

## Teste 06 - INE5405

### Rafael Begnini de Castilhos (20205642)

1) Obtenha, do(s) modelo(s) empírico(s)/dados agrupados\_DA e dos dados originais/dados não-agrupados coletados\_DNA, as principais estatísticas descritivas (valor central: a média, a mediana e a moda (se existir); dispersão: variância, desvio padrão, erro padrão da média e coeficiente de variação; forma: assimetria) para analisar as situações desejadas. Interprete os resultados:

a) obtendo o erro relativo entre as estatísticas fundamentais para DNA e DA;

b) das estatísticas obtidas do ponto de vista do consumidor.

Para obter as principais estatísticas descritivas como valor central, dispersão e forma, foi necessário utilizar os métodos científicos e fórmulas apresentadas pelo professor na aula síncrona.

Dados não agrupados:

#### Rio Grande do Sul:

Média ponderada	5,322	
Média	5,331	
Moda	5,299	
Mediana	5,299	
Variância	0,047	
Desvio padrão	0,218	
Erro padrão	0,013	
Coeficiente de variação	0,04	
Coeficiente de assimetria	0,148	Assim. Moderada

#### Paraná:

Média ponderada	5,238	
Média	5,225	
Moda	4,999	
Mediana	5,2	
Variância	0,028	
Desvio padrão	0,168	
Erro padrão	0,01	
Coeficiente de variação	0,032	
Coeficiente de assimetria	1,345	Assim. Muito Forte

Dados agrupados:

Média ponderada	5,280	
Média	5,280	
Moda	5,099	
Mediana	5,28	
Variância	0,041	
Desvio padrão	0,202	
Erro padrão	0,009	
Coeficiente de variação	0,038	
Coeficiente de assimetria	0,898	Assimet. Forte

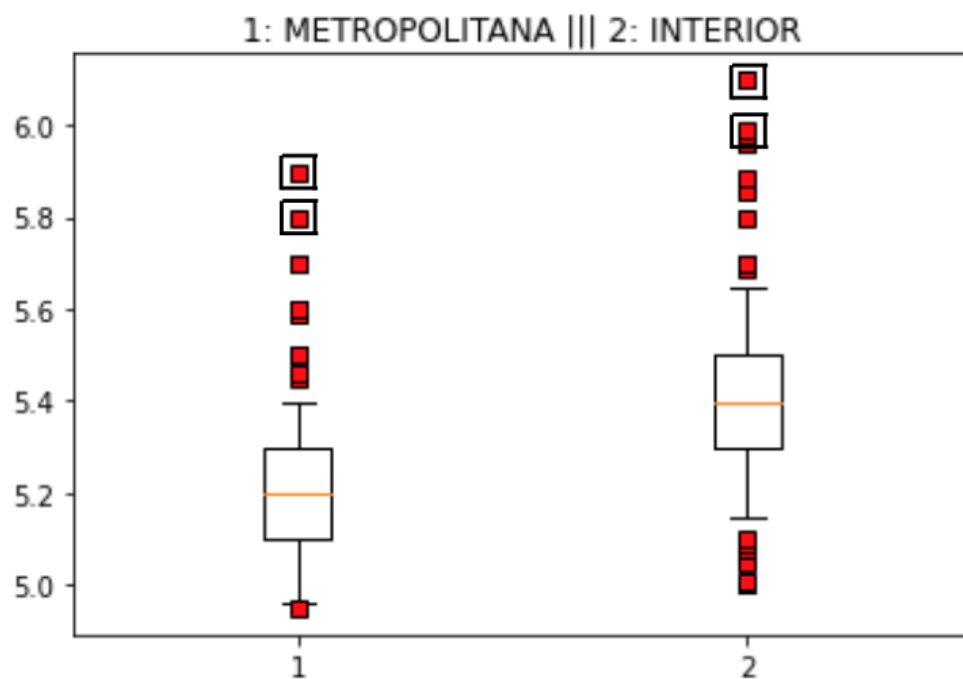
Erro relativo DNA e DA: Para encontrar o erro relativo, foi necessário utilizar a fórmula  $Er = (|X_s - X_p|) / X_s$ . Com isso, foi encontrado para os dados agrupados um erro relativo de 0,007% (muito próximo de zero), já para os agrupados foi encontrado 0,6% para o estado do Rio Grande do Sul e 0,4% para o Paraná. Quanto menor o erro relativo, melhor a precisão de uma medida em relação a outra, assim constatamos que é uma boa aproximação, não havendo muita variabilidade nos dados.

