

# Acesso ao Ensino Superior no Rio de Janeiro: Análise dos Indicadores Sociais e de Infraestrutura

*Autor: João Pedro Leyssieux Campanella*

*Acesse o [Github](#) do autor para visualizar o código usado para este relatório.*

Ao longo das últimas décadas, o Brasil passou por grandes avanços na política educacional, desenvolvimento acadêmico e disseminação do conhecimento. Contudo, restam diversas lacunas e desafios a serem enfrentados. O país possui uma vasta rede pública de ensino, profissionais capacitados e grande potencial intelectual, mas apenas 20% da população de 25 a 34 anos possui diploma de ensino superior. Outro ponto de destaque é a desigualdade de renda: segundo o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 40% dos estudantes universitários fazem parte da parcela dos 20% mais ricos da população.

Levando esse recorte para a segunda maior cidade do país, o Rio de Janeiro, a taxa de diplomados é ligeiramente superior à média nacional, chegando a 32%, de acordo com uma pesquisa do Instituto Pereira Passos (IPP) e da Prefeitura entre os anos de 2016 a 2020. Entretanto, mesmo que tal resultado pareça positivo, índices relacionados à segurança, saúde e bem-estar vem dificultando maiores progressos, uma vez que a flutuação destes ao longo dos anos representam desafios para os estudantes cariocas.

Diante deste cenário, o presente relatório busca compreender e explorar as principais causas para os baixos índices de acesso ao ensino superior na cidade do Rio de Janeiro.

## 1. INTRODUÇÃO

Este trabalho tem como objetivo verificar a relação dos principais índices sociais e de infraestrutura do Rio de Janeiro com o acesso da população carioca ao ensino superior. Para isso, parte-se da hipótese de que esses fatores são diretamente proporcionais entre si, logo que ambientes com maior bem-estar social e com mais oportunidades tendem a oferecer menos barreiras ao ingresso na universidade..

## 2. FERRAMENTAS UTILIZADAS

No desenvolvimento desta análise, para melhor visualização e manipulação dos dados foram utilizadas as seguintes ferramentas:

- Python
- Jupyter Notebook
- Numpy
- Matplotlib e Seaborn
- Pandas
- Microsoft Excel

## 3. ANÁLISE

### A. Entendendo o Dataset

Neste relatório foram utilizados os dados do [Índice de Progresso Social no Rio de Janeiro](#), fornecidos pela Prefeitura e Instituto Pereira Passos. Nesta base de dados estão presentes os seguintes indicadores:

- Dimensão 1: Necessidades Humanas Básicas
  - Nutrição e Cuidados Médicos Básicos
  - Água e Saneamento
  - Moradia
  - Segurança Pessoal
- Dimensão 2: Fundamentos do Bem-Estar
  - Acesso a Conhecimento Básico
  - Acesso a Informação e Comunicação
  - Saúde e Bem-Estar
  - Qualidade do Meio Ambiente
- Dimensão 3: Oportunidades
  - Direitos Individuais
  - Liberdade Individual e de Escolha
  - Tolerância e Inclusão
  - Acesso ao Ensino Superior

*obs1: A nota de cada dimensão está no intervalo de  $0 \leq n \leq 100$  é calculada pela média simples das notas de seus componentes, onde o número de componentes é sempre igual a 4.*

*obs2: A nota do Índice de Progresso Social (IPS) é calculada pela média simples das notas das dimensões, onde o número de dimensões é sempre igual a 3).*

O dataset utilizado possui 18 colunas originalmente, sendo elas:

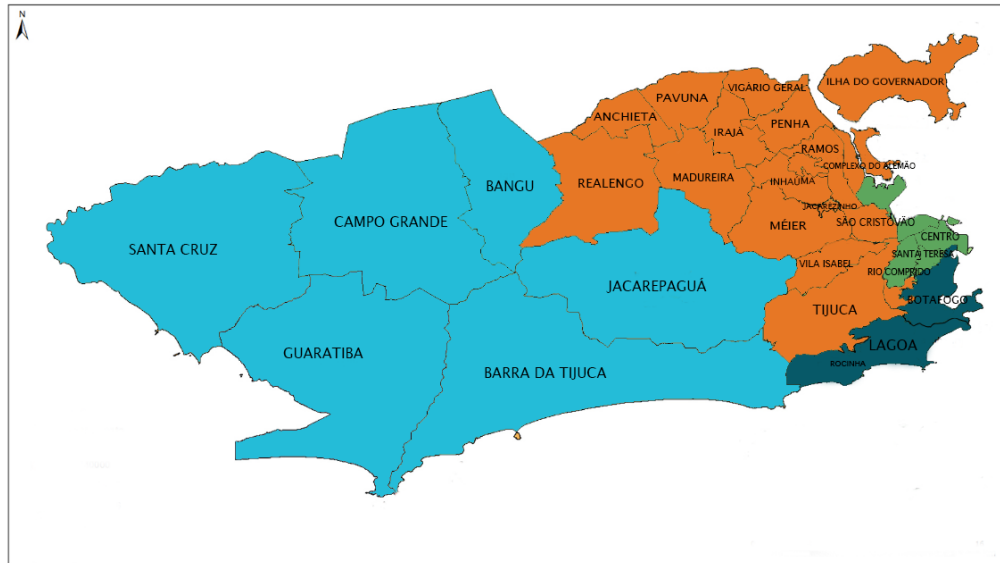
- 1 coluna 'Ano' (*int*)
- 1 coluna 'Região Administrativa' (*object*)
- 1 coluna 'IPS Geral' (*float*)
- 12 colunas 'Componentes' (*float*)
- 3 colunas 'Dimensões' (*float*)

Além disso, foi adicionada manualmente a coluna **Macrorregião** para fins de comparação nos gráficos posteriormente apresentados neste relatório.

Os índices desta pesquisa foram analisados para os anos de 2016, 2018 e 2020. Todas as regiões administrativas tem seus dados preenchidos para estes 3 anos, sem exceções. São **99 linhas** de dados ao todo, considerando todas as regiões administrativas e a cidade do Rio de Janeiro.

## B. Manipulando os Dados

Utilizando o Microsoft Excel, as regiões administrativas da cidade do Rio de Janeiro foram atribuídas às 4 macrorregiões, conforme a imagem abaixo:



**Fig. S1.** Mapa das Macrorregiões do Rio de Janeiro

### Legenda

- Zona Oeste (Azul Claro)
- Zona Sul (Azul Escuro)
- Zona Norte (Laranja)
- Centro (Verde)

Para evitar redundância na aferição dos dados, foram removidas as colunas que representam as **Dimensões**, que conforme explicitado anteriormente, são as média simples dos valores de seus respectivos componentes. Além disso, a coluna '*Nutrição Cuidados Médicos Básicos*' foi renomeada para '*Cuidados Médicos*' para fins estéticos nos gráficos, seguindo o trecho de código abaixo:

```
df_rj = pd.read_excel('planilha_rj.xlsx')

df_rj.drop(columns=['necessidades_humanas_basicas_notas_dimensao', \
                    'fundamentos_bem_estar_notas_dimensao', \
                    'oportunidades_notas_dimensao'], inplace=True)

df_rj.rename(columns={'nutricao_cuidados_medicos_basicos': \
                    'cuidados_medicos'}, inplace=True)
```

## C. Analisando os Dados

Inicialmente, foi verificado o valor médio do IPS na cidade do Rio de Janeiro nos anos de 2016, 2018 e 2020. Estes valores servirão de parâmetro de comparação a seguir.

```
df_rj[df_rj['macro_regiao'] == 'rj']
```

	ano	regiao_administrativa	macro_regiao	ips_geral
75	2016	Rio De Janeiro	rj	60.77
76	2018	Rio De Janeiro	rj	60.85
77	2020	Rio De Janeiro	rj	60.79

