

**FLORIANÓPOLIS, 07 DE DEZEMBRO**

Livia Corazza Ferrão

Análise estatística dos preços de marcas de bebidas

Trabalho da Disciplina INE 5405  
– Estatística e Probabilidade,  
sob Orientação do Prof. José  
Francisco D. de G. C. Fletes.

Florianópolis, 07 de Dezembro de 2023

## 1. Objetivo Geral

O que podemos avaliar quanto ao comportamento dos preços do ponto de vista do consumidor (de acordo com a tarefa definida como desafio individual) e segundo o modelo-base:  $X = u + \epsilon$ ?

1.  $X$ : Este é o preço de uma determinada bebida.
2.  $\mu$ : Representa a média dos preços. Essencialmente, é o valor central em torno do qual os preços variam
3.  $\epsilon$ : Refere-se ao erro ou a variação não explicada nos preços. É a diferença entre o preço real ( $X$ ) e a média ( $\mu$ ).

Ao analisar esse modelo, percebe-se que os preços individuais das duas marcas se desviam da média com uma pequena variação. Dessa forma, os erros são relativamente pequenos.

Excel:

[https://docs.google.com/spreadsheets/d/1VQyHeFzhS0AYyxCji1y\\_O3nDGpQkCJayLICIJqYvjN0/edit#gid=0](https://docs.google.com/spreadsheets/d/1VQyHeFzhS0AYyxCji1y_O3nDGpQkCJayLICIJqYvjN0/edit#gid=0)

## 2. Estatística descritiva e exploratória

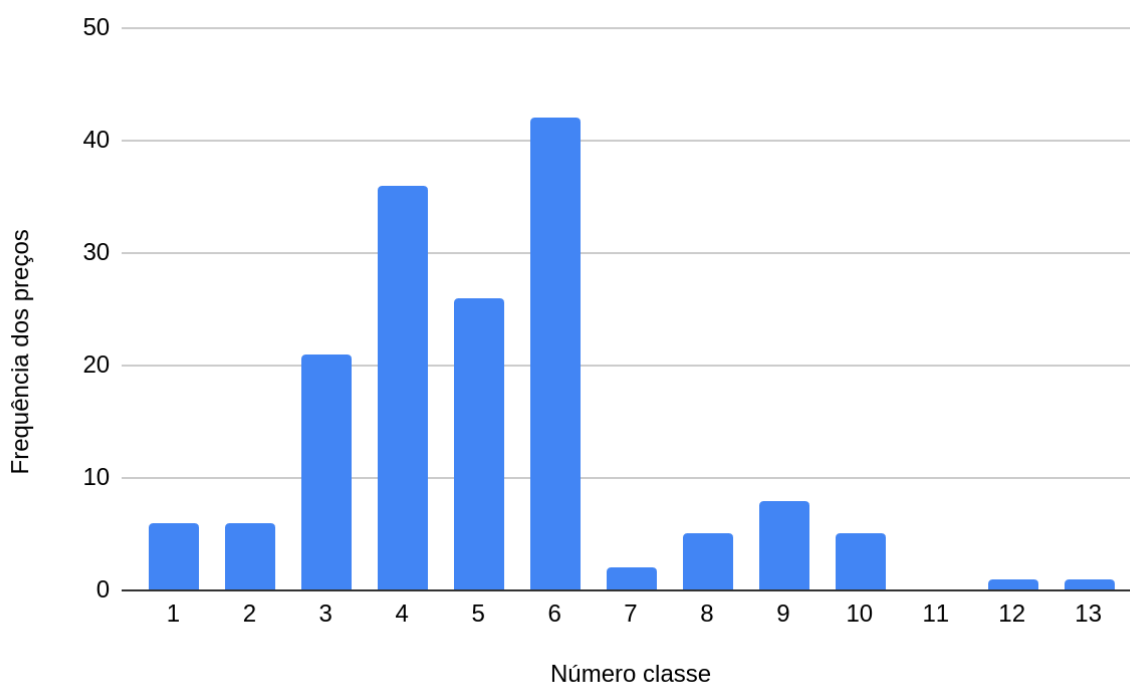
### 2.1. Modelos empíricos

- **Para Brahma duplo malte:**

- Menor  $X = 3,150$
- Maior  $X = 5,000$
- Range = 1,8500
- $k = \text{raiz}(n) = 13$
- $C = R/k = 0,1423$

