

Relatório técnico: Análise do preço da gasolina comum entre Minas Gerais e Paraná

João Vitor Maia Neves Cordeiro

25 de julho de 2021

Conteúdo

1	Introdução	2
2	Objeto	2
2.1	Tema	2
2.2	Delimitação do tema	2
2.3	Problema	2
2.4	Hipótese	3
3	Objetivos	3
3.1	Objetivos específicos	3
4	Justificativa	3
5	Metodologia	3
5.1	Método de abordagem	3
5.2	Técnicas de pesquisa	4
6	Apresentação e análise dos resultados	5
6.1	Análise preliminar da coleta	7
6.2	Análise descritiva dos dados	7
6.3	Análise exploratória dos dados	7
6.4	Conclusões	8

1 Introdução

Esse relatório é a recapitulação de um trabalho realizado ao longo do semestre para a disciplina INE5405, ministrada pelo Prof. José Francisco D. de G. C. Fletes, apresentado como método de avaliação para a unidade 1 da disciplina. O trabalho consistiu de uma coleta, normalização, análise e interpretação dos dados coletados através do site da ANP sobre a precificação da gasolina comum em postos dos estados Minas Gerais e Paraná.

2 Objeto

2.1 Tema

Como já dito na seção acima, o trabalho foi realizado em cima de dados coletados no site da ANP, que podem ser encontrados neste link, os dados estão disponíveis para consulta pública. Foram coletados os dados para os estados Minas Gerais e Paraná, apenas na gasolina comum.

2.2 Delimitação do tema

As análises foram feitas a partir de uma normalização inicial dos dados, para padronizar as planilhas coletadas em um formato mais trabalhável e de mais fácil leitura. Depois dessa normalização foi possível trabalhar tanto com dados agrupados quanto não agrupados, flexibilizando as análises e métricas possíveis de serem obtidas sobre o conjunto de dados estudado.

2.3 Problema

Preços de combustível utilizados em métodos de transporte cotidiano são de grande importância para a população em geral, pois além do impacto direto (transporte para o trabalho, viagens de carro e outros) existe também um impacto indireto, como o custo geral de qualquer produto que precise ser transportado até o local de venda, o custo de tarifas de transporte público e até o custo de produtos que normalmente não associamos a logística de transporte. Por isso, é necessário que sejam estudados e investigados valores discrepantes dentro de um estado, município ou bairro.

2.4 Hipótese

Queremos verificar se o preço do combustível possui alguma relação com a origem de sua bandeira, isto é, se por ser de uma bandeira nacional um combustível é mais caro ou mais barato do que em uma bandeira internacional, além de considerarmos também bandeiras brancas (postos que não possuem contrato).

3 Objetivos

Separar faixas de preço normalizadas em cada estado, aplicando técnicas para reduzir a discrepância entre dados dentro de um determinado estado e poder analisar o melhor custo-benefício para o consumidor em cada local.

3.1 Objetivos específicos

4 Justificativa

A partir do problema levantado em 2.3, pretende-se por meio da aplicações de técnicas de análises estatísticas produzir uma interpretação melhor do ponto de vista do consumidor sobre os dados coletados, auxiliando a população geral na hora de decidir em quais postos abastecer e quais bandeiras procurar.

5 Metodologia

5.1 Método de abordagem

1. Primeira etapa: coleta de dados

Utilizando a base de dados fornecida pelo site da ANP, será feita a análise sobre dados agrupados ou não, para que se possa identificar possíveis discrepâncias e taxas de maior frequência.

2. Segunda etapa: análise descritiva dos dados

Já com os dados normalizados, iremos buscar métricas estatísticas descritivas como moda, desvio padrão, variância e outras.

3. Terceira etapa: análise exploratória dos dados

Nessa parte será realizada a análise de valores discrepantes, principalmente classificados por bandeira, encontrados em etapas anteriores.