**Fig. S2.** IPS médio no Rio de Janeiro

Em seguida, foi *plotado* um gráfico de barras para comparação do acesso ao ensino superior entre as macrorregiões e a média da cidade

```
df_rj_agrupado = df_rj.groupby('macro_regiao', as_index=False)['\
    acesso_educacao_superior'].mean()

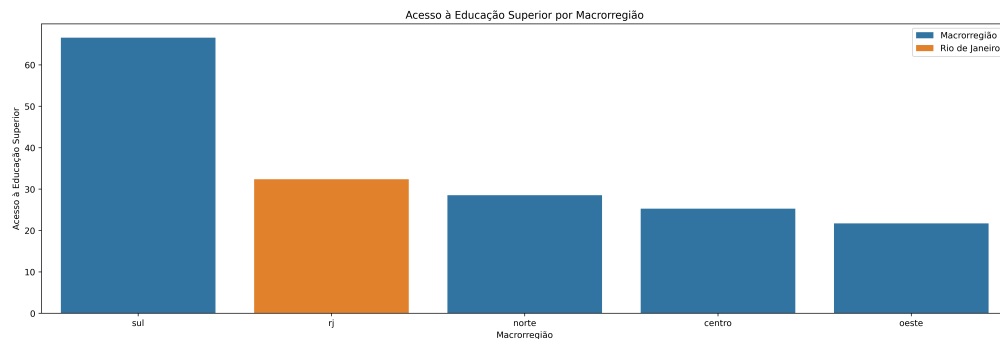
df_rj_agrupado['tipo'] = df_rj_agrupado['macro_regiao'].apply(\
    lambda x: 'Rio_de_Janeiro' if x == 'rj' else 'Macrorregiao')

df_rj_agrupado = df_rj_agrupado.sort_values(by='\
    acesso_educacao_superior', ascending=False)

plt.figure(figsize=(20, 6), dpi=400)

sns.barplot(x='macro_regiao', y='acesso_educacao_superior', hue='\
    tipo', data=df_rj_agrupado)

plt.title('Acesso_a_Educacao_Superior_por_Macrorregiao')
plt.ylabel('Acesso_a_Educacao_Superior')
plt.xlabel('Macrorregiao')
plt.legend()
plt.show()
```



**Fig. S3.** Acesso ao Ensino Superior por Macrorregião

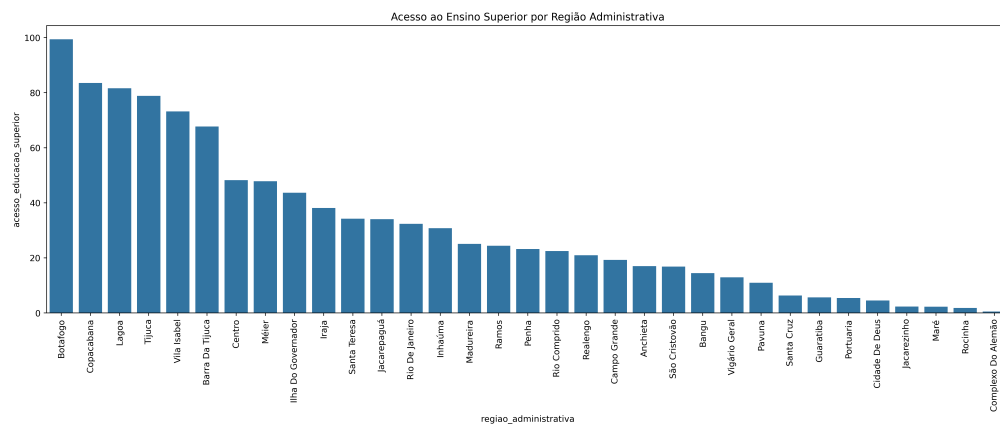
*obs: a média da cidade está sendo utilizada como parâmetro comparativo, estando representada pela coluna laranja*

A partir desta análise gráfica, fica evidente como a média Zona Sul está a frente (em termos educacionais) do resto da cidade. Inicialmente, pode-se partir de hipótese que isso está diretamente relacionado com o fato desta região ser a mais rica do Rio de Janeiro, o que certamente denota um padrão de vida elevado de seus habitantes.

Após isto, foi feito um gráfico de barras para melhor comparação das regiões administrativas de forma individual:

```
df_rj_decrescente = df_rj.sort_values(by='acesso_educacao_superior', ascending= False)

plt.figure(figsize=(20,6), dpi=400)
plt.xticks(rotation=90)
sns.barplot(x='regiao_administrativa', y='acesso_educacao_superior', data=df_rj_decrescente)
plt.title('Acesso_ao_Ensino_Superior_por_Regiao_Administrativa')
plt.show()
```



**Fig. S4.** Acesso ao Ensino Superior por Região Administrativa

Em primeiro lugar, encontra-se a região administrativa **'Botafogo'** - bairro frequentado pela classe média carioca e conhecido pela sua gastronomia e cultura - atingindo quase a nota máxima do índice em questão. Em último lugar, encontra-se a região administrativa **'Complexo do Alemão'** - famosa por abrigar um dos maiores conjuntos de favelas da Zona Norte da cidade.

