****

ESTUDO DE CASO A

Disciplina: Estatística Aplicada

Professor: Steven Ross

Alunos: Jéssica Nobre e Victor Lupinacci

2017.2

**Introdução**

O R é uma linguagem de computação estatística e gráfica, através do R é possível obter uma grande variedade estatística como modelagem linear e não-linear, testes estatísticos clássicos, análise de séries temporais, classificação entre outros, R é também altamente expansível com o uso dos pacotes, que são bibliotecas para funções específicas ou áreas de estudo específicas. Uma das possíveis aplicações desse software é de analises de caracterização de populações.

A caracterização de uma população se dá a partir de estudos estatísticos, que podem ser feitos por meio de pesquisas de opinião. Essas pesquisas visam um determinado grupo de pessoas, buscando analisar estatisticamente os costumes, situação econômica e preferências. Importantes variáveis analisadas em caracterização de população, são: idade, ocupação, salário, escolaridade, anos de estudo, situação conjugal, setor, etnia, gênero, horas de trabalho, local e IMC. Esses tipos de variáveis tornam-se essenciais para esse tipo de pesquisa, uma vez que são quesitos básicos para conhecer uma população e seus costumes para analisar as variáveis de forma individual e relacionadas.

Os tipos de variáveis existente são as variáveis qualitativa também conhecidas como variáveis categóricas, essas variáveis medem uma qualidade, não possui valor quantitativo, podendo ser ordinal quando possuem uma ordem natural, um exemplo índice de aprovação (ruim, regular, bom, ótimo),ou nominal onde não há uma ordem natural, como por exemplo sexo de uma pessoa, e as variáveis qualitativas que medem uma quantidade podendo ser discretas, onde os possíveis valores são contáveis, ou contínuas onde podem ser observados quaisquer valores de um intervalo. Através dessas variáveis é possível produzir gráficos e assim caracterizar e relacionar populações com diferentes variáveis.

A caracterização de populações se torna ainda mais importante para regiões com problemas relacionados às pessoas do local, pois é um mecanismo que auxilia no diagnóstico e solução das questões do ambiente. Um exemplo disso é a mesorregião Noroeste Fluminense, que é cenário de diversos problemas. O Noroeste Fluminense se localiza entre o [Espírito Santo](http://desciclopedia.org/wiki/Esp%C3%ADrito_Santo) e [Minas Gerais](http://desciclopedia.org/wiki/Minas_Gerais), mas oficialmente pertence ao [Rio de Janeiro](http://desciclopedia.org/wiki/Rio_de_Janeiro). Constituída pelos municípios de Itaperuna e Santo Antônio de Pádua, a mesorregião tem como principal problema a pobreza.

Visto isso, pode-se dizer que o estudo visa a caracterização da população adulta do Noroeste Fluminense e as possíveis relações entre as diferentes variáveis. A importância desse tipo de análise se dá no sentido de conhecimento da situação atual que a população estudada está submetida. Dessa forma, com base nos dados colhidos, conhecer as possíveis relações de causa para o cenário atual, propor e planejar possíveis mudanças sobre determinados problemas, como novas políticas públicas que visem a solução dos problemas da população.

**Objetivo**

Caracterização da população adulta do Noroeste Fluminense a partir de análises estatísticas de dados no software R. A partir disso, conhecer a situação atual da população para propor soluções a partir de políticas públicas.

**Metodologia**

Foram entrevistadas 32.561 pessoas da mesorregião Noroeste Fluminense. A partir de pesquisas de opinião foram coletadas informações, como: ocupação, salário, escolaridade, anos de estudo, situação conjugal, setor, raça, sexo, horas de trabalho na semana, local e IMC. Os dados foram organizados em uma planilha no software Excel para posterior análise no software R.

Para o estudo de caso utilizou-se a linguagem R para confeccionar e analisar suas informações e dados. A partir do banco de dados com essas informações foi possível fazer análises estatísticas e gráficas, utilizando os dados quantitativos e qualitativos de variáveis de maneira individual e relacionando-as entre si.

Foram realizadas as seguintes tipos de análises:

* Gráfico de Pizza
* Gráfico de Barras bivariados
* Histograma
* Boxplot
* Diagrama de Dispersão

Foram realizados alguns gráficos de pizza com diferentes variáveis qualitativas, uma dessas variáveis utilizada na confecção do gráfico de pizza foi variável raça, onde se obteve os percentuais da população que se considerava branca, preto, indígena, amarelo e outros. Foram também analisadas outras variáveis nos gráficos de pizza como sexo, escolaridade e ICM.

No gráfico de barras foram analisadas a relação de algumas variáveis como sexo e ICM. O histograma foi constituído das variáveis ocupação e escolaridade. Através das relações das variáveis a partir desses gráficos é possível explorar o perfil da população.