5.2 Técnicas de pesquisa

1. Primeira etapa: coleta de dados

Nessa etapa usamos histogramas para visualizar as classes de frequência, modelo visual que facilita a visualização de dados agrupados, esse trabalho foi feito com o auxílio do software Excel em uma planilha de dados.

2. Segunda etapa: análise descritiva dos dados

Já na segunda etapa utilizamos fórmulas matemáticas presentes nas referências bibliográficas da disciplina, novamente o software Excel foi utilizado para aplicar essas fórmulas no conjunto de dados e obter resultados mais precisos e menos propícios a erros de cálculo, isso foi feito já que o volume do conjunto de dados era bem significativo

3. Terceira etapa: análise exploratória dos dados

Na última etapa do trabalho, foram utilizados Box-Plots em conjunto com a regra dos quartis de Montgomery para explicitar a existência ou não de outliers e outros valores discrepantes, a fim de entregar uma análise completa do problema.

6 Apresentação e análise dos resultados

Classes de frequência	Frequência	Porcentagem	Moda	Media
5.640 - 5.720	23	7.72	5.669	5.679
5.720 - 5.800	74	24.83	5.799	5.799
5.800 - 5.880	21	7.05	5.85	5.859
5.880 - 5.960	48	16.11	5.899	5.913
5.960 - 6.040	51	17.11	5.999	5.993
6.040 - 6.120	43	14.43	6.099	6.093
6.120 - 6.200	18	6.04	6.149	6.174
6.200 - 6.280	4	1.34	6.279	6.270
6.280 - 6.360	8	2.68	6.288	6.301
6.360 - 6.450	8	2.68	6.398	6.402
Total	298	100.00	5.999	5.941

Figura 1: Tabela de frequências para os dados de Minas Gerais.

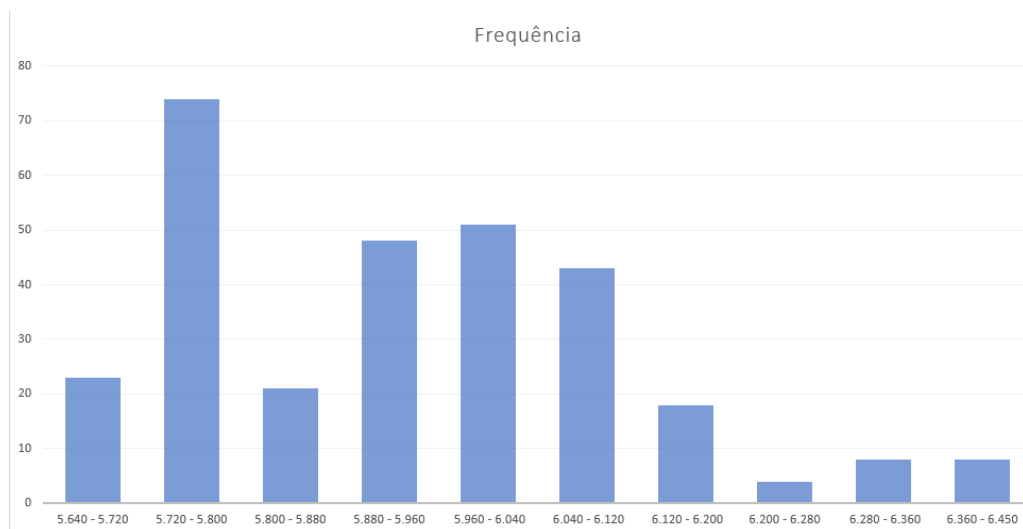


Figura 2: Histograma das frequências para os dados de Minas Gerais.

Classes de frequência	Frequência	Porcentagem	Moda	Media
5.149 - 5.239	27	10.67	5.199	5.191
5.239 - 5.329	53	20.95	5.299	5.291
5.329 - 5.419	45	17.79	5.39	5.375
5.419 - 5.509	37	14.62	5.49	5.478
5.509 - 5.599	40	15.81	5.59	5.577
5.599 - 5.689	17	6.72	5.68	5.655
5.689 - 5.779	18	7.11	5.69	5.709
5.779 - 5.869	12	4.74	5.79	5.793
5.869 - 5.990	4	1.58	5.89	5.940
Geral	253	100.00	5.299	5.456

Figura 3: Tabela de frequências para os dados do Paraná.