É relevante perceber que, de maneira geral, as regiões suburbanas e favelizadas ocupam posições piores nesta análise. Pode-se inferir que isso deriva fatores como renda, segurança pública e infraestrutura urbana. Isto será verificado posteriormente neste relatório.

Outro ponto de atenção neste gráfico seria a região administrativa da **'Barra da Tijuca'**, conhecida por ser uma área relativamente nova e com excelentes índices de desenvolvimento humano e de condições de vida. Apesar disso, no gráfico ela ocupa a 6ª posição, com uma nota consideravelmente distante do primeiro colocado nesta análise. Isto pode representar que os componentes do IPS podem não altamente correlatos, ou seja, que valores melhores de qualidade de vida não implicam com tanta intensidade no acesso às universidades.

Para entender isto melhor, foi criada uma Matriz de Correlação, representada graficamente por um Mapa de Calor.

**Matriz de Correlação:** é uma tabela que apresenta a relação entre diferentes variáveis de um conjunto de dados. Cada célula da matriz contém um valor que representa a força e a direção do relacionamento linear entre duas variáveis. Esses valores variam de -1 a 1, onde:

- 1 indica uma correlação positiva perfeita (quando uma variável aumenta, a outra também aumenta).
- -1 indica uma correlação negativa perfeita (quando uma variável aumenta, a outra diminui).
- 0 indica que não há correlação linear entre as variáveis.

Na diagonal principal da matriz, o valor será sempre 1, já que uma variável está perfeitamente correlacionada consigo mesma.

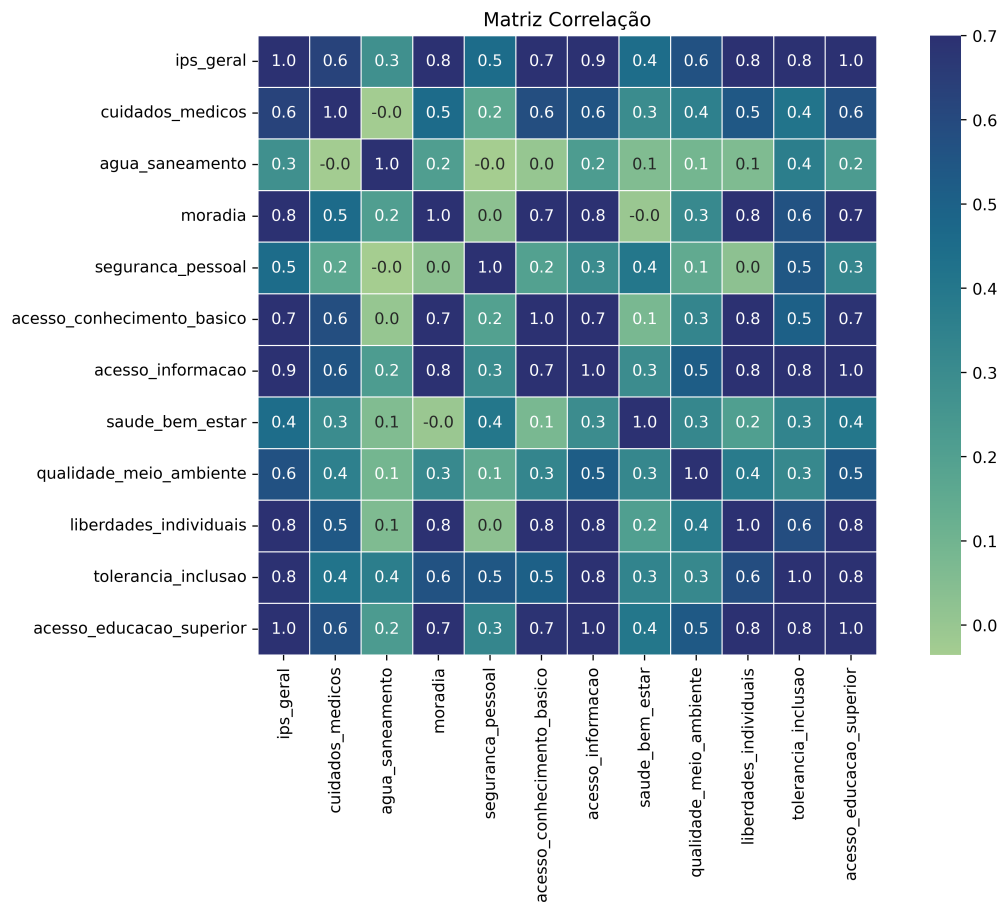
Para criar a matriz, foi usado o código:

```
df_rj_correlacao = df_rj.drop(columns=['ano', 'macro_regiao', '\
regiao_administrativa', 'direitos_individuais'])

matriz_correlacao = df_rj_correlacao.corr()

plt.figure(figsize=(12, 7), dpi=500)
sns.heatmap(matriz_correlacao, linewidth=.5, annot=True, vmax\
=.7, fmt=".1f", cmap="crest", square=True)
```