No ponto de vista do consumidor, essas estatísticas mostradas acima, retratam e expressam fielmente a realidade. O consumidor de Diesel S10 hodierno (assim como demais tipos e categorias), não percebe grande expressividade na diferença de preços entre bandeira do posto de gasolina e localização da região. Foi constatado que no estado do Paraná, todos os valores centrais estão abaixo comparado ao estado do Rio Grande do Sul, sendo assim, o consumidor do Paraná acaba sendo beneficiado em centavos, e até mesmo na terceira casa decimal dos centavos, entretanto, cabe analisar que tal precificação é por litro, desse modo, quando há um consumo de grande quantidade, até a terceira casa decimal irá influenciar no montante.

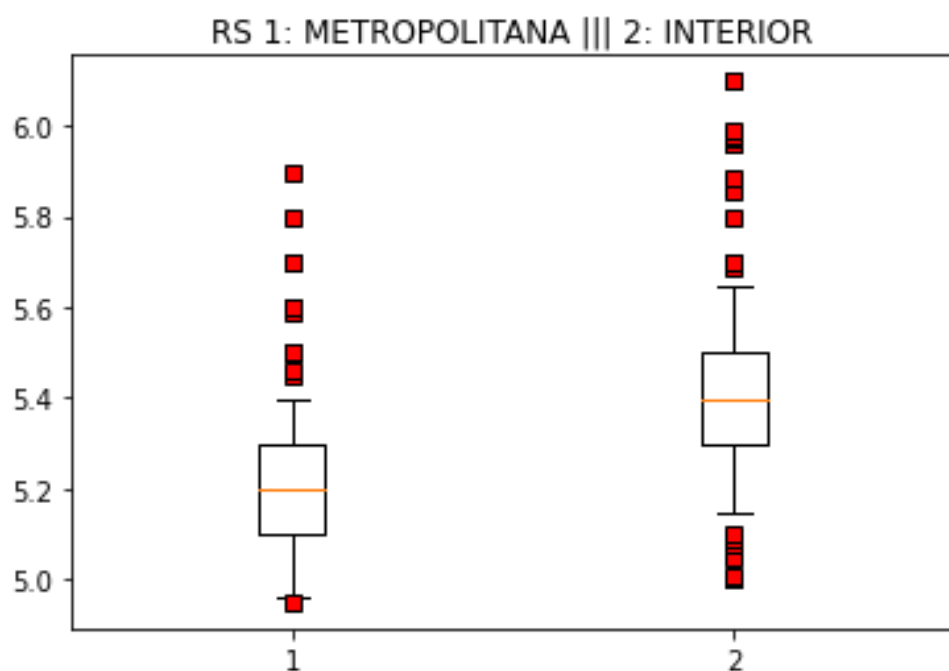
**2)** Faça uma análise exploratória dos dados: analisando os Box-plot's ou Diagramas em Caixa (esquema dos cinco números) comparativos, para a tarefa definida individualmente.

