso, a distribuição de indivíduos se mantém a mesma quanto ao Sexo, residência e em relação à participação do programa.

## Impacto das Variáveis Sociodemográficas

### Autoestima

Ao avaliar o impacto das variáveis sociodemográficas no score de autoestima, foi construído um modelo linear normal. A Tabela 7 apresenta os efeitos estatisticamente significantes das variáveis sociodemográficas.



**Tabela 7: Impactos significativos na pontuação de Autoestima do indivíduo**

Variáveis	Efeito
Intercepto	-0,448
Grupo Controle	-0,1766
Homem	0,2202
EM Completo ou Incompleto	0,5734
Superior Completo	0,883
R <sup>2</sup>	0,0844

Esse modelo explica apenas 8,44% da variabilidade dos scores calculados, de forma a avaliar o efeito de cada característica definida. Esse número é considerado bem baixo e indica que através dos dados do questionário, as variáveis sociodemográficas não tem poder explicativo em relação à variabilidade da Autoestima.

A partir da Tabela 7, pode-se afirmar que, se um indivíduo é mulher, possui apenas o Ensino Fundamental completo e participou do programa, ele terá uma pontuação de autoestima média estimada de -0,448, pontuação representada pelo Intercepto. Esse indivíduo é considerado como indivíduo base, de forma que, a mudança de uma dessas características, mantendo as demais, gera uma mudança na pontuação média referente ao efeito associado a escolha realizada. É através disso que é possível realizar inúmeras comparações e ter algumas interpretações.

Como o modelo acaba explicando apenas 8% da variabilidade do score, as conclusões a serem retiradas acabam não sendo muito significativas. Com isso, foram realizadas algumas análises no ponto de vista univariado através das variáveis sociodemográficas cujos impactos foram considerados significativos e as diferenças entre os grupos foram significativas.

#### **# Modelo Autoestima**

```
full.modelo4<- lm(autoestima~.-incapacidade-bem_estar-satisfacao_profissional,
data=data_modelo)
modelo4 <- MASS::stepAIC(full.modelo4,direction = "both", trace = FALSE)
summary(modelo4)
```

```
DAAG::vif(modelo4)# Multicolinearidade OK
```

```
#Normalidade: Pelo grafico vemos que os residuos são normais
par(mfrow=c(1,1),mar=c(4,4,2,0.5))
qqnorm(modelo4$residuals,pch=20,cex=1)
qqline(modelo4$residuals,lwd=3,col=2)
```

```
ks.test(x=rstudent(modelo4),"pnorm",mean=mean(rstudent(modelo4)),sd=sd(rstudent(modelo4)))
```

```
#Homocedasticidade:
st_res = rstudent(modelo4)
plot(modelo4$fitted.values,st_res)
abline(h=0,col="red")
```

```
lmtest::bptest(modelo4)
```

## Autoestima – Homem x Mulher

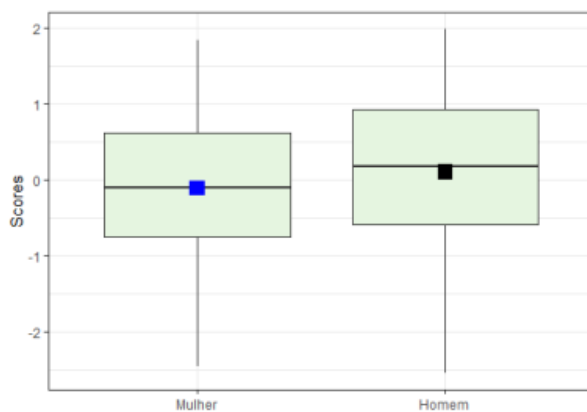


Figura 12: Boxplot de Autoestima por Sexo

Aplicado o Teste T, com p-valor = 0,06, tem-se, a um nível de significância de 5%, que as médias das pontuações de autoestima **são iguais** em ambos os sexos.

