

Universidade Federal do Estado do Rio de Janeiro – UNIRIO

Nome do curso: Biblioteconomia - Bacharelado

Nome da disciplina: Estatística Ano – 2021 - Semestre 01

Turno: Noturno

Nome do aluno: Claudia Mesquita Vera Cruz Matrícula: 20181332054

Título da atividade: Relatório

1. Sumário executivo

Este relatório tem o objetivo de apresentar dados Estatístico referente a amostra de educação, e estabelecer o resultado das análises, com apoio para a elaboração foi utilizado o sistema *R-commander*, elaborado os gráficos para a visualização das tendências e analisar comparativos, assim de acordo com o referencial teórico, foi comparado os dados e as variáveis. Concluindo que o nível da educação é influenciado pela região, IDH.

2. Introdução.

A Estatística pode ser dividida em duas áreas, a indutiva e a descritiva.

As variáveis de interesse no experimento são chamadas de variáveis de resposta ou variáveis dependentes. Outras variáveis no experimento que afetam a resposta e podem ser definidas ou medidas pelo experimentador são chamadas de variáveis preditoras, variáveis explicativas ou variáveis independentes. Por exemplo, você pode determinar o tempo de cozimento recomendado para uma mistura para bolo ou fornecer instruções sobre como cuidar de novas plantas híbridas.

Com base nessas definições o trabalho buscou comparar dados da educação de acordo com a amostra obtida.

3. Objetivos.

3.1 Objetivo Principal

- Fazer um diagnóstico sobre o banco de dados

3.2 Objetivos secundários

- Classificar corretamente as variáveis;
- Entender a importância de uma amostra aleatória;

4. Marco Teórico

Segundo Pertenelli (2013) referente a histograma, para alguns conjuntos de dados o número de valores distintos da variável em estudo é muito grande para serem considerados os tipos de apresentação gráfica apresentados acima. Em tais casos seria útil dividir os valores em grupos, ou intervalos de classe, e então plotar o número de valores dos dados correspondentes a cada intervalo de classe. Existem várias fórmulas para se estabelecer o número de classes, porém qualquer número de classes poderia ser utilizado, baseando-se nas seguintes observações:

- Não escolher muito poucas classes, para evitar perda de informação sobre os dados;
- Não escolher muitas classes, o que poderia fazer com que as frequências referentes a cada classe fossem tão pequenas a ponto de atrapalhar o discernimento de algum padrão de distribuição para a variável em estudo.

O que se faz na prática é tentar variados números de classes e verificar, com a ajuda de um computador, o número ideal para os dados em questão. Além disso, comumente usamos intervalos de classe de iguais amplitudes.

As variáveis estatísticas são recursos que permitem valores diferentes (número ou modal), um para cada unidade estatística. Eles são baseados na qualidade e não podem ser medidos por números. Quando o valor possível de uma variável é uma categoria, é qualitativo.

E para analisar tantos indicadores, probabilidades, variáveis foi criado o programa R. O R fornece uma ampla variedade de funções de visualização de dados, de recursos nativos mais simples a pacotes de recursos poderosos, torna o R a melhor ferramenta para criar visualizações complexas.

O R oferece uma variedade enorme de recursos de visualização de dados, desde os recursos nativos mais simples até pacotes com recursos poderosos, que tornam o R a melhor ferramenta para criação de visualizações sofisticadas, complexas (ALCOFARADO: LEVY: 2017).

5. Metodologia.

Com a finalidade de analisar e interpretar dados, foi feita uma pesquisa bibliográfica. Por meio da pesquisa bibliográfica foi possível avaliar os conceitos e as novas maneiras de se administrar pessoas. Foi realizado um método de pesquisa de natureza quantitativa, uma primeira visão do Programa R, uma análise estatística e uma análise gráfica do banco de dados.