- *obs1: Coluna ano e direitos individuais foram removidas pois não apresentaram correlação relevante.*
- *obs2: Região administrativa é representada por strings, portanto não é possível correlacionar a princípio.*



**Fig. S5.** Mapa de Calor - Matriz Correlação

*obs: As células escuras representam as maiores correlações. Foi definido que o tom mais escuro de azul aparecerá quando a correlação for maior ou igual a 0,7.*

A partir do mapa de calor, é possível perceber uma alta correlação entre as notas dos indicadores e do acesso ao ensino superior, o que prova que a hipótese previamente estabelecida na introdução deste relatório está de acordo com o que os dados informam. Com isso, a análise agora será direcionada ao caso dos *Outliers*, ou seja, pontos de dados que se desviam significativamente da tendência dos valores verificados.

Buscando entender as possíveis falhas da hipótese estabelecida neste relatório, a análise agora será restringida à região administrativa da Barra da Tijuca. Como mencionado anteriormente, apesar dos altos índices de desenvolvimento humano e de qualidade de vida, a nota de acesso ao ensino superior é abaixo da esperada. Iremos verificar quais são as possíveis causas disso por meio da análise dos dados.

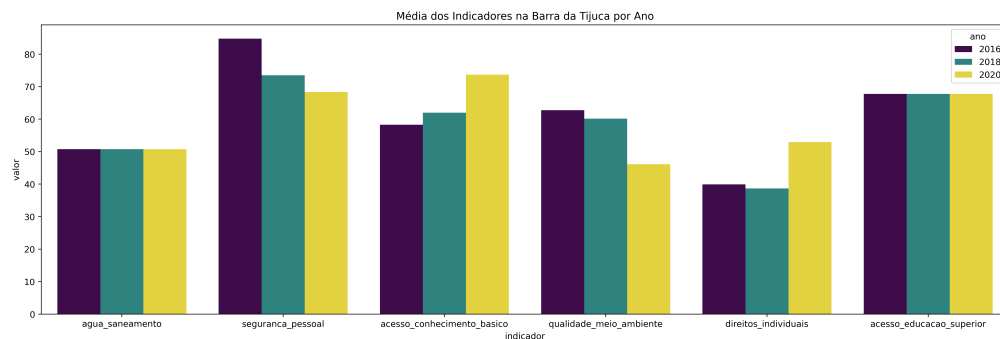
```
df_barra = df_rj[df_rj['regiao_administrativa'] == 'Barra_Da_\nTijuca'][['ano', 'agua_saneamento', 'seguranca_pessoal', '\nacesso_conhecimento_basico', 'qualidade_meio_ambiente', '\ndireitos_individuais', 'acesso_educacao_superior']]

df_barra_melt = df_barra.melt(id_vars='ano', var_name='indicador'\n, value_name='valor')

plt.figure(figsize=(20, 6), dpi=400)

sns.barplot(data=df_barra_melt, x='indicador', y='valor', hue='\nano', palette='viridis')

plt.title('Media_dos_Indicadores_na_Barra_da_Tijuca_por_Ano')
plt.show()
```



**Menções Relevantes:**

- 7

e 2023 (o que será mais abordado no próximo tópico). Na Barra da Tijuca, pode-se observar que os índices de segurança pessoal vêm piorando gradativamente. No entanto, esses valores estão mais relacionados ao crescente número de roubos de rua, o que não se prova ser um impeditivo no acesso à educação básica, por exemplo.