Número classe	Limite Inf	Limite Sup	Frequência acumulada	Fa	Fr	Xi	Xi * Fr	QDP
1	3,1500	3,2923	6	6	0,0377	3,2212	0,1216	0,0127
2	3,2923	3,4346	12	6	0,0377	3,3635	0,1269	0,0073
3	3,4346	3,5769	33	21	0,1321	3,5058	0,4630	0,0116
4	3,5769	3,7192	69	36	0,2264	3,6481	0,8260	0,0054
5	3,7192	3,8615	95	26	0,1635	3,7904	0,6198	0,0000
6	3,8615	4,0038	137	42	0,2642	3,9327	1,0388	0,0045
7	4,0038	4,1462	139	2	0,0126	4,0750	0,0513	0,0009
8	4,1462	4,2885	144	5	0,0314	4,2173	0,1326	0,0054
9	4,2885	4,4308	152	8	0,0503	4,3596	0,2194	0,0156
10	4,4308	4,5731	157	5	0,0314	4,5019	0,1416	0,0154
11	4,5731	4,7154	157	0	0,0000	4,6442	0,0000	0,0000
12	4,7154	4,8577	158	1	0,0063	4,7865	0,0301	0,0061
13	4,8577	5,0000	159	1	0,0063	4,9288	0,0310	0,0080
TOTAL				159	1,0000		3,8020	0,0930
							MÉDIA	VARIÂNCIA
								0,3049
								DESVPAD

Os preços concentram-se na classe 6, ou seja, a maior parte está dentro da faixa de valores de 3,8615 a 4,0048. Existem poucos preços acima de 7,00 reais.

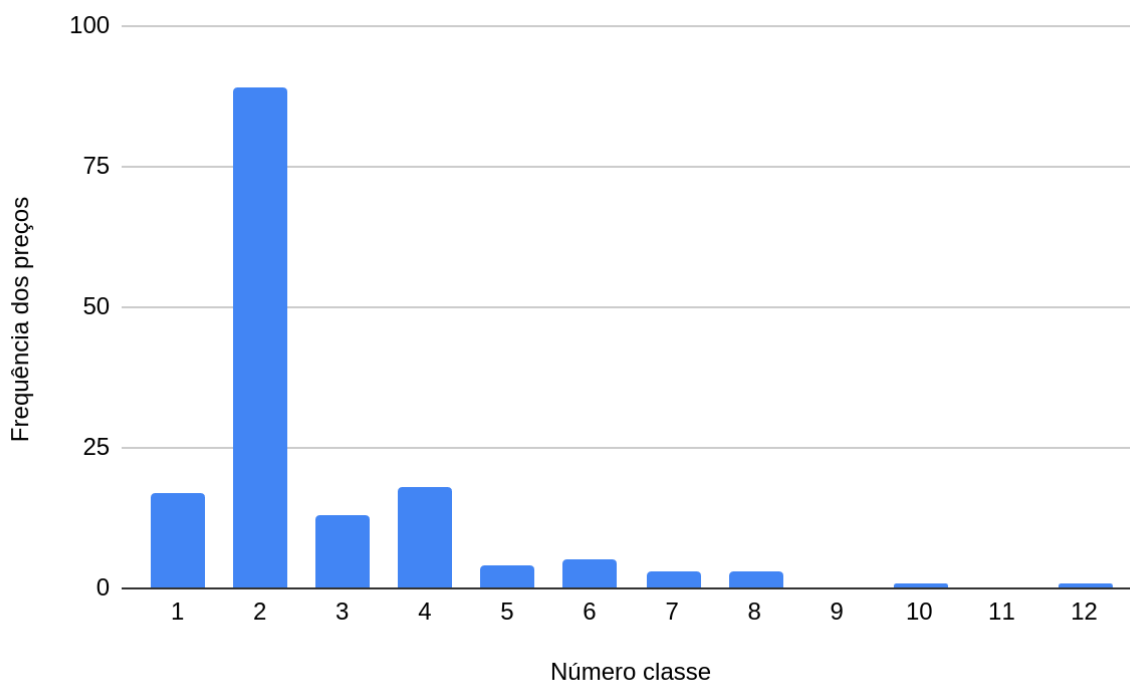


- **Para Schin Pilsen:**

- Menor X = 2,3500
- Maior X = 4,5000
- Range = 2,1500
- $k = \text{raiz}(n) = 12$
- $C = R/k = 0,1792$