## Autoestima – Escolaridade Diferentes

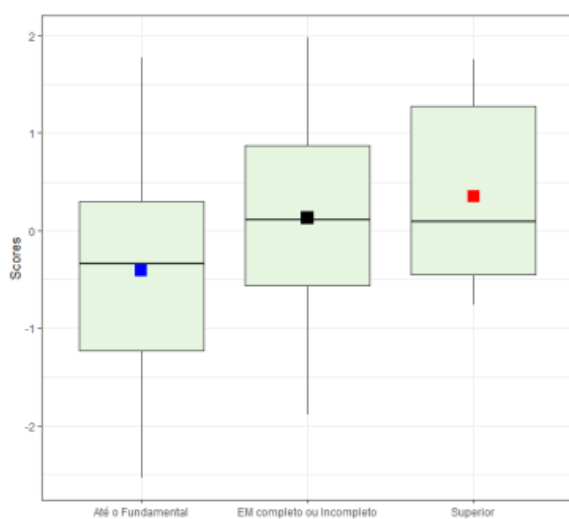


Figura 13: Boxplot de Autoestima por Escolaridade

Aplicado o Teste ANOVA, com p-valor < 0,01, tem-se, a um nível de significância de 5%, que as médias das pontuações de autoestima **são diferentes** entre as escolaridades. Através do teste de Bonferroni, tem-se que a média de autoestima de indivíduos com escolaridade "Até o Fundamental" se difere dos demais.

Mesmo que exista outras variáveis que possui impactos significantes no modelo, apenas essas duas apresentadas resultam em diferenças estatisticamente significantes dentro dos grupos.

```
ggplot(data_modelo, aes(y = autoestima, x = escolaridade)) + geom_boxplot(fill = palheta[2]) +  
stat_summary(fun="mean", color = c("blue", "black", "red"), shape = 15, size = 1) +  
labs(x = "", y = "Scores", title = "") +  
theme_bw()
```

### **## Escolaridade**

#### **#Testes de Normalidade**

```
ks.test(data_modelo$autoestima[data_modelo$escolaridade=="Até o  
Fundamental"], "pnorm", mean(data_modelo$autoestima[data_modelo$escolaridade=="Até o  
Fundamental"]), sd(data_modelo$autoestima[data_modelo$escolaridade=="Até o  
Fundamental"]))
```

```
ks.test(data_modelo$autoestima[data_modelo$escolaridade=="EM completo ou Incompleto"],  
"pnorm", mean(data_modelo$autoestima[data_modelo$escolaridade=="EM completo ou  
Incompleto"]), sd(data_modelo$autoestima[data_modelo$escolaridade=="EM completo ou  
Incompleto"]))
```

```
ks.test(data_modelo$autoestima[data_modelo$escolaridade=="Superior"],  
"pnorm", mean(data_modelo$autoestima[data_modelo$escolaridade=="Superior"]), sd(data_m  
odelo$autoestima[data_modelo$escolaridade=="Superior"]))
```

#### **#Teste de Comparação de Médias**

#Variancia OK

```
DescTools::LeveneTest(autoestima~escolaridade, data=data_modelo)
```

#### **#ANOVA**

```
oneway.test(autoestima~escolaridade, data=data_modelo, var.equal = T)
```

## **Impacto das Variáveis Sociodemográficas**

### **Satisfação Profissional**

Ao avaliar o impacto das variáveis sociodemográficas no score de Satisfação Profissional, foi construído um modelo linear múltiplo. A Tabela 8 apresenta os efeitos estatisticamente significantes das variáveis sociodemográficas:

**Tabela 8: Impactos significativos na Satisfação Profissional do Indivíduo**

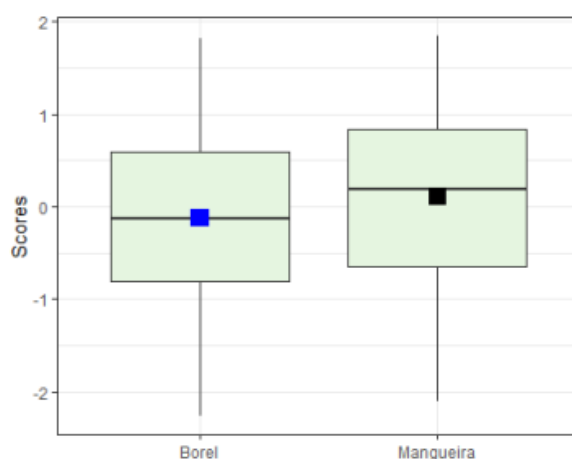
Variáveis	Efeito
Intercepto	0,0142
Grupo Controle	-0,201
Mangueira	0,2655
Raça: Branca	-0,4862
R <sup>2</sup>	0,0511