**2.1-** Entre dois Estados avaliados? (Entre duas regiões/estados (METROPOLITANA X INTERIOR)?)

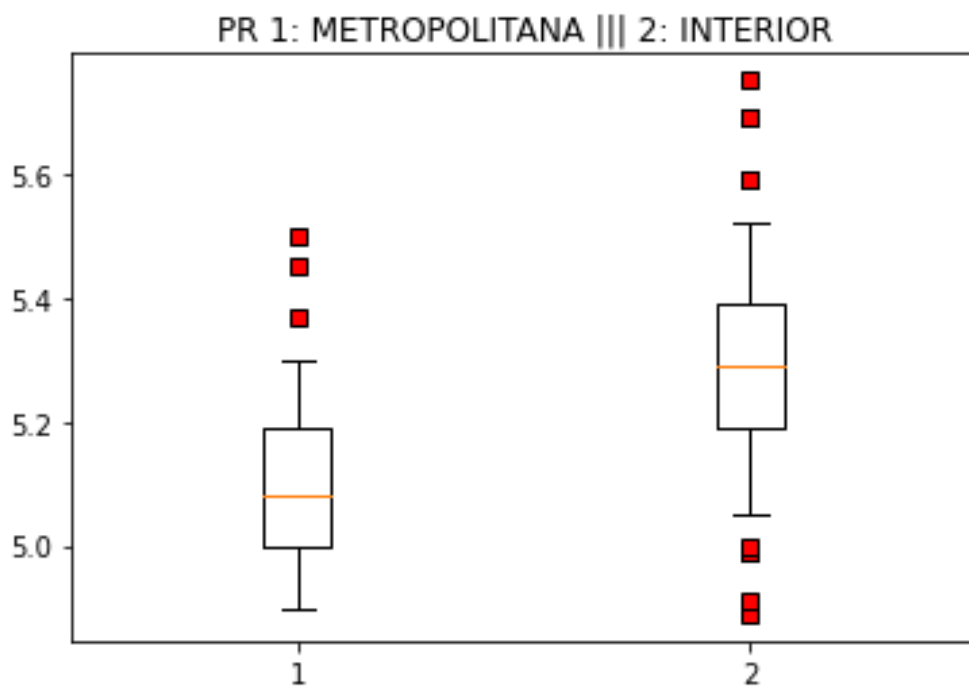
Região Metropolitana x Interior (Agrupados)						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Metropolitana	4,899	5,069	5,149	5,249	5,899	0,18
Interior	4,89	5,202	5,343	5,42	6,099	0,218



Região Metropolitana x Interior (RS)						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Metropolitana	4,949	5,099	5,197	5,298	5,899	0,199
Interior	4,999	5,298	5,399	5,499	6,099	0,201