5.1 Descrição da base de dados:

A base de dados aplicada neste trabalho foi uma tabela descritiva, a fonte de dados foi IBGE 2010, sendo as variáveis quantitativa discreta, com cerca de 2668 dados sendo eles:

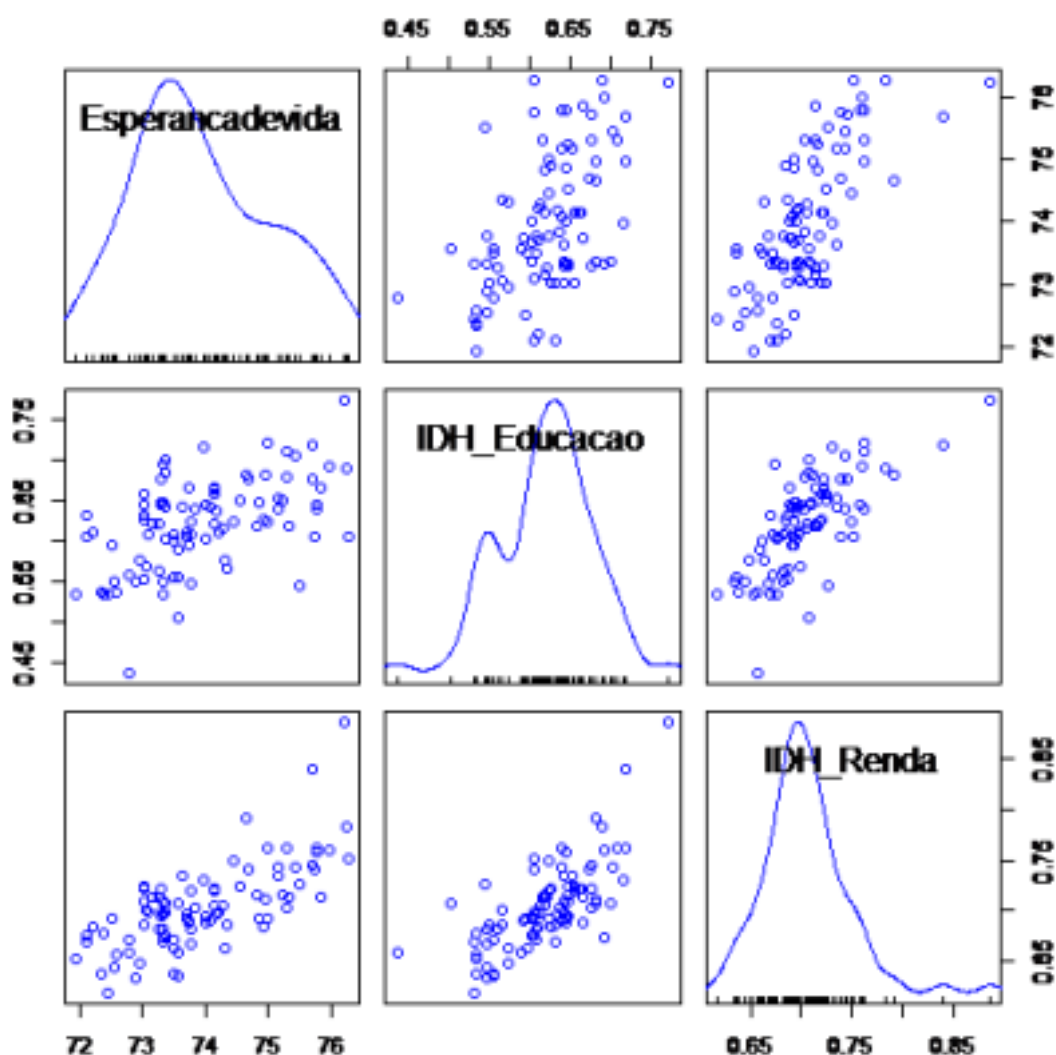
Indicador	Descrição
COD IBGE	COD IBGE
COD IBGE2	COD IBGE2
Gini	Índice de Gini (2010)
Agua	% da população em domicílios com água encanada (2010)
Banheiro	% da população em domicílios com banheiro e água encanada (2010)
Lixo	% da população em domicílios com coleta de lixo (2010)
Energia	% da população em domicílios com energia elétrica (2010)
Densidade	% da população em domicílios com densidade > 2 (2010)
Esperancadevida	Esperança de vida ao nascer (2010)
Mortalidade_infantil	Mortalidade infantil (2010)