Esse modelo explica apenas 5,11% da variabilidade dos scores calculados, de forma a avaliar o efeito de cada característica definida. Esse número é considerado bem baixo e indica que através dos dados do questionário, as variáveis sociodemográficas não tem poder explicativo em relação à variabilidade da Satisfação Profissional.

A partir da Tabela 8, pode-se afirmar que, se um indivíduo é de raça Branca, residente da Mangueira e participou do programa, ele terá uma pontuação de satisfação profissional média estimada de 0,0142, pontuação representada pelo Intercepto. Esse indivíduo é considerado como indivíduo base, de forma que, a mudança de uma dessas características, mantendo as demais, gera uma mudança na pontuação média referente ao efeito associado a escolha realizada. É através disso que é possível realizar inúmeras comparações e ter algumas interpretações.

Como o modelo acaba explicando apenas 5% da variabilidade do score, as conclusões a serem retiradas acabam não sendo muito significativas. Com isso, foram realizadas algumas análises no ponto de vista univariado através das variáveis sociodemográficas cujos impactos foram considerados significativos e as diferenças entre os grupos foram significativas.

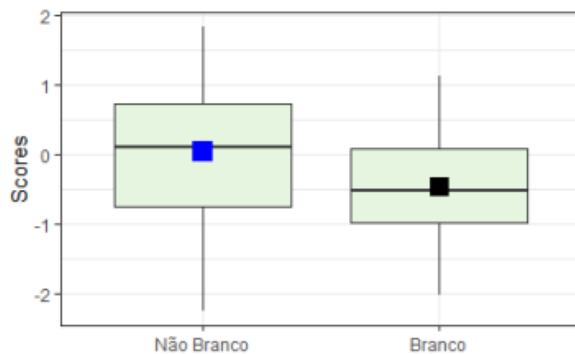
#### Satisfação Profissional – Borel x Mangueira



Aplicado o Teste T, com p-valor = 0,02, tem-se, a um nível de significância de 5%, que as médias das pontuações de Satisfação Profissional **são diferentes** entre indivíduos residentes da Mangueira e do Borel.

**Figura 14: Boxplot - Satisfação Profissional por Residência**

### Satisfação Profissional – Não Branco x Branco



Aplicado o Teste T, com p-valor < 0,01, tem-se, a um nível de significância de 5%, que as médias das pontuações de Satisfação Profissional **são diferentes** entre as pessoas de Raça Branca e Não Branca. Porém vale afirmar que a base possui 91% de indivíduos não brancos e 9% de indivíduos brancos.

Figura 15: Boxplot - Satisfação Profissional por Raça

Mesmo que exista outras variáveis que possui impactos significantes no modelo, apenas essas duas apresentadas resultam em diferenças estatisticamente significantes dentro dos grupos.

## Impacto das Variáveis Sociodemográficas

### Incapacidade

Ao avaliar o impacto das variáveis sociodemográficas no score de Incapacidade, foi construído um modelo linear múltiplo. A Tabela apresenta os efeitos estatisticamente significantes das variáveis sociodemográficas:

**Tabela 9: Impactos significativos na Incapacidade do Indivíduo**

Variáveis	Efeito
Intercepto	-0,3282
Grupo Controle	-0,0966
Mangueira	0,1919
Homem	0,1771
Possui filhos	0,155
EM Completo ou Incompleto	0,2188
Superior Completo	0,0248
R <sup>2</sup>	0,0305

Esse modelo explica apenas 3,05% da variabilidade dos scores calculados, de forma a avaliar o efeito de cada característica definida. Esse número é considerado bem baixo e indica que através dos dados do questionário, as variáveis sociodemográficas não tem poder explicativo em relação à variabilidade da Incapacidade.