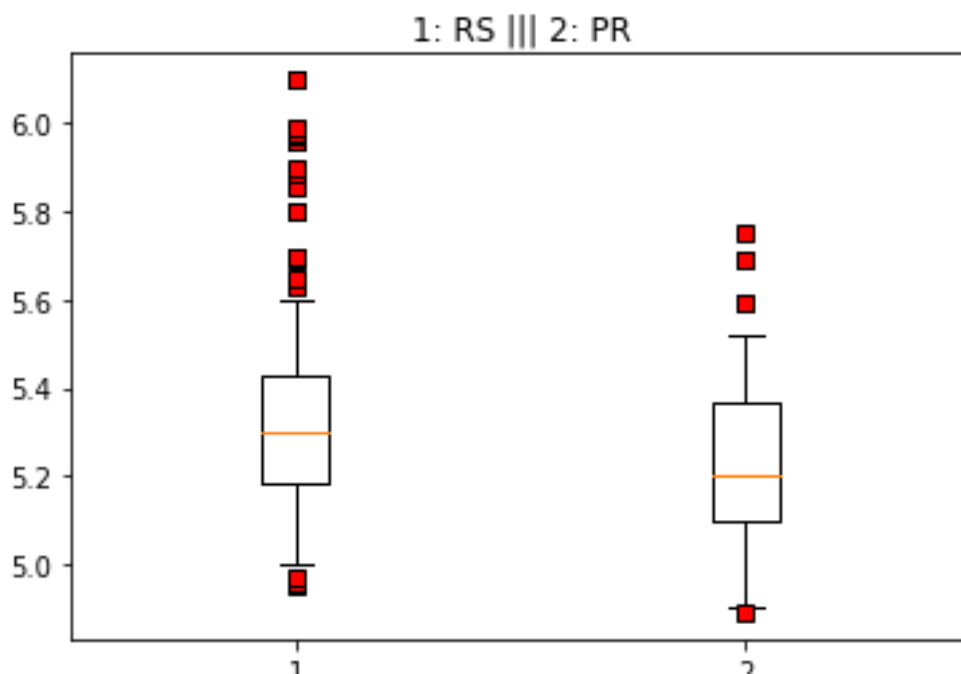


Região Metropolitana x Interior (PR)						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Metropolitana	4,899	4,999	5,079	5,19	5,499	0,191
Interior	4,89	5,19	5,29	5,39	5,75	0,2