- Acesso Conhecimento Básico:

- As notas deste item apresentaram melhora ao longo dos três anos medidos na Barra da Tijuca. Ele é altamente correlacionado com o acesso ao ensino superior, algo que também pode-se inferir de forma indutiva. Apesar da crescente neste componente, não é percebida uma melhora no acesso ao ensino superior no mesmo período de aferição.

- Direitos Individuais

- Este item apresentou uma considerável melhora no ano de 2020. Isso se deve ao fato de que, com o tempo, mais obras públicas de infraestrutura e mobilidade urbana foram sendo realizadas no bairro. Contudo, também não percebe-se influência disto no acesso ao ensino superior para a região administrativa em questão. Como dito anteriormente neste tópico, a coluna dos Direitos Individuais havia sido removida por haver baixíssima correlação com o que está sendo buscado.

Apesar da grande mudança em alguns dos índices mencionados acima, pode-se perceber que não houve praticamente nenhuma alteração no resultado final do acesso ao ensino superior nos 3 anos medidos para o bairro da Barra da Tijuca.

Isto aparece como um contraponto da hipótese previamente estabelecida neste relatório, logo que acreditava-se que todos os componentes do IPS estariam fortemente correlacionados com o acesso à universidade para a cidade do Rio de Janeiro.

### **C.2. Caso Cidade de Deus, Rocinha, Campo Grande e Centro**

Inicialmente, foi criado um gráfico de **regressão linear** para visualizar a relação entre segurança pública e acesso ao ensino superior.

**Regressão Linear:** Regressão linear é uma técnica que encontra a relação entre duas variáveis, ajustando uma linha reta aos dados. Essa linha é usada para entender como uma variável muda em resposta à outra e para fazer previsões baseadas nessa relação.

```
def regressao_linear(eixo_x, eixo_y, largura, altura):
    plt.figure(figsize=(largura, altura), dpi=400)
    sns.regplot(x = eixo_x, y = eixo_y);

    a, b = np.polyfit(eixo_x, eixo_y, 1)

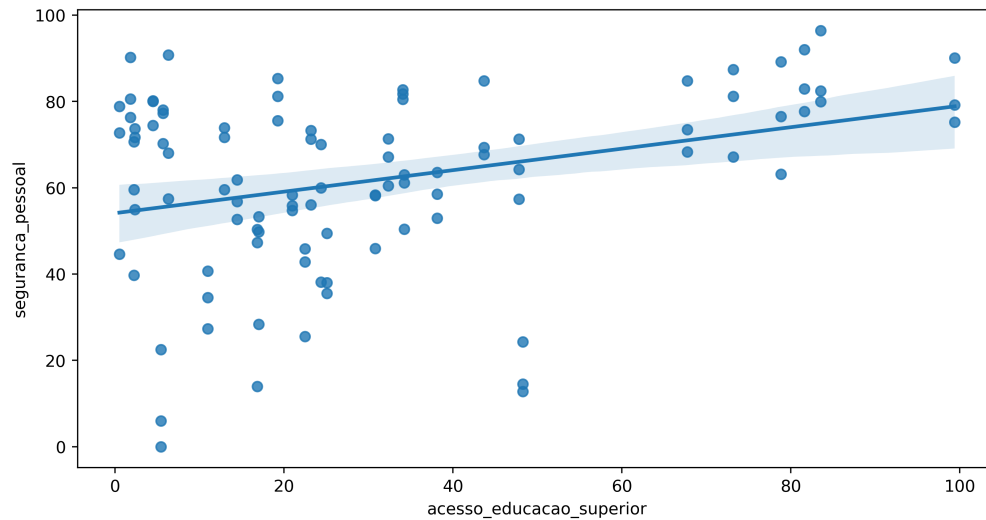
    print(f'Coeficiente_Angular:_{a:.2f}')
    print(f'Coeficiente_Linear:_{b:.2f}')

x = df_rj['acesso_educacao_superior']
y = df_rj['seguranca_pessoal']

regressao_linear(x, y, 10, 5)

plt.show()
```