A partir da Tabela 9, pode-se afirmar que, se um indivíduo é mulher, residente da Mangueira, possui apenas o Ensino Fundamental completo, não possui filhos e participou do programa, ele terá uma pontuação de Incapacidade média estimada de -0,3282, pontuação representada pelo Intercepto. Esse indivíduo é considerado como indivíduo base, de forma que, a mudança de uma dessas características, mantendo as demais, gera uma mudança na pontuação média referente ao efeito associado a escolha realizada. É através disso que é possível realizar inúmeras comparações e ter algumas interpretações.

Como o modelo acaba explicando apenas 3% da variabilidade do score, as conclusões a serem retiradas acabam não sendo muito significativas. Com isso, foram realizadas algumas análises no ponto de vista univariado através das variáveis sociodemográficas cujos impactos foram considerados significativos e as diferenças entre os grupos foram significativas.

Mesmo com muitas variáveis cujos impactos foram significativos, tem-se que ao realizar comparação entre os grupos, não há nenhuma característica do indivíduo que interfira na distribuição da pontuação de incapacidade.



## Impacto das Variáveis Sociodemográficas

### Bem-estar Pessoal

Ao avaliar o impacto das variáveis sociodemográficas no score de Autoestima, foi construído um modelo linear normal. A Tabela 10 apresenta os efeitos estatisticamente significantes das variáveis sociodemográficas:

Ao avaliar o impacto das variáveis sociodemográficas no score de Autoestima, foi construído um modelo linear normal. A Tabela 10 apresenta os efeitos estatisticamente significantes das variáveis sociodemográficas:

**Tabela 10: Impactos significativos no Bem Estar do indivíduo**

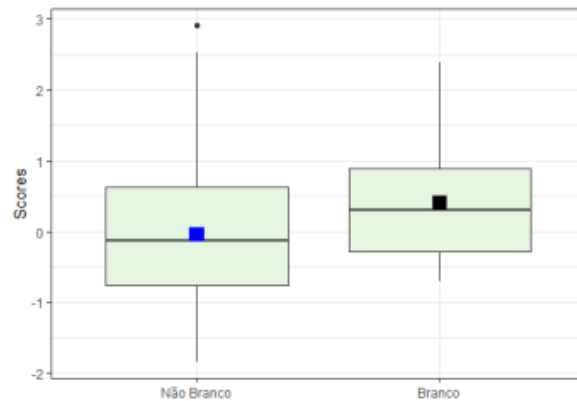
Variáveis	Efeito
Intercepto	0,1198
Grupo Controle	0,0352
Raça: Branca	0,4315
EM Completo ou Incompleto	-0,2304
Superior Completo	-0,4028
R <sup>2</sup>	0,0302

Esse modelo explica apenas 3,02% da variabilidade dos scores calculados, de forma a avaliar o efeito de cada característica definida. Esse número é considerado bem baixo e indica que através dos dados do questionário, as variáveis sociodemográficas não tem poder explicativo em relação à variabilidade do Bem Estar Pessoal.

A partir da Tabela 10, pode-se afirmar que, se um indivíduo é Branco, possui apenas o Ensino Fundamental completo e participou do programa, ele terá uma pontuação de Bem Estar média estimada de 0,1198, pontuação representada pelo Intercepto. Esse indivíduo é considerado como indivíduo base, de forma que, a mudança de uma dessas características, mantendo as demais, gera uma mudança na pontuação média referente ao efeito associado a escolha realizada. É através disso que é possível realizar inúmeras comparações e ter algumas interpretações.

Como o modelo acaba explicando apenas 5% da variabilidade do score, as conclusões a serem retiradas acabam não sendo muito significativas. Com isso, foram realizadas algumas análises no ponto de vista univariado através das variáveis sociodemográficas cujos impactos foram considerados significativos e as diferenças entre os grupos foram significativas.

