

Análise Estatística dos Filmes de Comédia dos EUA na década de 2000 e 2010

1. Análise da variável Score (IMDB)

Questão 1: A avaliação média dos filmes de comédia aumentou?

Minitab: Estat > Estatísticas Básicas > Exibição de Estatísticas Descritivas

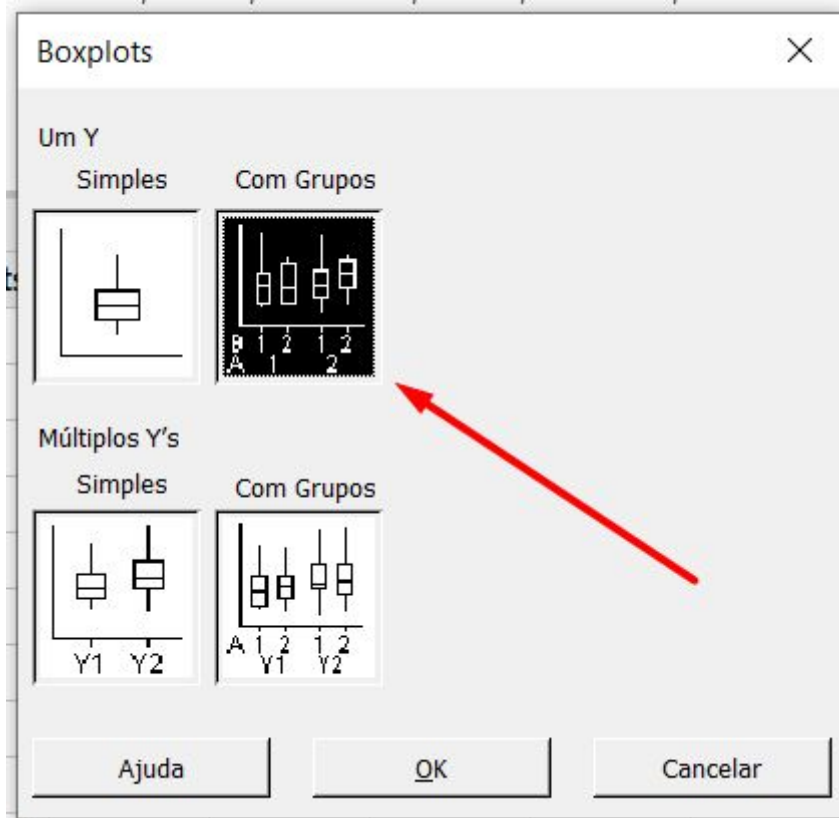
Estatísticas

Variável	year	N	N*	Média	EP	Média Desv	Pad	Mínimo	Q1	Mediana	Q3	Máximo
score	2000	72	0	6,215	0,113	0,962	3,600	5,600	6,250	6,900	8,300	
	2010	67	0	6,2612	0,0986	0,8073	3,8000	5,7000	6,3000	6,8000	7,7000	

Figura 1.