Prob_sobrevivencia	Probabilidade de sobrevivência até 40 anos (2010)
IDH	IDHM (2010)
IDH_Renda	IDHM Renda (2010)
IDH_Longevidade	IDHM Longevidade (2010)
IDH_Educacao	IDHM Educação (2010)
Probab_sobrev60	Probabilidade de sobrevivência até 60 anos (2010)
TFT	Taxa de fecundidade total (2010)
Taxa_envelhecimento	Taxa de envelhecimento (2010)
Taxa_analfabetismo	Taxa de analfabetismo - 25 anos ou mais (2010)
frequencia_liquida_EM	Taxa de frequência líquida ao médio (2010)
Expectativa_anos_de_estudo	Expectativa de anos de estudo (2010)
frequencia_liquida_Superior	Taxa de frequência líquida ao superior (2010)
perc_com_2_anos_de_de_atraso	% de 6 a 17 anos no básico com 2 anos ou mais de atraso (2010)
Renda_per_capita	Renda per capita (2010)
Renda_per_capita_nula	Renda per capita, exceto renda nula (2010)
Perc_pobres	% de pobres (2010)
Perc_extremamente_pobres	% de crianças extremamente pobres (2010)
Populacao	População total (2010)
ISS	Receita tributária - ISS - municipal - R\$ - Ministério da Fazenda - Secretaria do Tesouro Nacional - RISSM