Número classe	Limite Inf	Limite Sup	Frequência acumulada	Fa	Fr	Xi	Xi * Fr	QDP
1	2,3500	2,5292	17	17	0,1104	2,4396	0,2693	0,0108
2	2,5292	2,7083	106	89	0,5779	2,6188	1,5134	0,0103
3	2,7083	2,8875	119	13	0,0844	2,7979	0,2362	0,0002
4	2,8875	3,0667	137	18	0,1169	2,9771	0,3480	0,0059
5	3,0667	3,2458	141	4	0,0260	3,1563	0,0820	0,0042
6	3,2458	3,4250	146	5	0,0325	3,3354	0,1083	0,0110
7	3,4250	3,6042	149	3	0,0195	3,5146	0,0685	0,0113
8	3,6042	3,7833	152	3	0,0195	3,6938	0,0720	0,0173
9	3,7833	3,9625	152	0	0,0000	3,8729	0,0000	0,0000
10	3,9625	4,1417	153	1	0,0065	4,0521	0,0263	0,0110
11	4,1417	4,3208	153	0	0,0000	4,2313	0,0000	0,0000
12	4,3208	4,5000	154	1	0,0065	4,4104	0,0286	0,0178
TOTAL				154	1,0000		2,7525	0,0999
							MÉDIA	VARIÂNCIA
								0,3160
								DESVPAD

Os preços concentram-se na classe 2, ou seja, a maior parte está dentro da faixa de valores de 2,5292 a 2,7083. Existem poucos preços acima de 5,00 reais.



## 2.2. Estatísticas descritivas

### Brahma duplo malte

1) DADOS AGRUPADOS	
Média =	3,8020
Mediana =	3,7607
Moda =	3,9327
Variância =	0,0930
DesvPad =	0,3049
ErroPadMedia :	0,0242
Coef Variacao =	0,0802
Assimetria =	0,1355
1) DADOS NÃO AGRUPADOS	
Média =	3,8031
Mediana =	3,79
Moda =	3,99
Variância =	0,0970
DesvPad =	0,3115
ErroPadMedia :	0,0247
Coef Variacao =	0,0819
Assimetria =	0,0422

### Schin Pilsen

1) DADOS AGRUPADOS	
Média =	2,7525
Mediana =	2,65
Moda =	2,6188
Variância =	0,0999
DesvPad =	0,3160
ErroPadMedia :	0,0255
Coef Variacao =	0,1148
Assimetria =	0,3245
1) DADOS NÃO AGRUPADOS	
Média =	2,7772
Mediana =	2,69
Moda =	2,69
Variância =	0,0904
DesvPad =	0,3007
ErroPadMedia :	0,0242
Coef Variacao =	0,1083
Assimetria =	0,2900

## 2.3. Erro relativo

### Brahma duplo malte

Erro média =	0,0296
Erro mediana =	0,7731
Erro moda =	1,4361
Erro variância =	4,1903
Erro DesvPad =	2,1176
ErroPadMedia :	2,1176
Erro CoefVariac	2,0886
Erro Assimetria	-221,1477

