Marca_EESP_INGL

Alunos: Jeferson, Alfredo e Edevlado

Disciplina: Estatística

Professor: Bruno De Oliveira Cruz

Monitor: Thiago Mendes Rosa

Data de Entrega final: 12/12/2020

Trabalho em Grupo

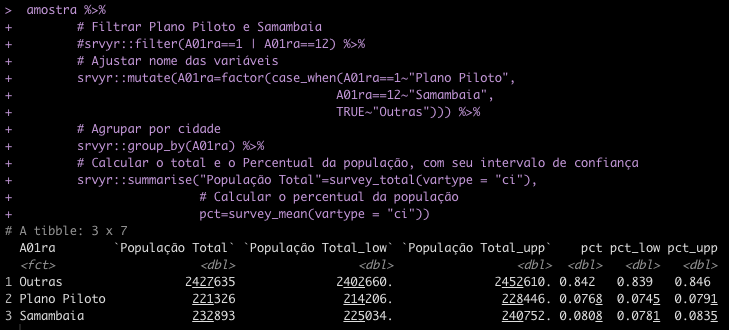
O trabalho será realizado em grupo (máximo 4 pessoas) e o objetivo é aplicar os conceitos aprendidos em sala de aula para uma base de dados real, a Pesquisa Distrital por Amostra de Domicílios (PDAD) da Codeplan. A PDAD é uma pesquisa domiciliar, composta por dois questionários, o de pessoas e o de domicílios. A base de dados, os questionários e o manual do entrevistador estão disponíveis no Github do curso e no site da Codeplan no seguinte endereço.

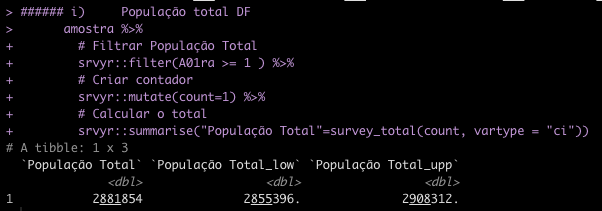
O grupo deverá escolher uma Região Administrativa para análise (aqui denominada RA X) e compará-la ao Plano Piloto (Região Administrativa 1). Os grupos irão realizar entregas parciais e discussões ao longo do curso, em especial nos dias de laboratório.

Parte 1 – Análise Exploratória de Dados

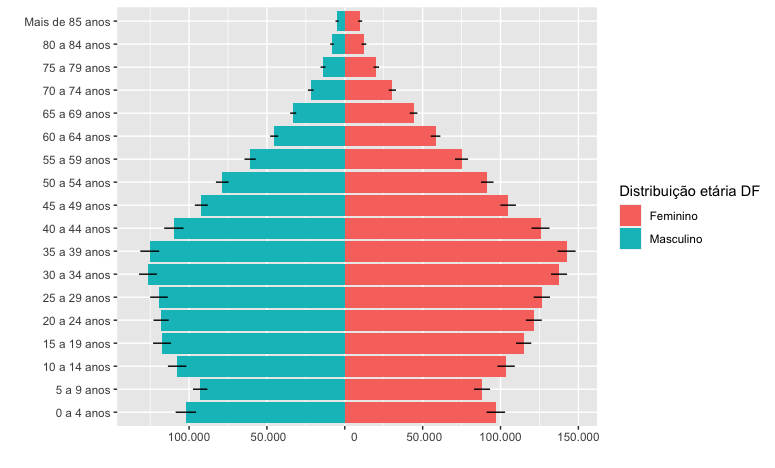
* 1. - Apresente um perfil da RA X, Plano Piloto e do Distrito Federal, estimando as seguintes variáveis:

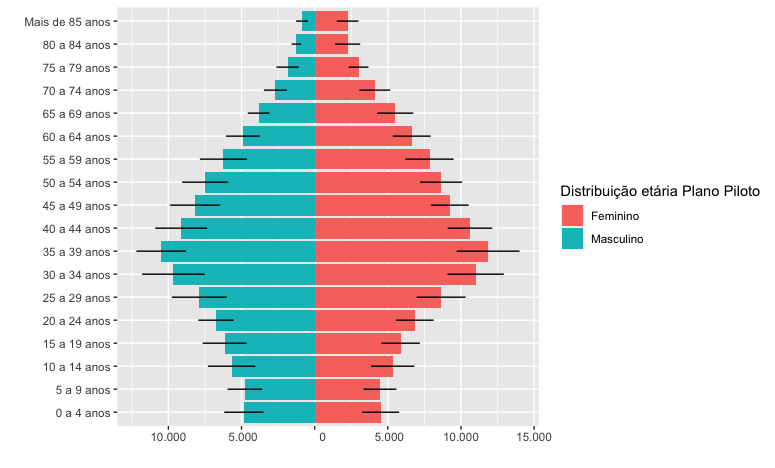
1. População total

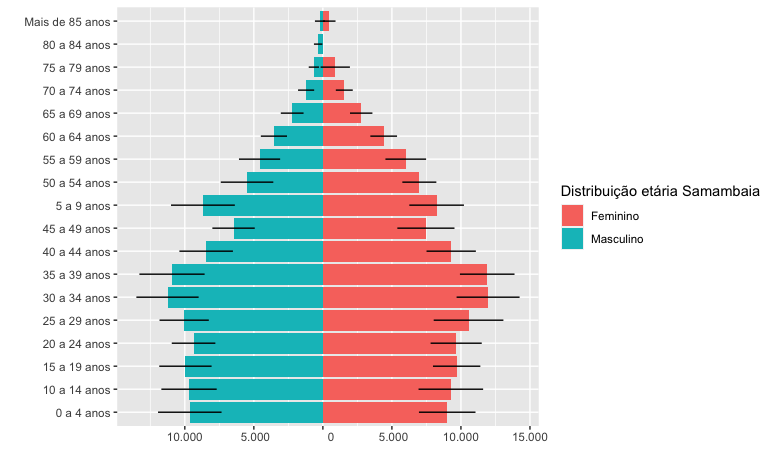




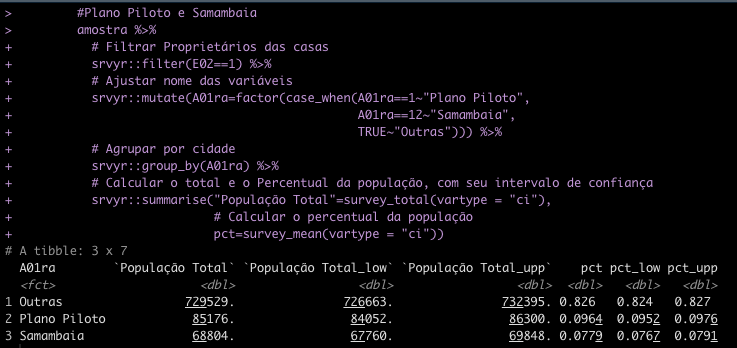
1. Distribuição etária da população (faça uma pirâmide etária, separando homens e mulheres, com classes variando de 5 em 5 anos de 0-4; 5-9; 10-14...)

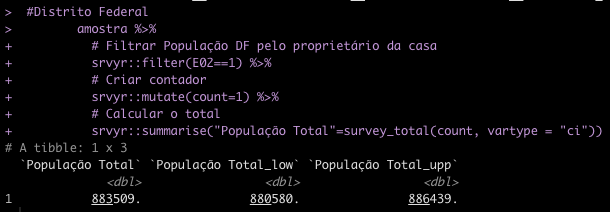




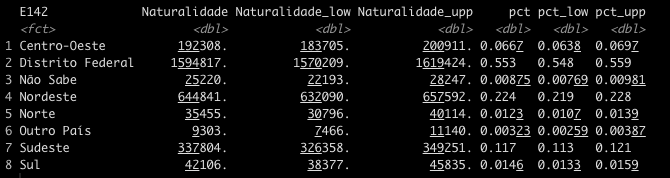


1. Número de domicílios





1. Naturalidade dos residentes (Região de Nascimento use as grandes regiões Norte, Nordeste, Centro-Oeste sem DF, Sudeste e Sul, crie uma categoria especial nascido DF)



* 1. – Calcule (quando possível) a média, mediana, variância, desvio padrão, para as seguintes variáveis com os seguintes filtros Distrito Federal como um todo, Plano Piloto e RA X:

1. Renda domiciliar per capita (calcule também Quantis Q1, Q3 e o percentil 99, ou seja, o valor do 1% mais rico daquela RA)
2. Renda do trabalho principal (calcule também Quantis Q1, Q3 e o percentil 99, ou seja, o valor do 1% mais rico daquela RA).
3. Escolaridade das pessoas de 25 anos ou mais. (note que é variável qualitativa ordenada)
4. Modo de transporte para o trabalho (apenas uma variável qualitativa)
5. Tempo gasto de deslocamento ao trabalho (veja que a variável está em classes, podemos colocar o ponto médio do intervalo para o cálculo de medidas de posição).
6. Número de automóveis no domicilio
7. Número de pessoas no domicílio
8. Número de cômodos como dormitório

Comente os resultados obtidos, compare a RA X´ com o Plano Piloto e o Distrito Federal. Descreva qual relação esperada entre a renda média, o modo de transporte e escolaridade das pessoas de mais de 25 anos.

* 1. – Faça um histograma (com barras e alisado) para as variáveis renda domiciliar per capita e número de automóveis no domicílio para a RA X´ com o Plano Piloto e o Distrito Federal.