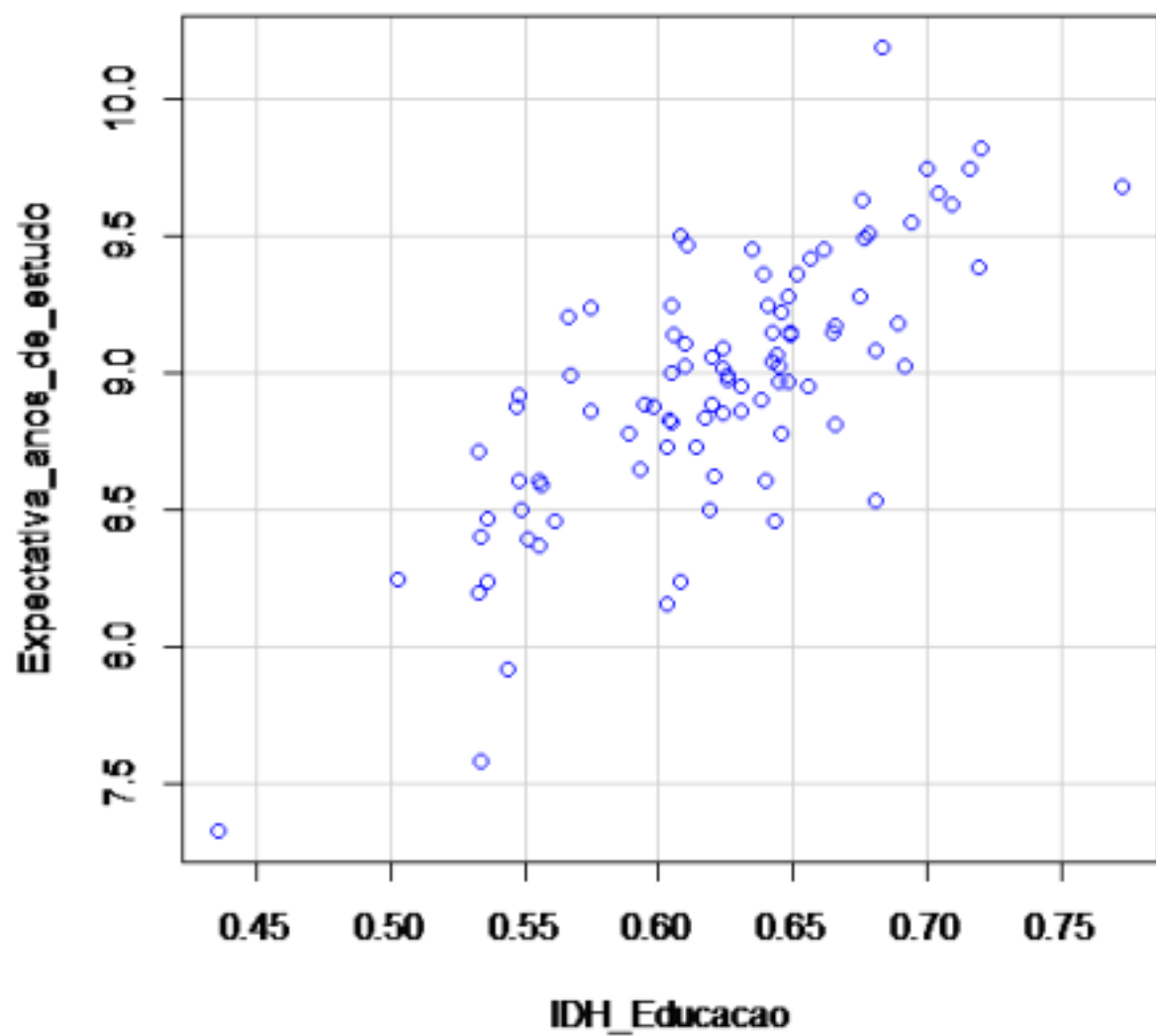
6. Análise de Resultados

A elaboração de gráficos no R é feito através de forma de script, com a descrição de acordo com a necessidade da análise. Abaixo segue gráficos comparativos, elaborados a partir do programa R, utilizando informações da base de dados, relacionados ao IDH Educação X Esperança de vida X IDH Renda. : A amostra provém de uma população Normal