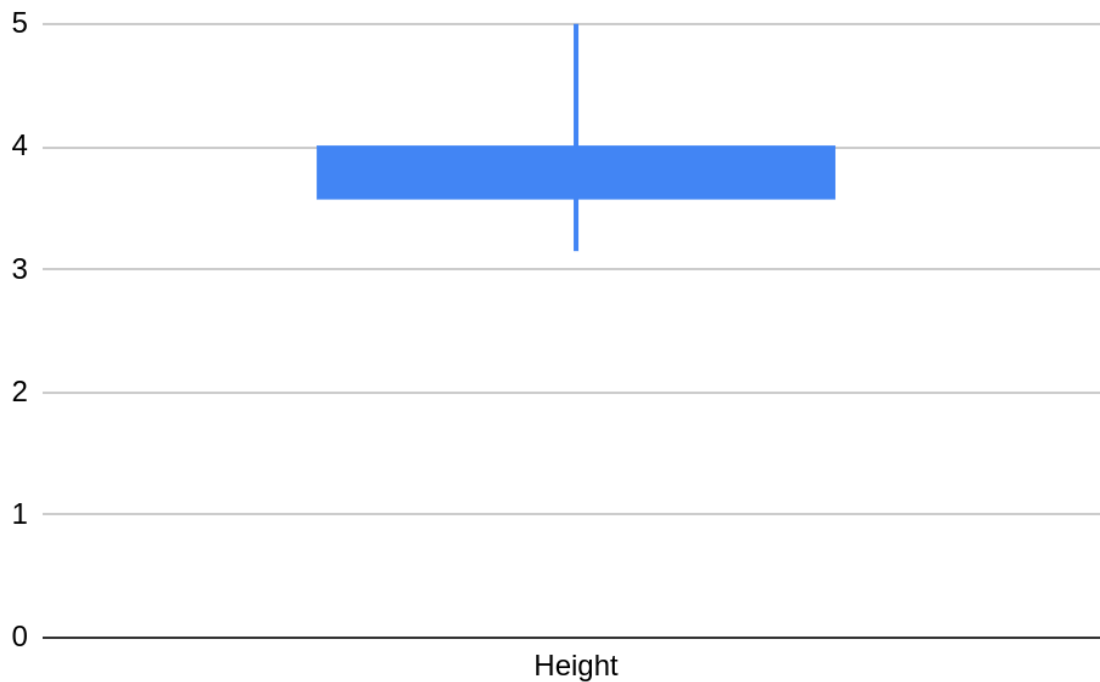
### Schin Pilsen

Erro média =	0,8881
Erro mediana =	1,4870
Erro moda =	2,6468
Erro variância =	-10,4780
Erro DesvPad =	-5,1085
ErroPadMedia :	-5,1085
Erro CoefVariac	-6,0503
Erro Assimetria	-11,8701