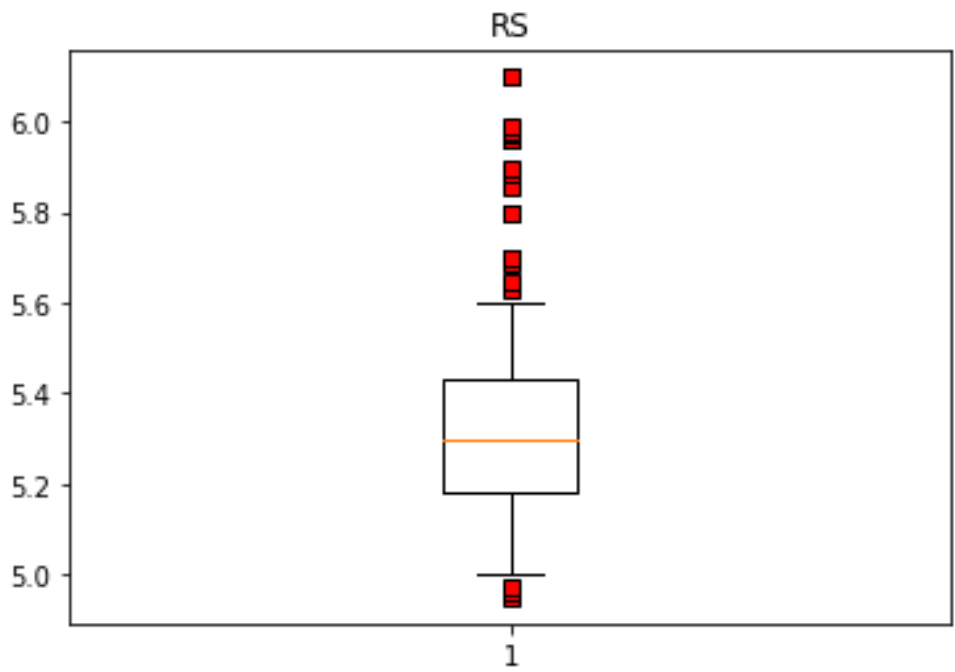


## 2.2- Para o Estado definido entre dois combustíveis?

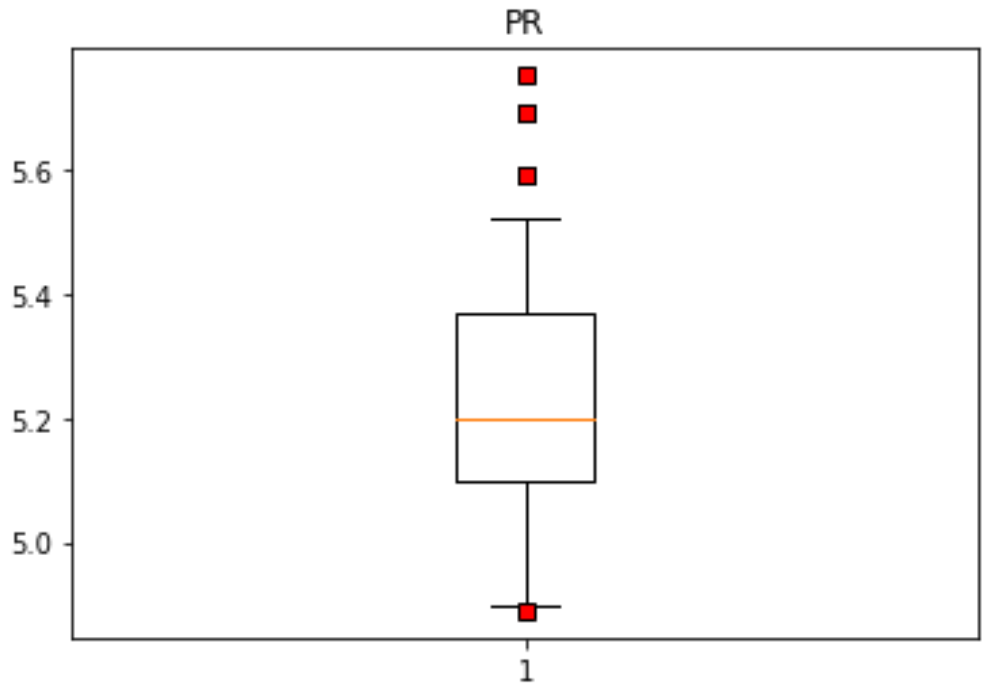
Estado Rio Grande do Sul x Paraná						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Rio Grande do Sul	4,949	5,181	5,298	5,432	6,099	0,251
Paraná	4,89	5,099	5,2	5,369	5,75	0,27



Estado RS						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Rio Grande do Sul	4,949	5,181	5,298	5,432	6,099	0,251

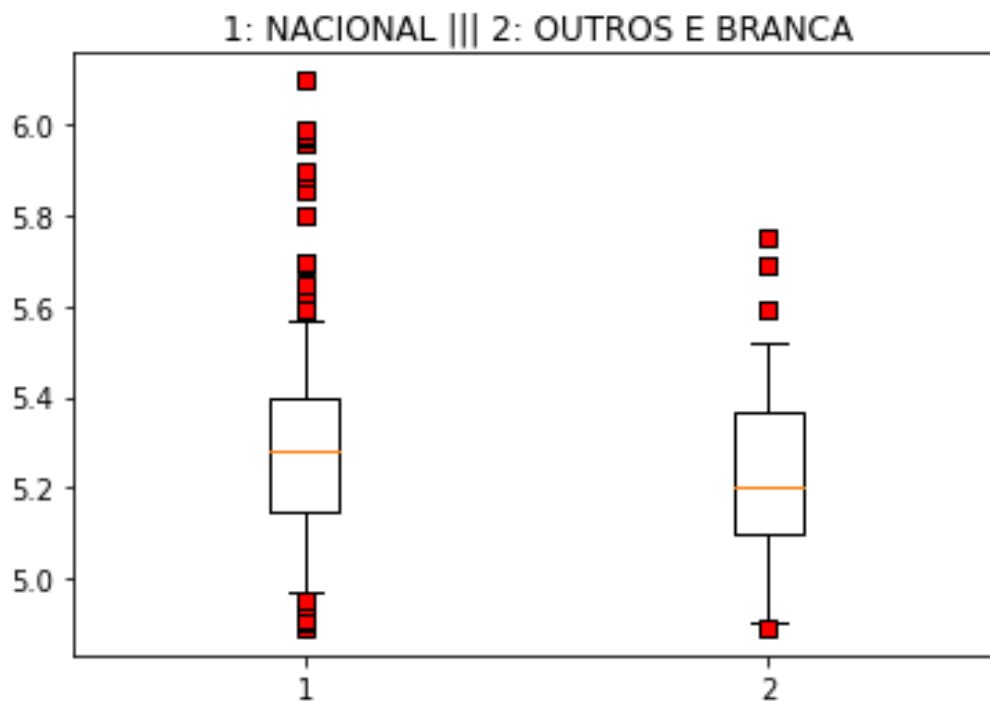


Estado PR						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Paraná	4,89	5,099	5,2	5,369	5,75	0,27

