

Segundo teste - INE5405

João Vitor Maia Neves Cordeiro (19100532)

A segunda entrega consistia de uma análise realizada em cima dos dados da ANP sobre o preço do combustível GASOLINA COMUM nos estados do Paraná e Minas Gerais. O primeiro passo foi a confecção de uma tabela de dados a partir dos arquivos exportados no site da agência nacional do petróleo.

Como os dados dos dois estados não estão relacionados nesse primeiro momento, foi optado por separar em duas tabelas de dados diferentes para cada um.

1. Paraná

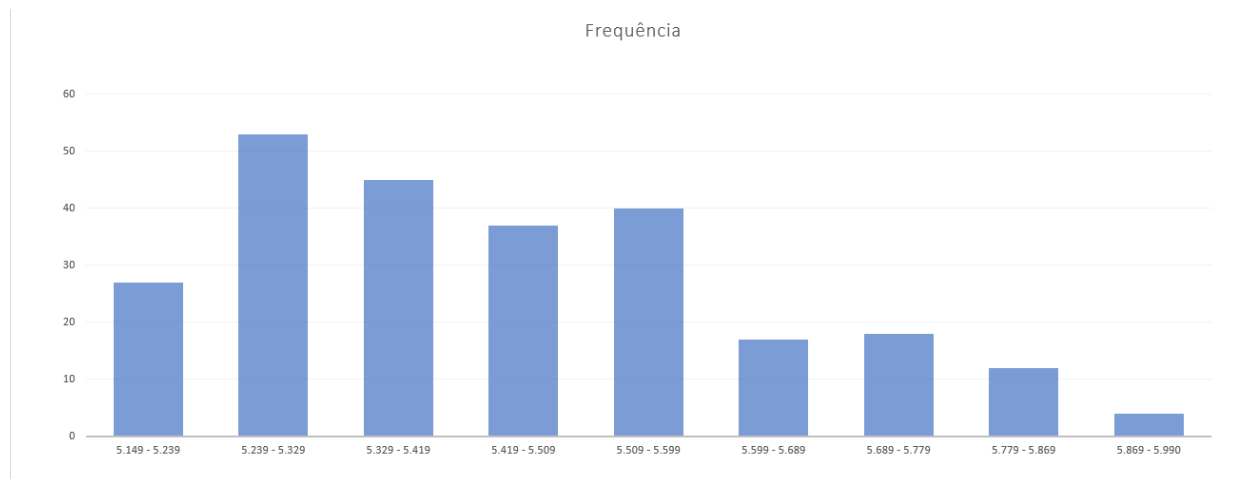
Foram encontradas 253 observações de preços da gasolina comum espalhadas pelo estado, após um breve tratamento de formatação dos dados foi construída uma tabela com [Bandeira, preço da venda, data da coleta dos dados, razão social do posto, endereço e bairro].

Em posse desses dados, encontramos o valor de $X_n = 5.990$ e $X_1 = 5.149$, com um range = **0.841**. Como a quantidade de observações era significativa optei por usar a fórmula de Sturges para obter a quantidade de classes, ao aplicar a fórmula chegamos a 8.978 que ao ser arredondado nos dá 9 classes. Utilizando a fórmula para a amplitude de cada classe chegamos ao valor de **0.09**.

A partir daí o trabalho é concentrado no agrupamento das observações dentro das frequências criadas.

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

Depois disso com os dados das frequências pudemos gerar um histograma com o auxílio do Excel. Ao aplicar os conceitos vistos em aula detectamos que o pico de frequências se encontra na faixa **5.239 - 5.329** contando com 53 aparições. Além disso percebemos que é um gráfico com assimetria positiva, concentrando a moda a esquerda do eixo X com uma cauda tendendo para a direita.