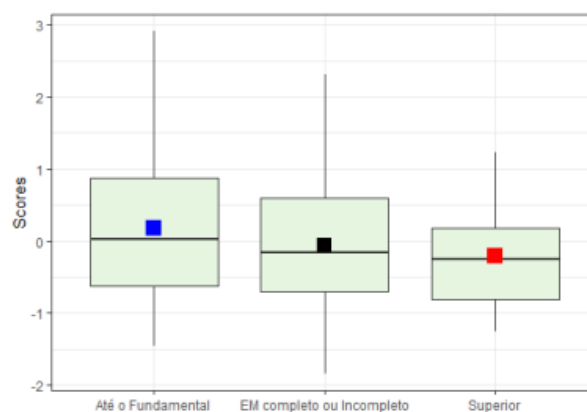
## Bem-estar Pessoal – Não Branco x Branco



Aplicado o Teste T, com p-valor = 0,017, tem-se, a um nível de significância de 5%, que as médias das pontuações de Bem Estar Pessoal **são diferentes** entre indivíduos de Raça Branca e Não Branca.

Figura 16: Boxplot - Bem Estar Pessoal por Raça

## Bem-estar Pessoal – Escolaridades Diferentes



Aplicado o Teste ANOVA, com p-valor=0,13, tem-se, a um nível de significância de 5%, que as médias das pontuações de Bem Estar Pessoal **são iguais** entre as pessoas de escolaridades diferentes.

Figura 17: Boxplot - Bem Estar Profissional por Escolaridade

Mesmo que exista outras variáveis que possui impactos significantes no modelo, apenas essas duas apresentadas resultam em diferenças estatisticamente significantes dentro dos grupos.

## Conclusão

A presente consultoria consistiu em uma análise sobre os dados de indivíduos que participaram do projeto Jovens Construtores nas comunidades Borel e Mangueira. A análise de consistência da base indicou que 0,9% dos dados possuíam alguma característica faltante. Como essa porcentagem não foi considerada significativa, esses dados foram descartados. A base continha algumas características descritivas de cada indivíduo e suas respostas do questionário aplicado a eles ao final do projeto. Esse questionário estava baseado em escala Likert, ou seja, as respostas dos participantes especificavam seu nível de concordância ou de satisfação com determinadas afirmações. Dessa forma, os objetivos desta consultoria eram os seguintes: verificar o que as respostas do questionário poderiam estar medindo, e, em seguida, verificar como isso se relaciona com as características descritivas de cada participante.

Diante dos resultados obtidos através da análise fatorial, tem-se que as respostas dos questionários estavam medindo 4 fatores distintos, sendo eles: Autoestima, Satisfação Profissional, Incapacidade e Bem-estar Pessoal. Essa análise fatorial foi construída a partir de 13 das 17 questões presentes no questionário. Utilizou-se o método de mínimos quadrados generalizados e a rotação oblíqua promax para a estimação desses fatores, que juntos explicam 44% da variabilidade dos dados.

Após a construção dos scores desses fatores, avaliou-se o impacto conjunto das variáveis sociodemográficas nos scores, onde foram criados 4 modelos lineares múltiplos, sendo cada um avaliando os impactos referentes aos respectivos scores. Porém os quatro modelos, mesmo seguindo seus pressupostos, não apresentaram bons resultados, com  $R^2$  entre 3 e 8% apenas.

Com o objetivo de tirar algumas conclusões, foi realizado alguns testes com as variáveis que apresentaram impactos significativos. Com isso se obteve as seguintes conclusões: as médias de autoestima são diferentes entre indivíduos com escolaridades diferentes, as médias de satisfação profissional são diferentes entre indivíduos de raças diferentes e entre indivíduos que residem em locais diferentes e as médias de Bem-estar Pessoal são diferentes entre indivíduos de Raças diferentes.

**OBS.: os códigos estão em uma versão resumida, ideal apenas para uma abordagem inicial. Qualquer dúvida ou sugestão é bem-vinda.**