Box Plot

Minitab: Grafico > BoxPlot



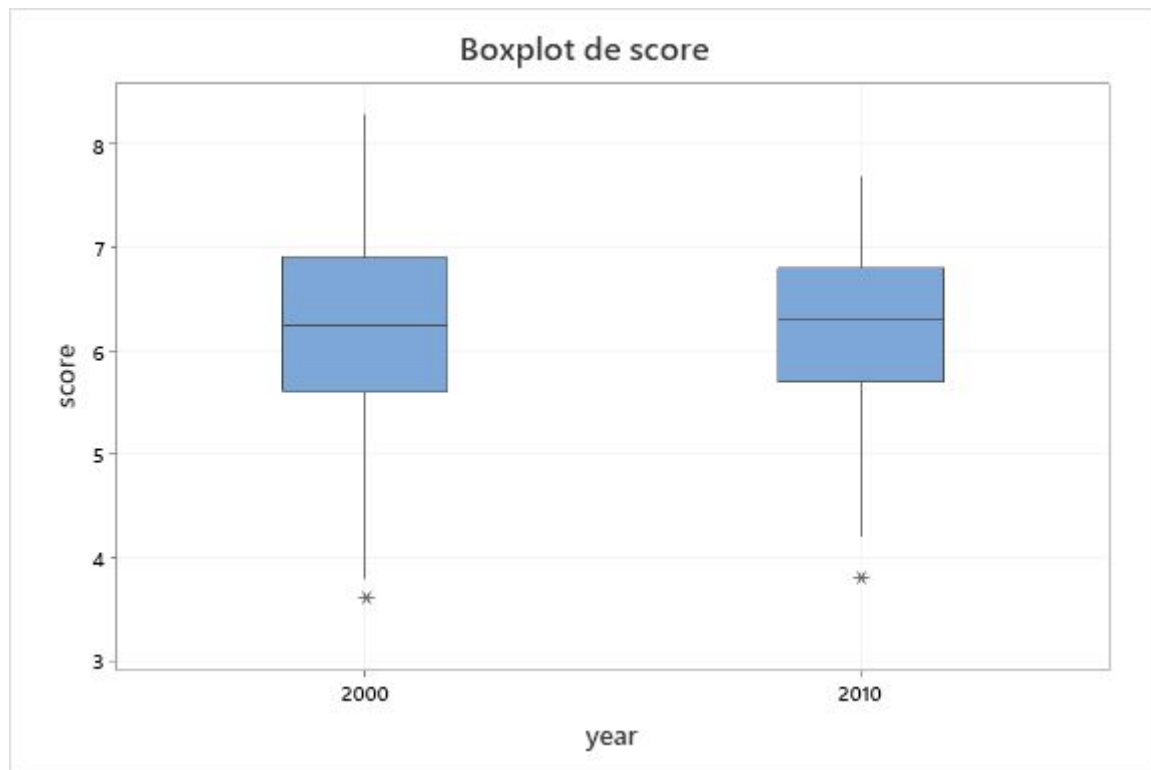


Figura 2.

Através da análise descritiva da Figura 1, pode-se perceber que a média das avaliações dos filmes é bem próxima uma da outra. Porém, apenas com essa verificação, não é possível concluir que as médias se mostraram constantes. A Figura 2 apresenta o gráfico Box Plot para ter uma análise visual inicial do comportamento das amostras. Pelo gráfico, pode-se inferir a presença de outliers, mas que não apresentam problemas para análise.

Para analisar corretamente a questão estabelecida, deve-se executar um teste estatístico apropriado. A primeira consideração a ser feita para uma análise mais sofisticada, seria relativa a normalidade e homocedasticidade (variância constante) das amostras utilizadas. A Figura 3 apresenta o gráfico do teste de normalidade com nível de significância a 5%.

Teste de Normalidade

Minitab: Estat > Estatísticas Básicas > Teste de Normalidade

H0: Dados possuem distribuição normal

H1: Dados não possuem distribuição normal

Significância = 5% (= 0,05)

Teste de Normalidade

Variável:

Linhas de Percentil

☒ Nenhum

☐ Em valores de Y:

☐ Em valores de dados:

Testes para Normalidade

☐ Anderson-Darling

☒ Ryan-Joiner (Similar ao Shapiro-Wilk)

☐ Kolmogorov-Smirnov

Selecionar

Título:

Ajuda OK Cancelar

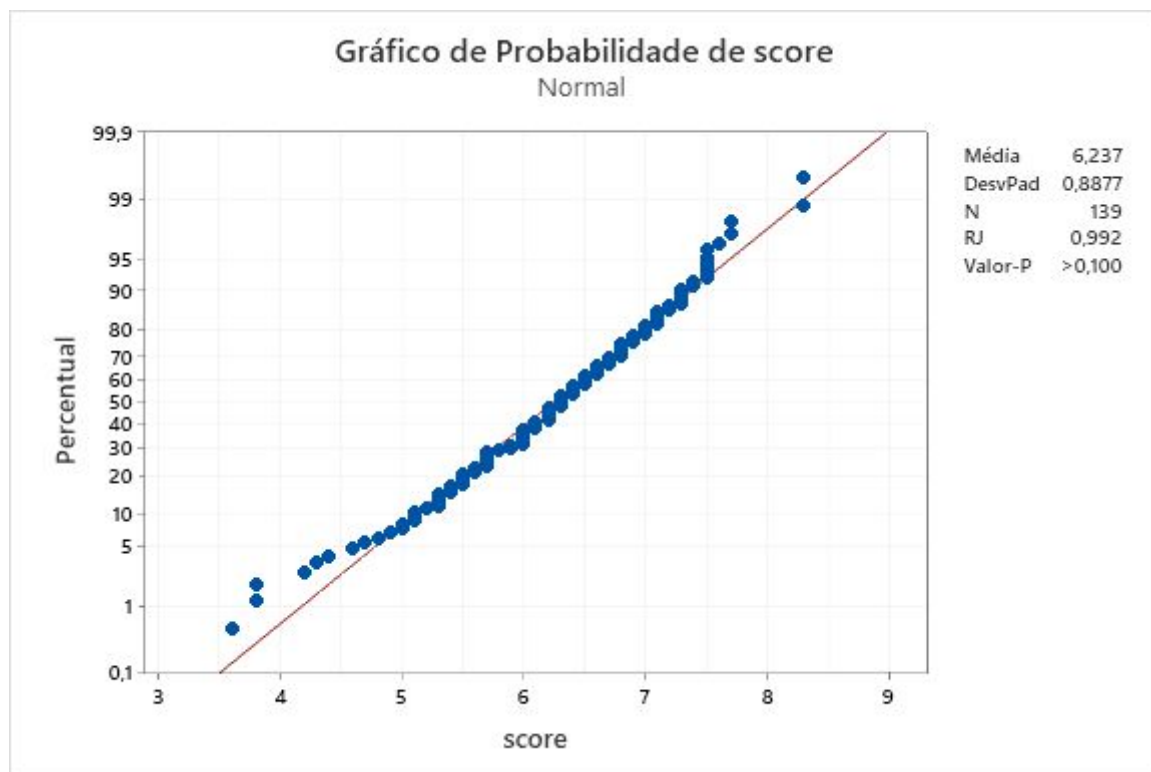


Figura 3.

Como valor- $p > 0,100$ é superior ao nível de significância estabelecido de 5% (0,05), pode-se então aceitar a Hipótese nula de que os dados possuem distribuição normal.

Para utilização de um teste paramétrico na comparação das médias, além do teste de normalidade, precisa-se efetuar o teste de homocedasticidade (variâncias iguais).

Teste de Homocedasticidade (Igualdade de variâncias)