## 2.4. Diagrama de caixas

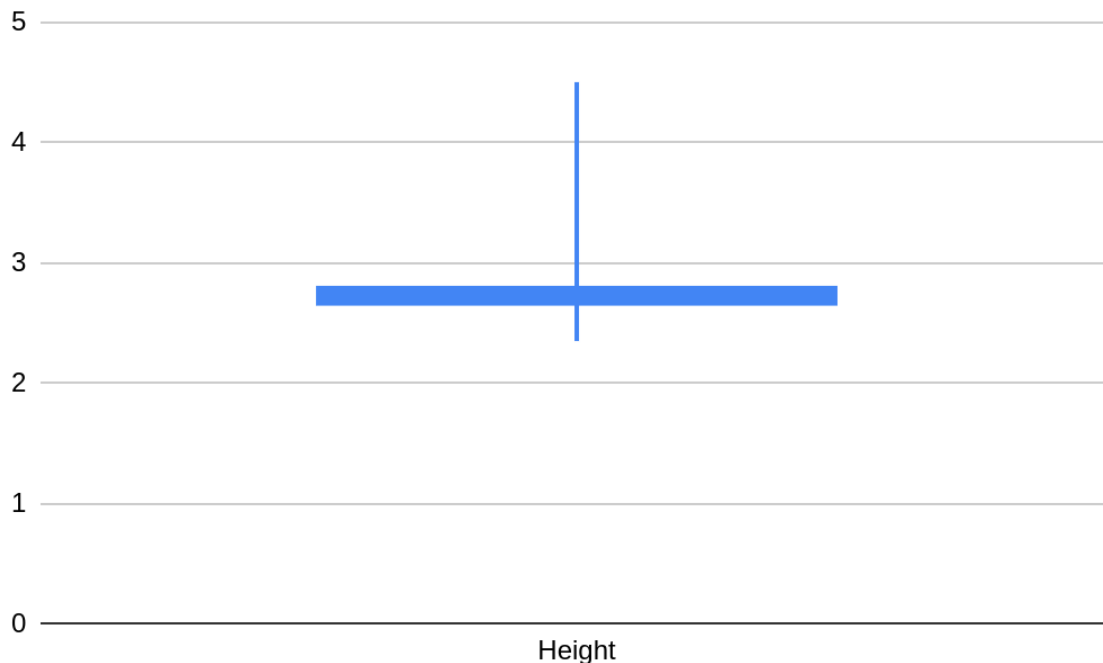
- **Para Brahma duplo malte:**

Existe outsider, pois existem valores acima de 4,59, indicando do ponto de vista do consumidor que os valores estão mais altos do que o esperado. Não existem outliers, indicando que os valores não são muito discrepantes



- **Para Schin Pilsen:**

Existe outlier, pois existem valores que ultrapassam o 3,1725, indicando do ponto de vista do consumidor valores muito caros para a bebida.



### 3. Estatística não paramétrica

- **Para Brahma duplo malte:**
- **SITUAÇÃO DE INTERESSE:** A situação de interesse é verificar se os dados de preços de bebidas seguem uma distribuição normal;
- **HIPÓTESE NULA:**  $H_0$  : Os dados mostram aderência à frequência esperada da distribuição normal;
- **HIPÓTESE ALTERNATIVA:**  $H_1$  : Os dados não mostram aderência à frequência esperada da distribuição normal;
- **ESTATÍSTICA:** A estatística do teste no caso do qui-quadrado é calculada usando a fórmula:  $\chi^2 = \sum ((O_i - E_i)^2 / E_i)$ ;
- **CRITÉRIO REJEIÇÃO  $H_0$ :** Se o  $\chi^2$  calculado for maior que o valor crítico, a hipótese nula é rejeitada. O valor crítico depende do número de graus de liberdade, que é geralmente  $k-1$ , onde  $k$  é o número de categorias ou intervalos.

QQ:

Z (Lim Inf)	Z (Lim Sup)	P (Z <= LI)	P(Z <= LS)	Pi	Ei (Esperado)	Oi (Observ)	QQ
-2,1384	-1,6717	0,0162	0,04729	0,03105	4,9372	6	0,2288
-1,6717	-1,2050	0,0473	0,11411	0,06682	10,6237	6	2,0124
-1,2050	-0,7382	0,1141	0,23018	0,11607	18,4559	21	0,3507
-0,7382	-0,2715	0,2302	0,39300	0,16281	25,8870	36	3,9508
-0,2715	0,1952	0,3930	0,57738	0,18439	29,3174	26	0,3754
0,1952	0,6619	0,5774	0,74599	0,16861	26,8085	42	8,6086
0,6619	1,1286	0,7460	0,87047	0,12449	19,7933	2	15,9954
1,1286	1,5954	0,8705	0,94468	0,07421	11,7992	5	3,9180
1,5954	2,0621	0,9447	0,98040	0,03572	5,6788	8	0,9488
2,0621	2,5288	0,9804	0,99428	0,01388	2,2065	5	3,5366
2,5288	2,9955	0,9943	0,99863	0,00435	0,6921	0	0,6921
2,9955	3,4622	0,9986	0,99973	0,00110	0,1752	1	3,8822
3,4622	3,9290	0,9997	0,99996	0,00023	0,0358	1	25,9640
						159	70,4635
						Calc QQ	

Novo Ei	Oi (Observ)	QQ
15,5609	12	0,8149
18,4559	21	0,3507
25,8870	36	3,9508
29,3174	26	0,3754
26,8085	42	8,6086
19,7933	2	15,9954
11,7992	5	3,9180
8,7885	15	4,3902
	159	38,4039
		Calc QQ

Com 5 graus de liberdade (8-2-1) e nível de significância de 0,05 temos 11,07 como área

Como  $\chi^2 = 38,4039$ , então cai na região de rejeição. O teste rejeita  $H_0$  em favor de  $H_1$ .