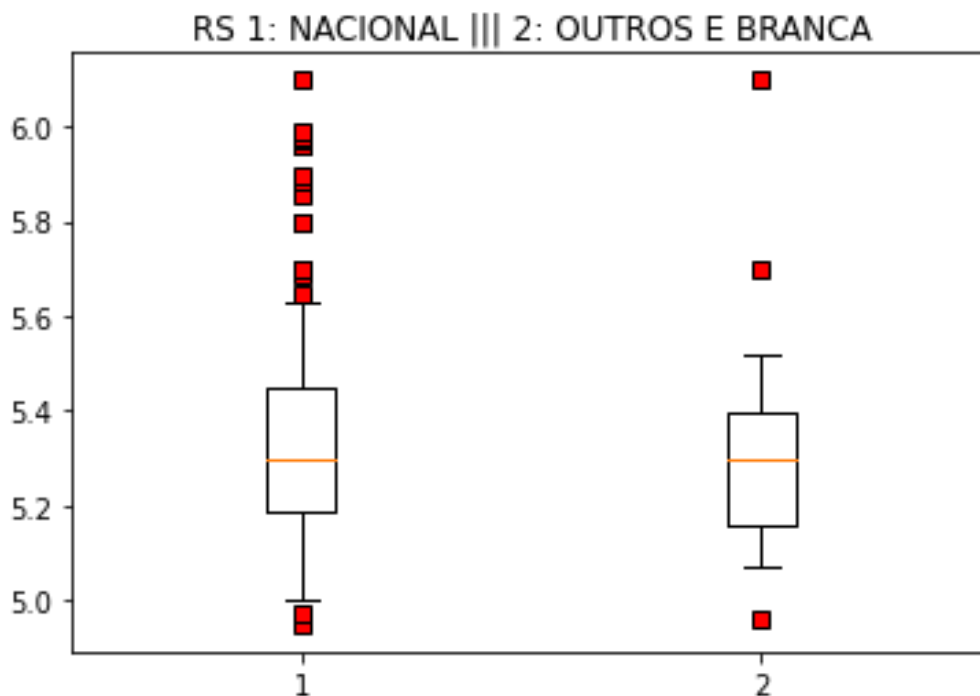


### 2.3- Para os dois grupos de postos de combustíveis?

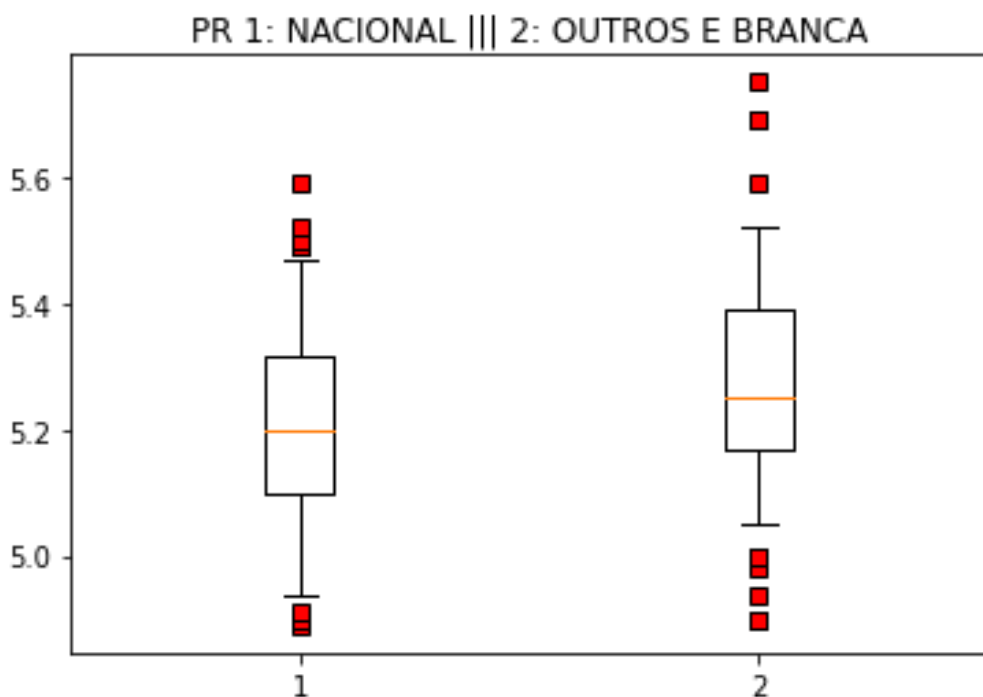
Tipo Bandeira Nacional x Outras e branca						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Nacional	4,89	5,149	5,279	5,399	6,099	0,25
Outras e branca	4,89	5,099	5,2	5,369	5,75	0,27