Minitab: Estat > Estatísticas Básicas > Teste para 2 Variâncias

H0: Dados possuem igualdade de variâncias

H1: Dados não possuem igualdade de variâncias

Significância = 5%

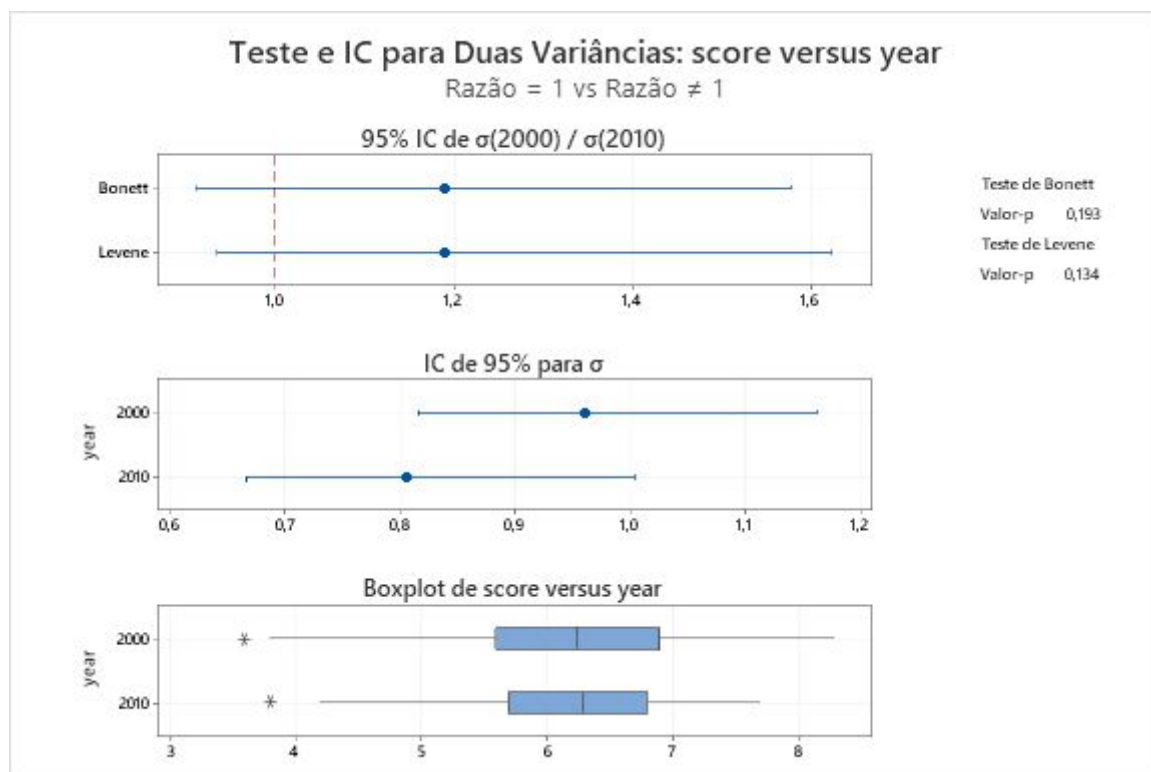
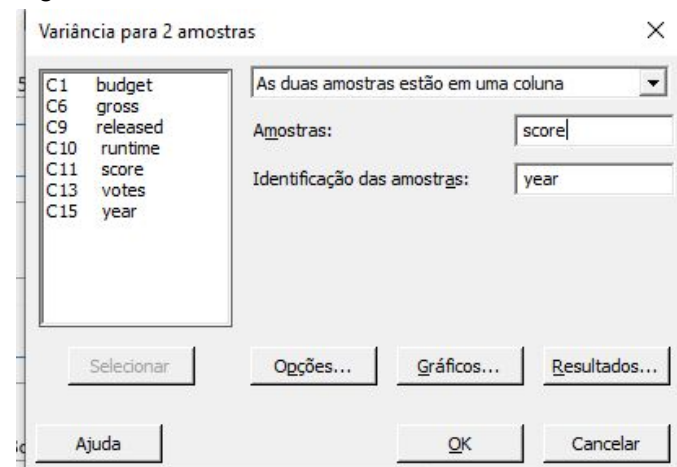


Figura 4.

Aplicando o Teste de Levene conforme a Figura 4, pode-se observar um valor-p = 0,134. Como esse valor é superior ao nível de significância estabelecido de 5%, pode-se aceitar a hipótese nula de que os dados são homocedásticos, ou seja, possuem igualdade de variâncias.

Uma vez constatado que os dados possuem distribuição normal e são homocedásticos, pode-se utilizar um método paramétrico para a comparação das médias. Como o estudo possui um delineamento experimental de 1 fator e 2 tratamentos, um método importante para comparar dois grupos é o Teste T de Student.

Comparação das Médias

Minitab: Estat > Estatísticas Básicas > Teste t para 2 amostras

H0: Médias são iguais

H1: Médias são diferentes

Significância = 5%

Teste		
Hipótese nula	$H_0: \mu_1 - \mu_2 = 0$	
Hipótese alternativa	$H_1: \mu_1 - \mu_2 \neq 0$	
Valor-T	GL	Valor-p
-0,31	135	0,760

Figura 5.

Utilizando o Test T, pode-se observar na Figura 5 um valor-p = 0,760 que é superior ao nível de significância estabelecido, portanto pode-se aceitar a hipótese nula de que não existe diferença significativa entre as médias a um nível de significância de 5%.