QQ tab: 11,0705

Distribuição logNormal:

2.7 DISTRIBUIÇÃO LOG NORMAL							
ln Z (Lim Inf)	ln Z (Lim Sup)	P (Z <= LI)	P(Z <= LS)	Pi	Ei (Esperado)	Oi (Observ)	QQ
-8,7062	-8,5613	0,0000	0,00000	0,00000	0,0000	6	0
-8,5613	-8,4225	0,0000	0,00000	0,00000	0,0000	6	0
-8,4225	-8,2894	0,0000	0,00000	0,00000	0,0000	21	0
-8,2894	-8,1614	0,0000	0,00000	0,00000	0,0000	36	0
-8,1614	-8,0383	0,0000	0,00000	0,00000	0,0000	26	0
-8,0383	-7,9196	0,0000	0,00000	0,00000	0,0000	42	0
-7,9196	-7,8050	0,0000	0,00000	0,00000	0,0000	2	0
-7,8050	-7,6944	0,0000	0,00000	0,00000	0,0000	5	0
-7,6944	-7,5873	0,0000	0,00000	0,00000	0,0000	8	0
-7,5873	-7,4836	0,0000	0,00000	0,00000	0,0000	5	0
-7,4836	-7,3831	0,0000	0,00000	0,00000	0,0000	0	0
-7,3831	-7,2856	0,0000	0,00000	0,00000	0,0000	1	0
-7,2856	-7,1909	0,0000	0,00000	0,00000	0,0000	1	0
						0	
						Calc QQ	

- **CONCLUSÃO:** Os dados observados não se encaixam bem na distribuição normal esperada. Há evidências estatísticas de que as frequências observadas diferem significativamente das frequências esperadas de acordo



com a distribuição normal. Também não se encaixa bem na distribuição logNormal.

- **Para Schin Pilsen:**
- **SITUAÇÃO DE INTERESSE:** A situação de interesse é verificar se os dados de preços de bebidas seguem uma distribuição normal;
- **HIPÓTESE NULA:**  $H_0$  : Os dados mostram aderência à frequência esperada da distribuição normal;
- **HIPÓTESE ALTERNATIVA:**  $H_1$  : Os dados não mostram aderência à frequência esperada da distribuição normal;
- **ESTATÍSTICA:** A estatística do teste no caso do qui-quadrado é calculada usando a fórmula:  $\chi^2 = \sum ((O_i - E_i)^2 / E_i)$ ;
- **CRITÉRIO REJEIÇÃO  $H_0$ :** Se o  $\chi^2$  calculado for maior que o valor crítico, a hipótese nula é rejeitada. O valor crítico depende do número de graus de liberdade, que é geralmente  $k-1$ , onde  $k$  é o número de categorias ou intervalos.

QQ:

Z (Lim Inf)	Z (Lim Sup)	P (Z <= Li)	P(Z <= LS)	Pi	Ei (Esperado)	Oi (Observ)	QQ
-1,2738	-0,7068	0,1014	0,23983	0,13846	21,3233	17	0,8766
-0,7068	-0,1399	0,2398	0,44437	0,20454	31,4987	89	104,9696
-0,1399	0,4270	0,4444	0,66533	0,22096	34,0271	13	12,9937
0,4270	0,9940	0,6653	0,83988	0,17456	26,8819	18	2,9346
0,9940	1,5609	0,8399	0,94073	0,10084	15,5300	4	8,5603
1,5609	2,1279	0,9407	0,98333	0,04260	6,5599	5	0,3709
2,1279	2,6948	0,9833	0,99648	0,01315	2,0255	3	0,4688
2,6948	3,2617	0,9965	0,99945	0,00297	0,4571	3	14,1485
3,2617	3,8287	0,9994	0,99994	0,00049	0,0753	0	0,0753
3,8287	4,3956	0,9999	0,99999	0,00006	0,0091	1	108,2687
4,3956	4,9626	1,0000	1,00000	0,00001	0,0008	0	0,0008
4,9626	5,5295	1,0000	1,00000	0,00000	0,0001	1	19569,0474
						154	19822,7153
							Calc QQ