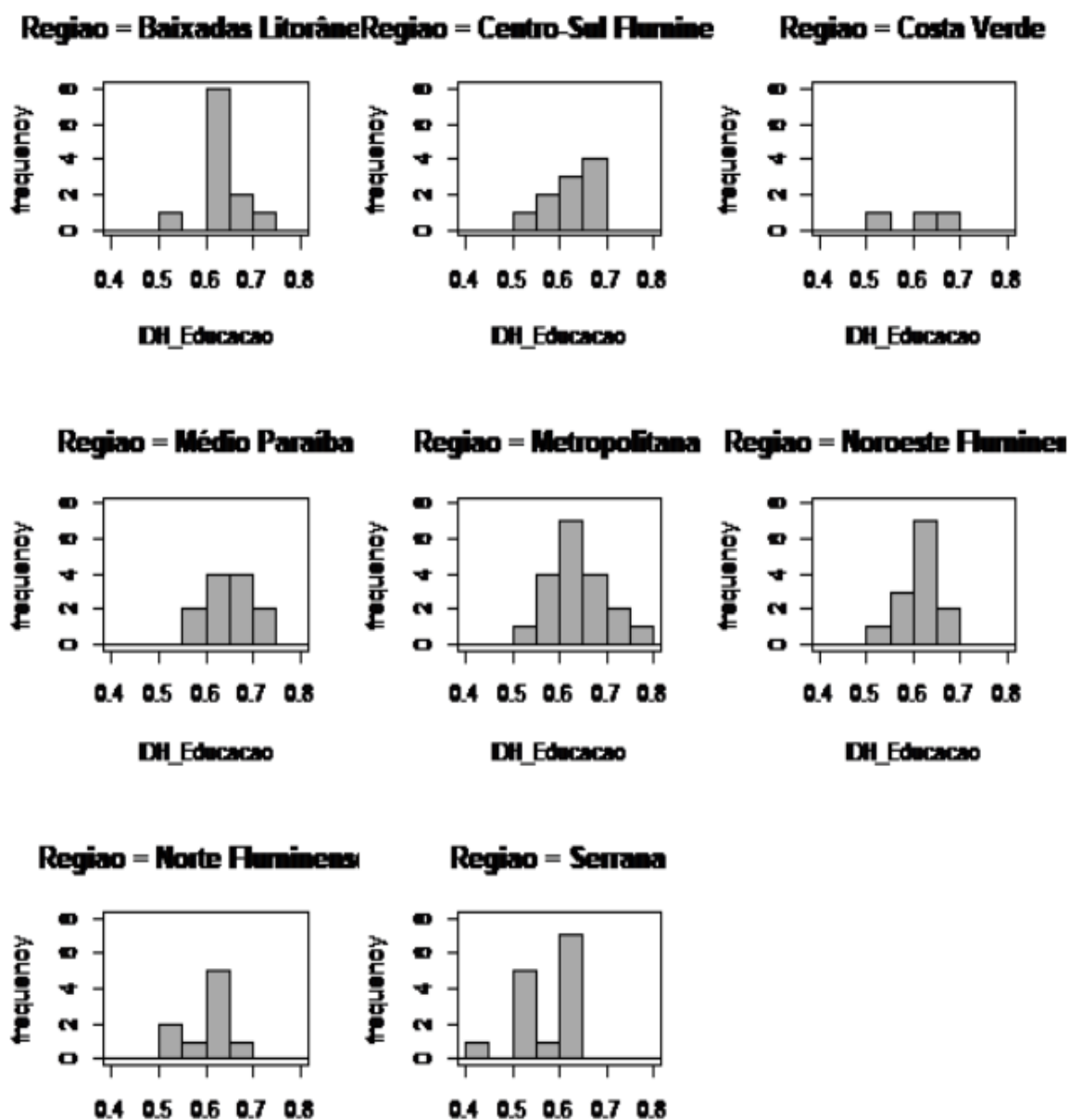
Matriz de Dispersão com Histograma



Expectativa de vida IDH Educação



Média da Frequência escolar por Região X IDH Educação



7. Discussão

Ao realizar a pesquisa bibliográfica e utilizar o programa R, analisando os resultados, é possível concluir que os resultados obtidos afirmam a teoria. Sendo importantes as funções do programa R, assim como antecipadamente conhecer o programa, mas também ter embasamento teórico para saber analisar os gráficos, as variantes.

É possível notar que não apresenta outlier na educação e na renda, no nível máximo, porém conhecendo a realidade brasileira é possível concluir que é devido estes patamares serem acessíveis a menor parcela da população; há concentração nos níveis média, mostrando que a maioria da população na pesquisa corresponde a indivíduos com IDH de renda e educação médio; há correlação do nível de escolaridade com o IDH renda, quanto maior o IDH melhor os índices da educação e possui simetria.

8. Conclusão

De acordo com a análise da base de dados regiões com maior IDH tem o IDH em educação maior, sendo a frequência e anos de vida de estudo maiores, consequentemente, maior o número de pessoas com nível superior e alfabetizados.

Mostra que o meio tem um impacto direto na vida dos estudantes, mas que mesmo assim há chances para os indivíduos conquistarem o ensino superior, sabe-se que o investimento na educação é preciso, para que a acessibilidade e equidade de fato alcance aos que querem estudar. E neste trabalho é evidenciado que de acordo com o meio o ensino é impactado, sendo importante o estudo e a estatística desse tema para analisar quais grupos e regiões, as ações governamentais devem atuar. Ressaltando que não só para a educação é aplicável e necessário esse método.

Também se ressalta que o programa R, tem extrema importância para estudos que envolvem a estatística, pois facilita a importação e análise de dados, assim como a elaboração de gráficos, estudo de variantes.

Referências

Mello, P. M.; Peternelli, L. A. **Conhecendo o R - Uma Visão mais que estatística**. 1ª ed. 2013.

ALCOFARADO, L. F.; LEVY, A. **Visualização de dados com o software R**. 1ª ed. 2017. Disponível em: <www.estadisticacomr.uff.br>. Acesso em: 08 Set 2021.