2. Minas Gerais

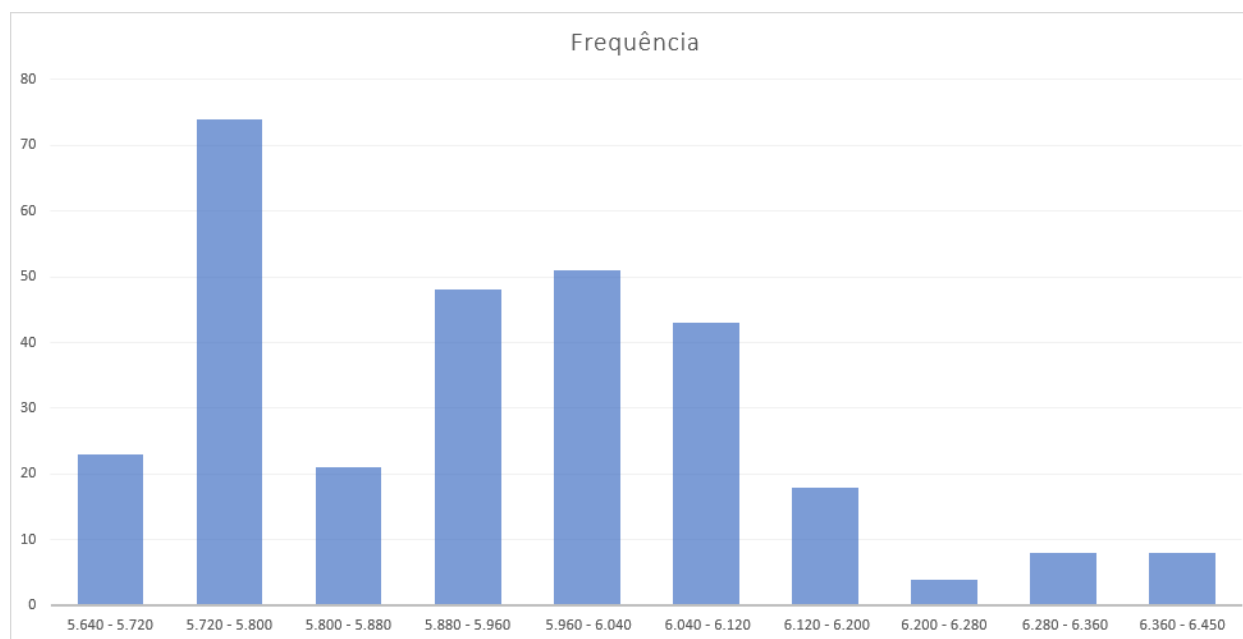
Foram encontradas 298 observações de preços da gasolina comum espalhadas pelo estado, após um breve tratamento de formatação dos dados foi construída uma tabela com [Bandeira, preço da venda, data da coleta dos dados, razão social do posto, endereço e bairro].

Em posse desses dados, encontramos o valor de $X_n = 6.449$ e $X_1 = 5.640$, com um range = **0.809**. Como a quantidade de observações era significativa optei por usar a fórmula de Sturges para obter a quantidade de classes, ao aplicar a fórmula chegamos a **9.214** que ao ser arredondado nos dá **10** classes. Utilizando a fórmula para a amplitude de cada classe chegamos ao valor de **0.08**.

Novamente agrupamos as frequências encontradas em uma nova tabela

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

Depois disso com os dados das frequências pudemos gerar um histograma com o auxílio do Excel. Dessa vez ao observar o gráfico notamos que o pico de frequências se encontra na faixa **5.720 - 5.800** contando com 74 aparições. Novamente temos um um gráfico com assimetria positiva, concentrando a moda a esquerda do eixo X com uma cauda tendendo para a direita.



Ao final do trabalho de coleta, formatação e tratamento inicial dos dados ainda não consegui obter uma boa análise do ponto de vista do consumidor, já que levando em consideração apenas os dados coletados acabam ficando muitas lacunas e variáveis incertas que na minha opinião dificultam uma análise assertiva.