Tipo Bandeira Nacional x Outras e branca (RS)						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Nacional	4,949	5,189	5,298	5,449	6,099	0,26
Outros e branca	4,959	5,159	5,298	5,399	6,099	0,24



Tipo Bandeira Nacional x Outras e branca (PR)						
Quartis	Q0	Q1	Mediana	Q3	Q4	IQR
Nacional	4,89	5,099	5,199	5,315	5,59	0,216
Outros e branca	4,899	5,17	5,25	5,39	5,75	0,22



**3)** Verifique se existem valores discrepantes (“outliers” e/ou “outliers extremos”), usando a regra básica segundo os quartis e interprete o significado destes do ponto de vista do consumidor.

Utilizando o critério de quartis, foi possível realizar análise de discrepantes e suspeitos. Em todos os diagramas de caixa acima, foi realizado o esquema dos 5 números, utilizando a regra básica, também determinando o intervalo inter-quartis de cada uma.

Esmiuçando os dados, em todos os casos foram encontrados outsiders (possíveis suspeitos), que são representados por quadrados vermelhos. Em uma situação específica, foram encontrados outliers (suspeitos), que são representados por quadrados vermelhos rodeados por um quadrado branco de borda preta. Na comparação de Região Metropolitana com Interior nos dados Agrupados, foi possível identificar 2 pontos da Região Metropolitana e 2 pontos da Região Interior que estão fora da curva normal, estando localizados na extrema direita do gráfico de curva.

Com isso, é possível identificar quais são os dados que se diferenciam drasticamente de todos os outros, ou seja, um valor que foge da normalidade e provavelmente irá causar anomalias nos resultados obtidos.

No ponto de vista do consumidor, recomenda-se evitar totalmente o consumo de outliers, e analisar a devida situação dos outsiders à esquerda da moda, pois pode haver

promoções e descontos, e caso o posto em questão possuir uma boa reputação, ou bandeira conceituada, não se tornaria um problema efetivo.

Em ambos estados, a região metropolitana apresenta uma média em torno de 20 centavos de diferença com o interior. No estado do Paraná os postos que possuem bandeira nacional, apresentam uma leve diferença, em torno de 5 centavos para menos, com os que possuem bandeira branca e outros. Já no Rio Grande do Sul não há diferença no valor médio entre as bandeiras, porém deve ser considerado que os postos de bandeiras nacionais possuem uma quantidade expressiva de outsiders.

Tendo como base os fatos acima, se torna possível constatar que a melhor opção de compra é na região metropolitana do Paraná nos postos cuja bandeira é nacional. Já no estado do Rio Grande do Sul, a melhor opção de compra é na região metropolitana, não havendo preferência por bandeira, pois há um equilíbrio no preço.