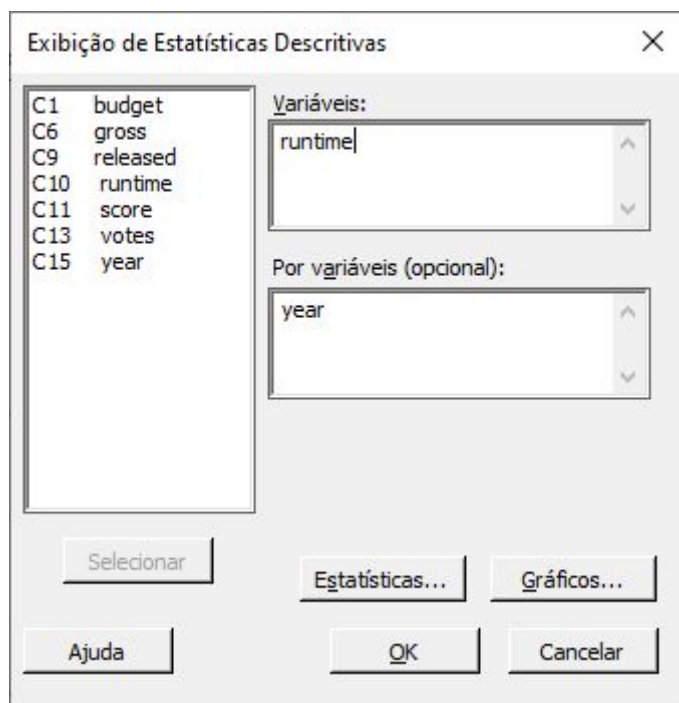
Com essa análise, pode-se concluir que estatisticamente a média de score dos filmes de comédia dos Estados Unidos em 2000 e 2010 são iguais. Esse estudo só é válido para esse conjunto de dados.

Dataset utilizado disponível em: github.com/jjthegomes/analise_estatistica_2020.

2. Análise Estatística de Dados Variável Runtime (Duração em minutos)

Questão 2: O tempo de duração médio dos filmes de comédia aumentou?