Distribuição logNormal:

ln Z (Lim Inf)	ln Z (Lim Sup)	P (Z <= Li)	P(Z <= Ls)	Pi	Ei (Esperado)	Oi (Observ)	QQ
-6,0063	-5,7738	0,0000	0,00000	0,00000	0,0000	17	0
-5,7738	-5,5572	0,0000	0,00000	0,00000	0,0000	89	0
-5,5572	-5,3545	0,0000	0,00000	0,00000	0,0000	13	0
-5,3545	-5,1640	0,0000	0,00000	0,00000	0,0000	18	0
-5,1640	-4,9843	0,0000	0,00000	0,00000	0,0000	4	0
-4,9843	-4,8143	0,0000	0,00000	0,00000	0,0001	5	0
-4,8143	-4,6530	0,0000	0,00000	0,00000	0,0001	3	0
-4,6530	-4,4995	0,0000	0,00000	0,00000	0,0003	3	0
-4,4995	-4,3530	0,0000	0,00001	0,00000	0,0005	0	0
-4,3530	-4,2131	0,0000	0,00001	0,00001	0,0009	1	0
-4,2131	-4,0791	0,0000	0,00002	0,00001	0,0016	0	0
-4,0791	-3,9505	0,0000	0,00004	0,00002	0,0026	1	0
							0
							Calc QQ

- **CONCLUSÃO:** Os dados observados não se encaixam bem na distribuição normal esperada. Há evidências estatísticas de que as frequências observadas diferem significativamente das frequências esperadas de acordo com a distribuição normal. Também não se encaixa bem na distribuição logNormal.

#### 4. Estatística paramétrica

##### 4.1) Razão de variâncias

- **SITUAÇÃO DE INTERESSE:** A situação de interesse é verificar a razão de variâncias entre as marcas.
- **HIPÓTESE NULA:** H0: As variâncias de preços para as duas marcas são iguais.
- **HIPÓTESE ALTERNATIVA:** H1: As variâncias de preço das duas marcas são diferentes.
- **ESTATÍSTICA:** A estatística do teste é calculada por:  $f = s_1^2 / s_2^2$  Teste F
- **CRITÉRIO REJEIÇÃO H0:** Rejeita a hipótese nula se o valor-p <  $\alpha$  ou rejeita a hipótese nula se o valor absoluto da estatística de teste (f) for maior que o valor crítico.
- **CÁLCULOS:**

Var marca 1:	0,0930	denominador
Var marca 2:	0,0999	numerador
f =	1,0742	
gl no numerad =	154	
gl no denomin =	163	
Valor p:	0,6717	
significancia = 0,05, então $\alpha=0,025$ (bilateral)		
fc =	1,3034	

- **CONCLUSÃO:** Como  $f < f_c$ , o teste aceita  $H_0$ . Portanto as variâncias são iguais. Pela outra abordagem: valor de  $p > 0,05$ , então aceite  $H_0$ .

#### 4.2) DIFERENÇAS DE MÉDIAS ENTRE AS MARCAS

- **SITUAÇÃO DE INTERESSE:** A situação de interesse é verificar a diferença de média entre as marcas.
- **HIPÓTESE NULA:**  $H_0$ : As diferenças de médias de preços para as duas marcas são iguais.
- **HIPÓTESE ALTERNATIVA:**  $H_1$ : As diferenças de médias de preços para as duas marcas são diferentes.
- **ESTATÍSTICA:** A estatística do teste é calculada pelo teste T:

$$t = (\text{media}(X_1) - \text{media}(X_2)) / (S_a * \text{raiz}(1/n_1 + 1/n_2))$$

$$S_a^2 = ((n_1-1) s_1^2 + (n_2-1) s_2^2) / (n_1 + n_2 - 2)$$

- **CRITÉRIO REJEIÇÃO H0:** Rejeita a hipótese nula se o valor-p < α ou rejeita a hipótese nula se o valor absoluto da estatística de teste (t) for maior que o valor crítico.
- **CÁLCULOS:**

Média marca 1:	3,8020			
Média marca 2:	2,7525			
1) Dif médias:	1,0495			
2) Calc denom:	0,0351			
3) Calc t:	29,9019			
4) Calc tc:	1,6498			
gl: 159 + 154 - 2	311			
Valor p:	0,0000	Como teste é bilateral a área deve ser dobrada		
PELA ABORDAGEM CLÁSSICA:				
a = 0,05/2 = 0,025				
Pela tabela 4, com 311 graus de liberdade, tc = 1,649767922				
t > tc, H0 rejeitado				

- **CONCLUSÃO:** Como p é menor que o valor de significância (0.05), então H0 é rejeitado e aceita-se H1 Como t > tc, o teste rejeita H0. Assim, existem diferenças de preços entre as marcas.