1. Qual sua conclusão sobre o grau de simetria de cada distribuição? Justifique a resposta comparando a média e mediana das distribuições? Faça um gráfico QQ da RA X´ e o Plano Piloto; RA X´ e o Distrito Federal.
2. Vamos aplicar duas transformações nas variáveis log e raiz cúbica. Faça novamente os histogramas. Houve alteração em relação ao grau de simetria.
3. Vamos padronizar as variáveis transformadas da seguinte forma: onde x é a renda domiciliar per capita transformada (log (renda domiciliar per capita) ou (renda domiciliar per capita)1/3), é média e é o desvio-padrão. Calcule a média, variância e mediana da variável z. Faça um gráfico QQ da variável com uma normal padrão N(0,1). O que podemos dizer das transformações realizadas?
   1. Análise agora uma distribuição conjunta com tabelas de contingência (dupla entrada):
4. Faça uma tabela de dupla entrada com escolaridade (redefina a variável para fundamental completo ou incompleto; médio completo ou incompleto; superior ou mais completo ou incompleto) por sexo, para pessoas de 25 anos ou mais para a RA X´ e para o Plano Piloto. Crie uma tabela com frequências nas linhas, colunas e total. Calcule o coeficiente de correlação de Pearson (. A partir destas tabelas, é possível fazer alguma inferência sobre a diferença de escolaridade entre Homens e Mulheres na RA X´, compare os resultados com o Plano Piloto.
5. Faça um *box-plot*, calcule a média e variância para as variáveis renda do trabalho principal e região de origem (naturalidade, use apenas as grandes regiões Norte/Nordeste; Centro-Oeste sem DF; Sudeste/Sul, crie uma categoria especial nascido DF para a RA X´ e para o Plano Piloto. Calcule o R2, qual a interpretação desses valores? Existe alguma diferença do local nascimento e renda do trabalho em cada uma das RA´s ?
6. Repita o exercício anterior comparando a variável sexo com a variáveis tempo de tarefas domésticas para a RA X´ e o Plano Piloto. Existe diferença entre o tempo dedicado a afazeres domésticos entre homem e mulher? Em qual RA isso e mais acentuado?

Parte 2 – Teste de Hipótese e Intervalo de Confiança

2.1 – Calcule o intervalo de confiança a 95% de confiança da renda média domiciliar e do tempo dedicado a afazeres domésticos para a RA X´.

2.2 – Faça um teste de hipótese que o número médio de automóveis por domicílio na RA X´ é igual a 1, com grau de confiança de 95% (bi-caudal).

2.3 – Faça um teste de média e compare a renda domiciliar média da Ra X com o Plano Piloto. Repita o mesmo exercício comparando RA X´ e o Distrito Federal.

2.4 – Faça agora um teste de hipótese compara a variância da renda média domiciliar da RA X´ e do Plano Piloto. Repita o mesmo teste comparando a RA X´ e o Plano Piloto.

Parte 3 – Tópicos Especiais

3.1 – Calcule o Gini da renda domiciliar per capita para o DF, RA X´ e o Plano Piloto. Qual a sua conclusão?

3.2 – Vamos supor que o grupo seja convidado a opinar sobre a discussão da gestão da pandemia de Covid-19:

Com bases nos dados da PDAD, qual seria a resposta aos gestores sobre a reabertura ou não de escolas na RA X´?

1. Para subsidiar a discussão, primeiro calcule o número de crianças e adolescentes (0 a 18 anos) que mora com pessoas mais de 60 anos. Calcule o intervalo de confiança desta estimativa. Faça isso para RA X´ e para o DF. Crie um filtro para alunos de escolas particulares e públicas.
2. Calcule o número médio de pessoas por dormitório desses domicílios com crianças para a RA X´ e o DF. Crie um filtro para alunos de escolas particulares e públicas.
3. Calcule o número de crianças, adolescentes e jovens (0 a 29 anos) por nível escolar. Calcule também quantos destes utilizam o transporte coletivo para ir à escola (ônibus, metrô, transporte escolar público ou transporte escolar privado). Calcule o intervalo de confiança desta estimativa. Faça também um filtro para unidades públicas e particulares.
4. Estime o percentual de domicílios conectados à internet para o Distrito Federal. Faça também um filtro estudantes de escolas públicas e particulares, independentemente da idade. Faça também um teste de hipótese a 95% de confiança para comparar se as duas proporções são equivalentes ou não.
5. Pela variável G06 (Qual atividade da empresa que lhe paga o trabalho principal?), caracterize os trabalhadores do setor educação (não necessariamente todos estão em sala de aula, mas é uma boa proxy do perfil dos trabalhadores do setor). Calcule o número de pessoas nessas atividades, o tempo médio de deslocamento ao trabalho e modo de transporte, além da distribuição etária. Calcule também quantos destes trabalhadores moram com pessoas com mais de 60 anos.
6. Estime uma variável de uso do tempo: some o total de horas trabalhadas, o tempo declarado com deslocamento e o tempo declarado com afazeres domésticos para o responsável pelo domicílio e pelo cônjuge. Faça um filtro para domicílios com e sem crianças (menores de 12 anos) que frequentam escola e por sexo.

Com base nesses resultados, qual seria sua recomendação para os gestores?