Minitab: Estat > Estatísticas Básicas > Exibição de Estatísticas Descritivas

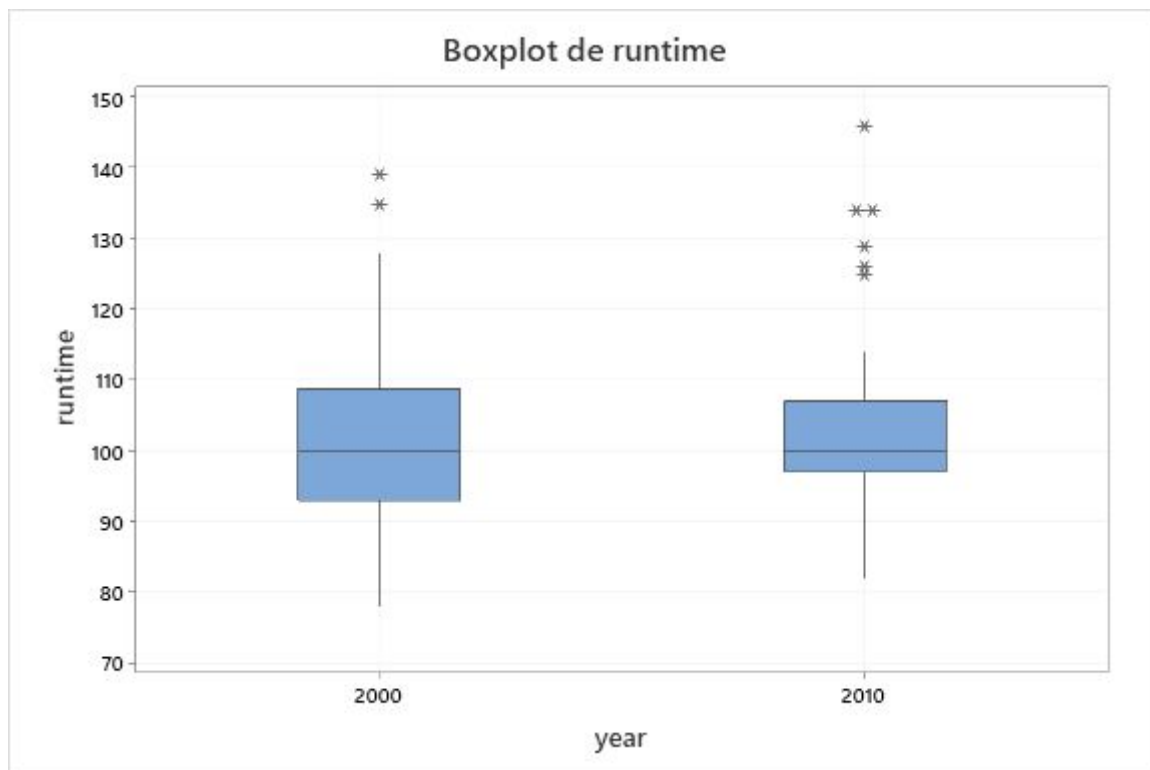
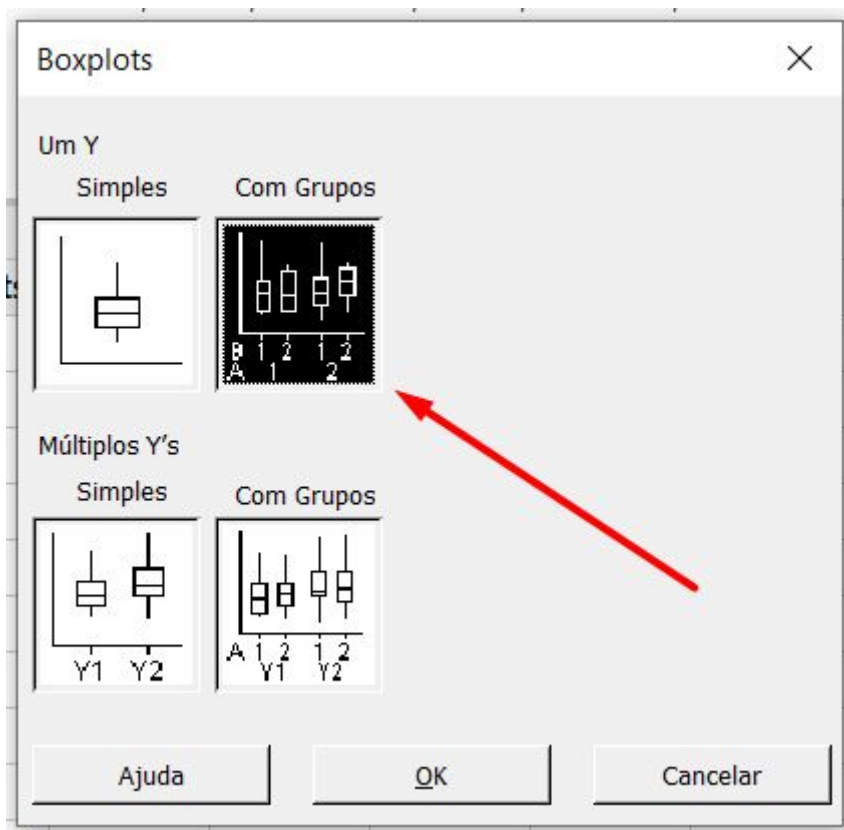


Estatísticas

Variável	year	N	N*	Média	EP	Média	DesvPad	Mínimo	Q1	Mediana	Q3	Máximo
runtime	2000	72	0	101,53	1,51	12,82	78,00	93,00	100,00	108,75	139,00	
	2010	67	0	102,58	1,45	11,90	82,00	97,00	100,00	107,00	146,00	