**4.3) Construir um Intervalo de Confiança de 95% para a razão entre variâncias e para a diferença de médias de sua tarefa individual, modelos empíricos construídos.**

- Para razão de variâncias:

Fórmula para a razão de variâncias de duas distribuições normais.			
Razão das variâncias $\sigma_1^2/\sigma_2^2$ de duas distribuições normais	$\frac{s_1^2}{s_2^2}$	$\frac{s_1^2}{s_2^2} f_{1-\alpha/2, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} f_{\alpha/2, n_2-1, n_1-1}$	
		em que $f_{1-\alpha/2, n_2-1, n_1-1} = \frac{1}{f_{\alpha/2, n_1-1, n_2-1}}$	
Tomando-se um nível de confiança de 95%:			
[F a/2 (n2-1, n1-1) * S1/S2, F 1- a/2 (n2-1, n1-1) * S1/S2]		Var marca 1:	0,0930
		Var marca 2:	0,0999
F[0.05/2 (11, 12) * 0,0999/0,0930, F 1-0,05/2 (11, 12) * 0,0999/0,0930]		n1:	13
		n2:	12
IC = [	3,5680	0,31322	]

- Para diferença de médias:

Fórmula para a diferença das médias populacionais de duas distribuições normais			
diferentes e desconhecidas.			
Diferença nas médias de duas distribuições normais $\mu_1 - \mu_2$ , com variâncias $\sigma_1^2 \neq \sigma_2^2$ e desconhecidas	$\bar{x}_1 - \bar{x}_2$	$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2$	
		$\leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	
		em que $v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 + 1} + \frac{(s_2^2/n_2)^2}{n_2 + 1}} - 2$	
1) Grau de liberdade corrigido:	20,6774	Média 1:	3,8020
		Média 2:	2,7525
2) Valor Crítico da Distribuição t:	0,7281		
3) Intervalo de Confiança:	IC = [	0,9589	1,1400 ]

## 5) CONCLUSÕES GERAIS

5.1) IV– O que V pode concluir da análise dos preços, do ponto de vista do consumidor, quanto às partes anteriores de estatística descritiva/exploratória, testes de hipóteses e os intervalos de confiança para a razão entre variâncias

## **e a diferença de médias? Qual a diferença e qual a semelhança entre os dois métodos (Testes de hipóteses e Intervalo de confiança)?**

Com base nos resultados dos testes e análises, e do comportamento dos preços do ponto de vista do consumidor, as duas marcas possuem estatísticas parecidas, a média, a variância e o desvio padrão são valores bem próximos. Pelo próprio histograma pode-se verificar que as duas marcas não seguem a distribuição normal. Também nenhuma delas também se adere à distribuição LogNormal.

Em relação às estatísticas descritivas e exploratórias, as duas possuem valores baixos para a variância e desvio padrão, os gráficos de caixas mostram poucos valores fora do padrão.

Os testes de hipótese mostraram a não aderência à distribuição normal pelas marcas, mostrando valores do QQ bem acima do esperado.

Em relação às estatísticas paramétricas, verifica-se que existem diferenças entre as médias, mas as variâncias se mantêm bem próximas. Já o intervalo de confiança indica os intervalos para a diferença de médias em razão de variâncias.

### **Teste de Hipótese:**

- Objetivo: Determinar se uma afirmação (hipótese nula) é verdadeira ou não com base nos dados.
- Interpretação: Rejeitar a hipótese nula sugere que existe uma diferença ou efeito significativo.
- Exemplo: No caso da razão de variâncias, o teste de hipótese verificará se as variâncias são iguais ou diferentes.

### **Intervalo de Confiança:**

- Objetivo: Estimar a faixa de valores em que um parâmetro populacional específico provavelmente está.
- Interpretação: Se o intervalo de confiança contém zero (no caso de diferenças de médias), sugere que não há uma diferença significativa.

- Exemplo: O intervalo de confiança para a diferença de médias indica uma faixa de valores prováveis para a diferença real entre as médias."