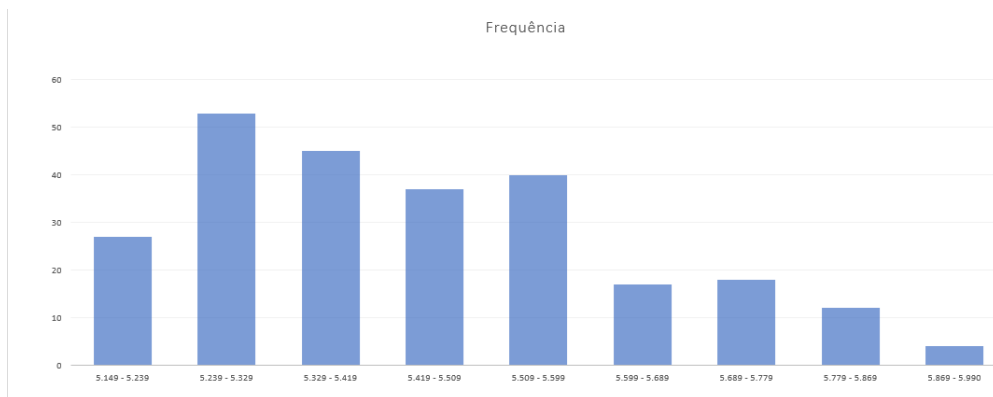


Figura 4: Histograma das frequências para os dados do Paraná.

6.1 Análise preliminar da coleta

Após coletar os dados e aplicar o modelo empírico, foram geradas as tabelas de frequências e histogramas (Figuras de 1 a 4), com isso pudemos ver que o pico de frequências em preços da gasolina comum se encontra em 5,720-5,800 para Minas Gerais e 5,239-5,329 para o Paraná. Ou seja, analisando apenas o dado de frequência de preços podemos notar que há uma maior incidência de preços altos em Minas Gerais do que no Paraná, o que provavelmente impacta fortemente no consumo final do produto.

6.2 Análise descritiva dos dados

Com os dados já organizados foi possível utilizar o software Excel e sua funcionalidade de aplicações de fórmulas para calcular moda, média e outros dados. O cálculo foi baseado em fórmulas presentes na bibliografia da disciplina e se encontram na próprias tabelas de frequências mostradas anteriormente.

6.3 Análise exploratória dos dados

Nessa última etapa, foi realizada uma comparação primeiramente entre estados, e em um segundo momento a análise foi feita em cima de uma separação de 3 bandeiras no estado do Paraná. Para isso, foram utilizados Box-Plots juntamente com a definição de quartis de Montgomery. Em nenhum dos casos foram encontrados outliers.