Box Plot

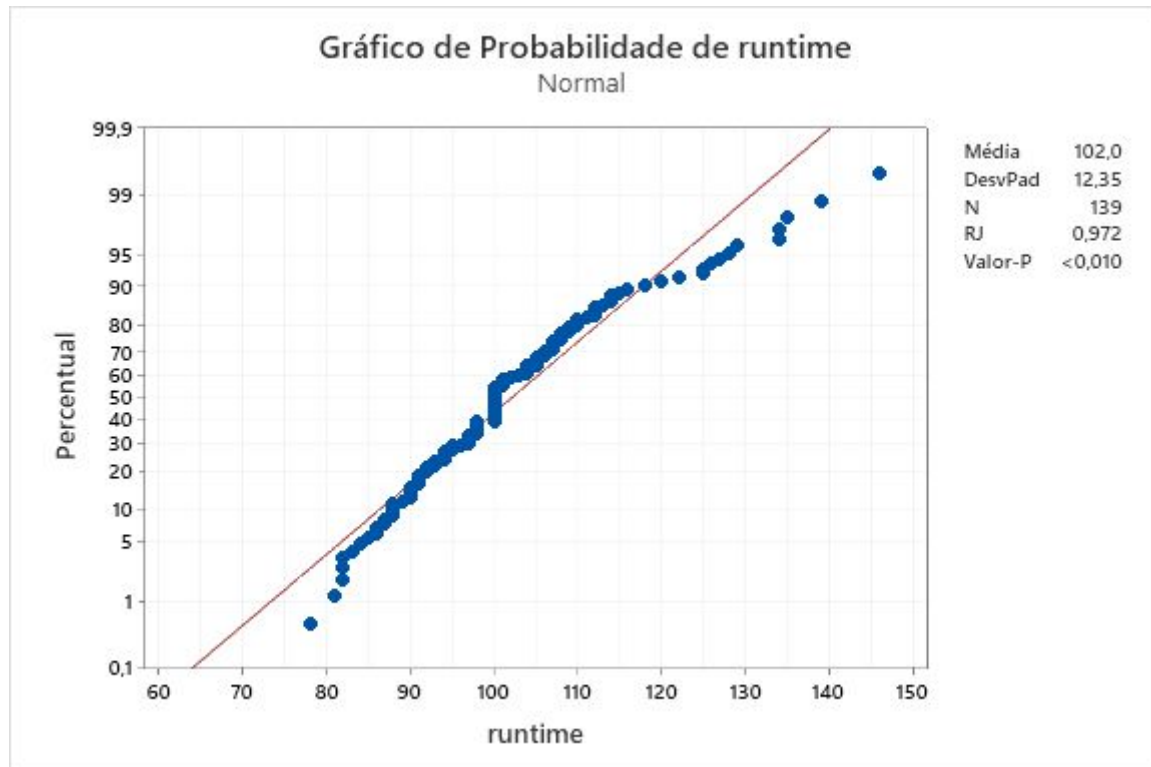
Minitab: Grafico > BoxPlot



Teste de Normalidade

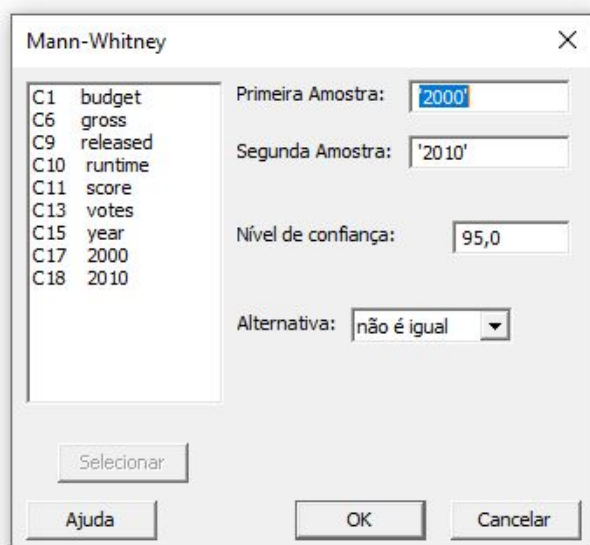
Minitab: Estat > Estatísticas Básicas > Teste de Normalidade

H0: Dados possuem distribuição normal
H1: Dados não possuem distribuição normal
Significância = 5% (= 0,05)



Como valor- $p < 0,010$ é inferior ao nível de significância estabelecido de 5% (0,05), pode-se então aceitar a Hipótese alternativa de que os dados **não** possuem distribuição normal. Para utilização de um teste não paramétrico na comparação das médias, precisa-se efetuar o teste de Mann-Whitney.

Minitab: Estat > Não-paramétrico > Mann-Whitney



H0: Dados possuem igualdade de variâncias

H1: Dados não possuem igualdade de variâncias

Significância = 5%

Teste

Hipótese nula $H_0: \eta_1 - \eta_2 = 0$

Hipótese alternativa $H_1: \eta_1 - \eta_2 \neq 0$

Método	Valor W	Valor-p
Não ajustado para empates	4917,50	0,607
Ajustado para empates	4917,50	0,606

Utilizando o de Mann-Whitney, pode-se observar um valor-p = 0,607 que é superior ao nível de significância estabelecido, pode-se aceitar a hipótese nula de que não existe diferença significativa entre as médias a um nível de significância de 5%.