As variáveis aplicadas no boxplot foram raça x anos de estudo, sexo x anos de estudo e situação conjugal x horas de trabalho.

As variáveis sexo x anos de estudo e sexo x idade foram as variáveis utilizadas para produzir os diagramas de dispersão e assim avaliar suas correlações.

Fundamentado nos resultados das estatísticas e dos gráficos gerados foi possível verificar os dados da população e assim formular e assumir resultados.

**Resultados e Discussão**

Com as informações da população adulta do Noroeste Fluminense produziu-se gráficos através do programa R.

Com ferramenta R foi executável a produção de um o gráfico de pizza onde verificou-se que a população adulta é composta em sua maioria pela raça branca com 85,43%, seguida com 9,59% composta por preto, 3,19% amarelo e quase 1% indígena, outra raça correspondeu a 0,88%. Com o gráfico de pizza também foi possível analisar a porcentagem de homens e mulheres da população adulta da região, observou-se que cerca de 66% é composta de homens contra 33,08% de mulheres.

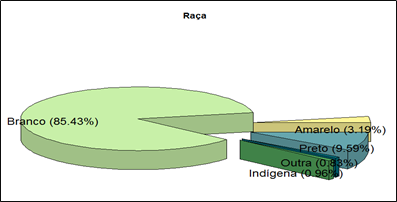


Gráfico de pizza 1-Porcentagem da raça da população adulta do Noroeste Fluminense.

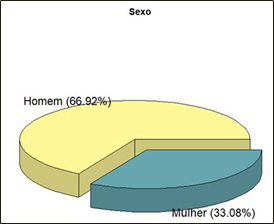


Gráfico de pizza 2. Porcentagem de homens e mulheres na população adulta no Noroeste Fluminense.

A escolaridade da população também foi avaliada e com o gráfico de pizza foi possível visualizar o perfil escolar. Por meio do gráfico podemos constatar que grande parte da população, 32,25%, possui o ensino médio completo e o índice de analfabetismo é bem baixo com apenas 0,16% da população.

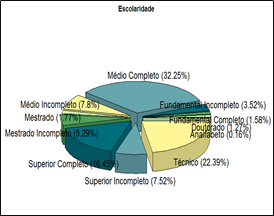


Gráfico de pizza 3. Escolaridade da população do Noroeste Fluminense.

O IMC, que é a sigla para ìndice de massa corporal, é uma forma de avaliar o peso do indivíduo e assim verificar se o mesmo está dentro do peso ideal, acima ou abaixo do peso desejado. O IMC é calculado através da fórmula: peso ÷ altura x altura. O IMC varia com a idade e o estilo de vida de cada pessoa, para adultos um IMC entre 18,6 e 24,9 indica um peso normal saudável, abaixo de 18,5 já considerado abaixo do peso ideal e acima de 25 até 29,9 é considerado sobrepeso, um IMC 30 ou superior classifica como obesidade.

O gráfico do Índice de massa corporal mostra que 75,92% da população adulta está com o IMC igual ou abaixo de 30, o que nos diz que a maioria da população está com o dentro dos parâmetros peso normal, sobrepeso ou abaixo do peso. Das pessoas entrevistadas 24,08% estão com o valor de IMC acima de 30, o que demonstra uma preocupação já que tal índice indica obesidade.

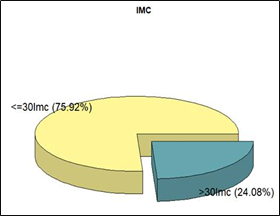


Gráfico de pizza 4. IMC da população do Noroeste Fluminense.

### 

Com as informações do banco de dados sobre a população também foi possível analisar duas variáveis ao mesmo tempo e ter informações mais profundas e comparáveis da população. Com um gráfico de barra bivariados obteve-se alguns resultados sobre o IMC entre homens e mulheres da população, segundo o gráfico a o número de homens que possuem IMC igual ou menor que 30 é superior ao número de mulheres, cerca de quase 10.000 mulheres possuem esse IMC,

O IMC acima de 30 tem ocorrência maior no sexo masculino, mais de 6.000 homens, esse índice nas mulheres é bem mais baixo quando comparado, devido a diferença amostral, número de indivíduos do sexo masculino na população é maior do que o sexo feminino, podemos dizer que eles possuem os maiores IMC.

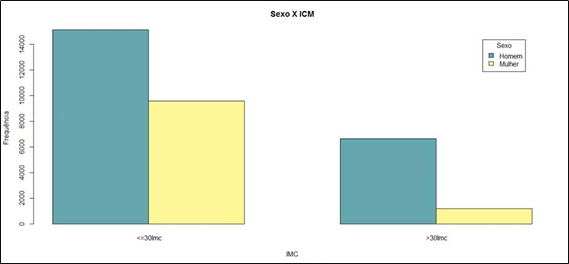


Gráfico de barras bivariadas 4. IMC entre homens e mulheres.