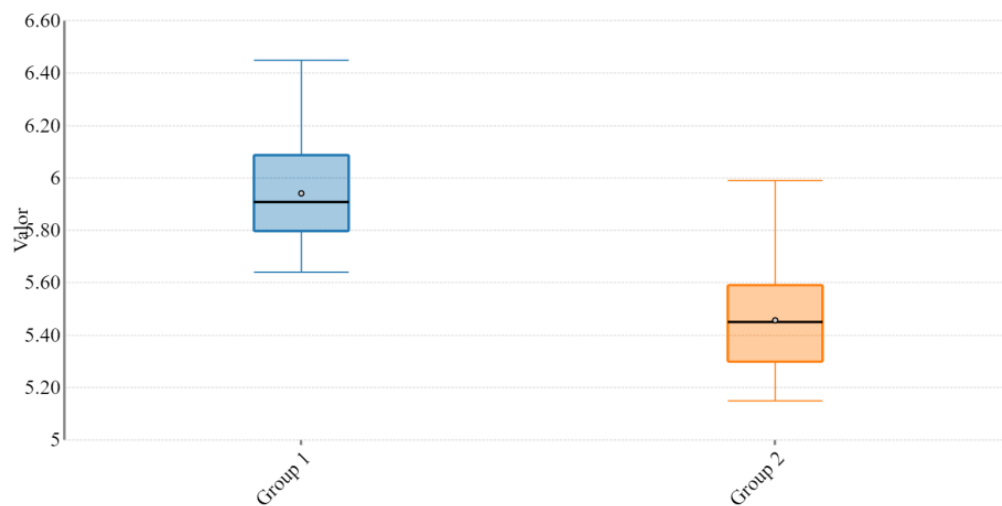


Figura 5: Box plot comparativo entre os estados.

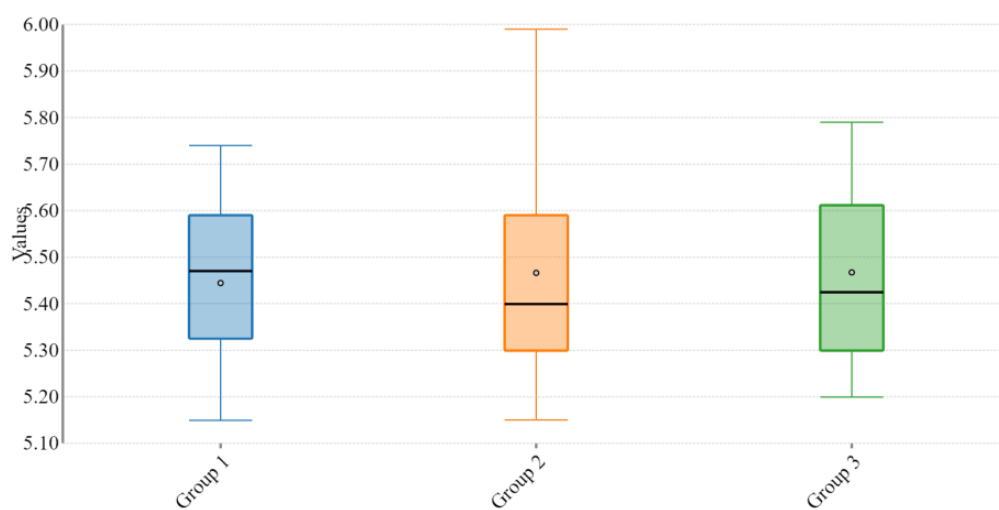


Figura 6: Box plot comparativo entre bandeiras no Paraná.

6.4 Conclusões

Ao realizar uma análise comparativa entre os dois estados podemos perceber uma diferença considerável, onde o estado de Minas Gerais possui preços

em média muito mais altos, o Box-Plot deixa isso ainda mais visível com a análise de quartis. Do ponto de vista do consumidor, além dos preços mais altos nos postos, é possível que preços de produtos comuns que sejam transportados apenas dentro do estado também sejam encarecidos por essa alta no combustível.

A comparação entre as bandeiras nos mostra a primeira vista que as 3 possuem uma posição geral consistente, com suas medianas muito próximas. Além disso, os valores mínimos e os quartis também diferem em poucos centavos, nos dando a entender que existe uma boa equiparação de preços (ou possivelmente um cartel) entre as bandeiras.

Um dado acaba ficando discrepante no gráfico, o valor máximo da bandeira Ipiranga beira os 6 reais, aproximadamente 20 centavos acima dos valores máximos de suas concorrentes. De fato, ao voltarmos para nossa tabela de dados vemos que dentre todas as bandeiras, a Ipiranga possui os maiores valores máximos. Para o consumidor que tenha contato apenas com os postos onde o valor é alto, isso pode indicar que essa bandeira pratica preços mais altos do que as rivais, mas ao analisarmos a distribuição como um todo notamos que isso não é uma verdade.