**Fig. S7.** Relação entre Acesso ao Ensino Superior e Segurança Pessoal

Apesar da inclinação positiva do gráfico, pode-se perceber que existem muitos dados que estão distantes da linha de tendência. Tomaram-se 2 casos para análise:

1. Segurança Pessoal > 80 & Acesso a Educação Superior < 20
2. Segurança Pessoal < 20 & Acesso a Educação Superior > 40

Que foram obtidas a partir dos código abaixo:

```
df_teste = df_rj[(df_rj['seguranca_pessoal']>80) & (df_rj['\
acesso_educacao_superior']<20)][['ano', '\
regiao_administrativa']]
df_teste
```

Que obteve:

	ano	regiao_administrativa
63	2016	Campo Grande
65	2020	Campo Grande
67	2018	Cidade De Deus
68	2020	Cidade De Deus
77	2020	Santa Cruz
96	2016	Rocinha
98	2020	Rocinha

**Fig. S8.** Segurança Pessoal > 80 & Acesso a Educação Superior < 20

E do código

```
df_teste = df_rj[(df_rj['seguranca_pessoal']<20) & (df_rj['\
acesso_educacao_superior']>40)][['ano', '\
regiao_administrativa']]
df_teste
```

Que obteve:

	ano	regiao_administrativa
1	2018	Centro
2	2020	Centro

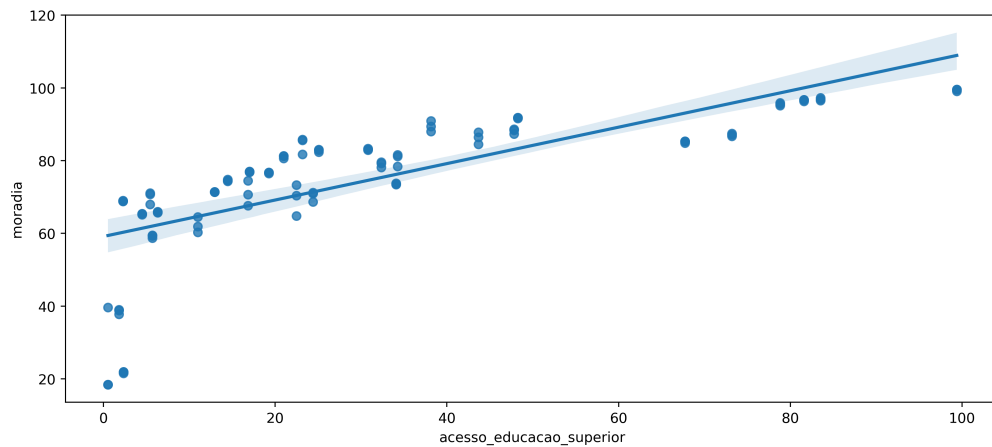
**Fig. S9.** Segurança Pessoal < 20 & Acesso a Educação Superior > 40

As regiões administrativas que se encontram nas tabelas acima são consideradas *outliers*, por divergirem da curva de tendência da Regressão Linear.

Os *outliers* irão aparecer em praticamente todas as comparações que forem feitas com o **Acesso ao Ensino Superior**, como no exemplo abaixo, em que foi utilizada a categoria '**Moradia**' (que possui correlação de 0,7).

```
x = df_rj[ 'acesso_educacao_superior' ]
y = df_rj[ 'moradia' ]

regressao_linear(x, y, 12, 5)
```



**Fig. S10.** Moradia x Acesso a Educação Superior

Podem ser observados os *outliers* para valores de **moradia** < 50 e de **acesso educação superior** < 10.

#### 4. CONCLUSÃO

Na introdução deste relatório, foi levantada a hipótese de que todos os componentes do IPS deveriam possuir alta correlação com o Acesso ao Ensino Superior no Rio de Janeiro. Contudo, ao longo das análises, fica claro que há grandes exceções em todas as comparações dos componentes. Além disso, percebe-se que melhorias em índices como segurança individual, moradia e acesso ao ensino básico, por exemplo, não são fatores que influenciam diretamente no acesso ao ensino superior no curto prazo. No fundo, será a soma desses parâmetros com questões culturais e históricas que influenciará as escolhas individuais de cada cidadão em cursar uma universidade..