Com as ferramentas do R também é possível confeccionar histogramas. Levantando uma análise da ocupação e escolaridade obtivemos o seguinte histograma. Indivíduos que possuem o ensino médio completo é a maioria nas diferentes ocupações. Através do histograma podemos perceber que na região estudada todos os entrevistados já trabalharam na vida. A empresa privada é a ocupação que mais possui diferentes tipos de escolaridade.

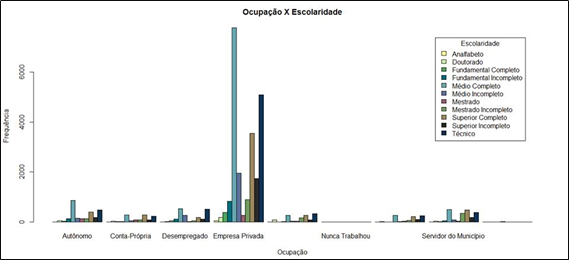


Gráfico 5.. Histograma - Ocupação x escolaridade.

Criou-se também para representar a variação de dados observados de uma variável numérica por meio de quartis. Tendo em vista o box plot comparando raças com seus respectivos anos de estudo, é possível dizer que as medianas das diferentes raças são parecidas. Contudo os quartis dos box plots indígenas, outro e pretos são menores que os demais e possuem mais outliers. Isso indica que existem mais pessoas dessas etnias destoando da média dos entrevistados da mesma raça.

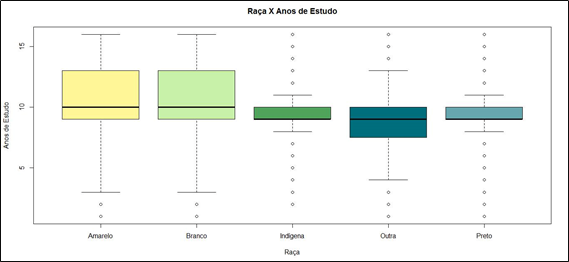


Gráfico 6 Box plot Raça x anos de estudo.



Tabela 1. Anos de estudo e raça.

Verificando o box plot criado com as variáveis anos de estudo e sexo é possível afirmar que não existe uma diferença abrupta de anos de estudo entre os sexos. Com isso, pode-se dizer que o gênero não é significativo nos anos de estudo de maneira geral.



Gráfico 7 Boxplot - Sexo x anos de estudo.

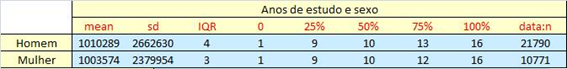


Tabela 2. Anos de estudo x sexo.

O diagrama de dispersão é uma representação gráfica de valores simultâneos de duas variáveis relacionadas a um mesmo processo, mostrando o que acontece com uma variável quando a outra se altera, ajudando desta forma a verificar a relação entre elas. Com o R fez-se um diagrama de dispersão ou correlação, entre os dados de horas trabalhadas e os anos de estudo, pode-se dizer que a linha de tendência possui uma fraca inclinação, comprovando que não existe uma correlação entre as variáveis apresentadas.

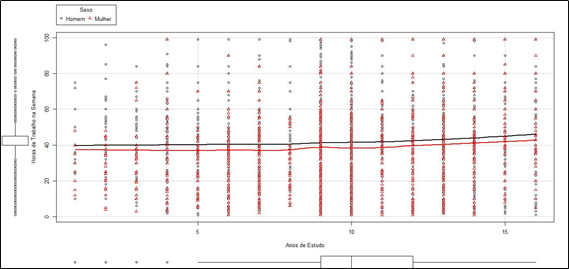


Gráfico 8. Diagrama de dispersão- anos de estudo x horas de trabalho na semana.

**Considerações Finais/ recomendações**

Política Pública

A saúde no Brasil são direitos universais de todos os brasileiros, tendo isso como base e os dados das análise das informações da população do Noroeste Fluminense, uma sugestão de política pública é em relação a saúde das pessoas, como foi observado em gráficos o IMC da população não é tão ruim, porém visando uma melhoria de vida o incentivo a atividades físicas e o cuidado com a saúde podem atender níveis de IMC mais desejáveis. A implementação de ações ao combate a obesidade, o incentivo a atividade física, cuidados com a alimentação através da informação e a disponibilidade de instrumento para prática de exercícios físicos podem ajudar a população, principalmente os indivíduos do sexo masculino, a ter um estilo de vida mais saudável.

Com o auxílio da ferramenta R foi possível a criação de diversos gráficos e análise de diferentes variáveis.