Exemplo 1.
Numa turma do 10º ano de uma determinada escola secundária, os alunos registaram o número de irmãos, tendo-se obtido a seguinte amostra:

1	2	2	1	3	0	0	1	1	2
1	1	1	0	0	3	4	3	1	2

Esta variável (número de irmãos) é uma variável discreta e, como tal, uma representação gráfica adequada é o diagrama de barras.

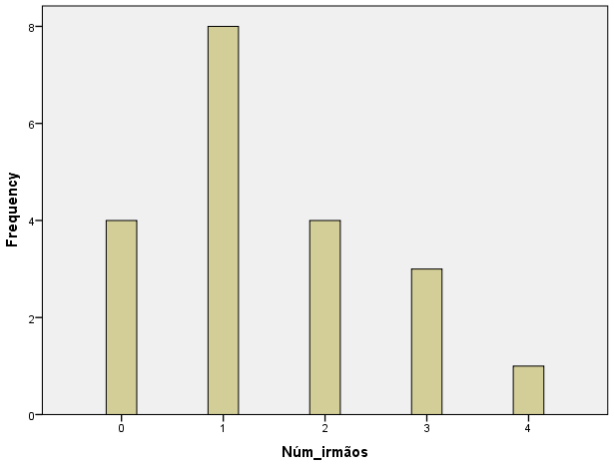
Núm_irmãos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	20.0	20.0	20.0
	1	8	40.0	40.0	60.0
	2	4	20.0	20.0	80.0
	3	3	15.0	15.0	95.0
	4	1	5.0	5.0	100.0
	Total	20	100.0	100.0	

Statistics

Núm_irmãos

N	Valid	20
	Missing	0
Mean		1.45
Median		1.00
Mode		1
Std. Deviation		1.146
Variance		1.313
Range		4
Minimum		0
Maximum		4
Percentiles	25	1.00
	50	1.00
	75	2.00



Exemplo 2.
No final de uma sessão de um filme de *Harry Potter* interrogaram-se 48 espectadores sobre a sua idade, tendo-se obtido os seguintes valores:

32	34	33	12	57	13	58	16
23	23	62	65	35	15	17	20
14	11	51	33	31	13	11	58
23	10	63	34	12	15	62	13
40	11	18	62	64	30	42	20
21	56	11	51	38	49	15	21

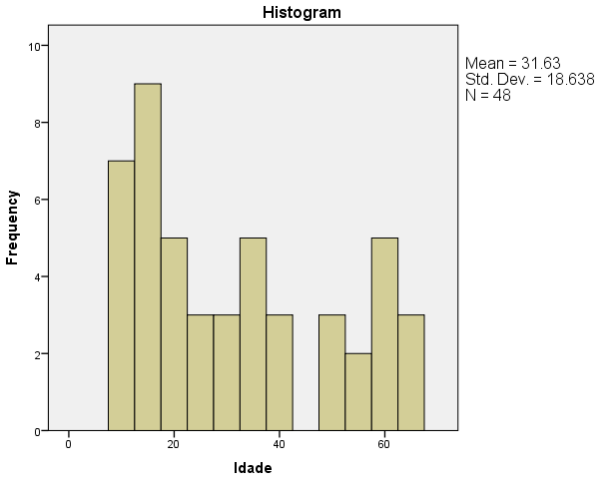
Estes dados são de natureza contínua. Assim, uma representação gráfica adequada é o histograma.

Exemplos para as aulas Teóricas

Statistics

Idade

N	Valid	48
	Missing	0
Mean		31.63
Median		26.50
Std. Deviation		18.638
Variance		347.388
Skewness		.547
Kurtosis		-1.171
Range		55
Minimum		10
Maximum		65
Percentiles	25	15.00
	50	26.50
	75	50.50

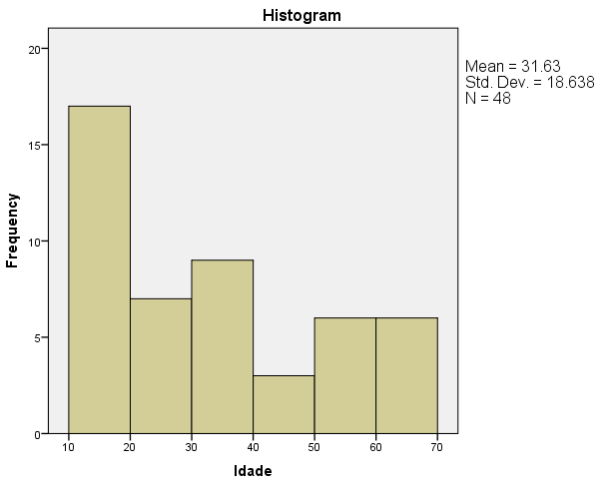


Em primeiro lugar é necessário escolher um número de classes, seja k, para classificar os dados.
Regra de *Sturges*:

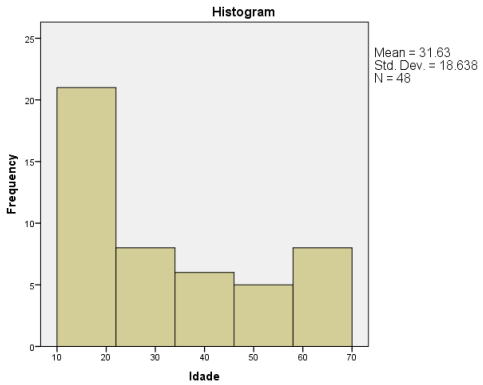
$k = \lceil \log_2 n \rceil + 1 = \lceil \ln(n)/\ln(2) \rceil + 1$ (aproximadamente o mesmo que tomar o menor inteiro k : $2^k \geq n$)

$k = \lceil \log_2 48 \rceil + 1 = \lceil 5.58 \rceil + 1 = 6$ ($2^5 = 32 < 48; 2^6 = 64 > 48 \Rightarrow k = 6$)

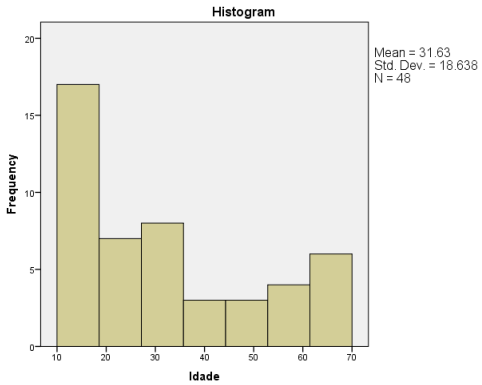
k = 6



k = 5

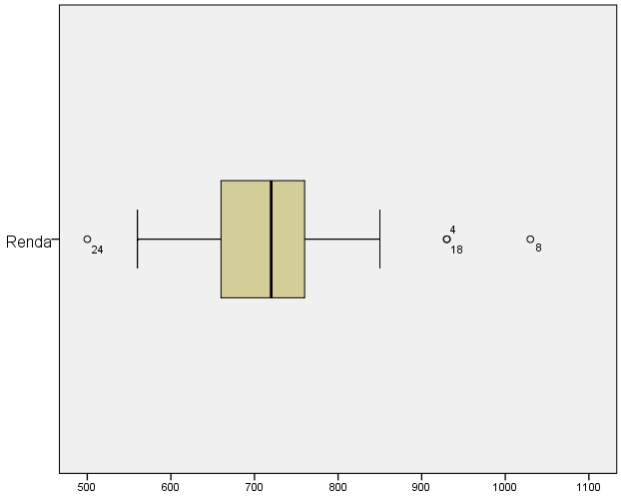


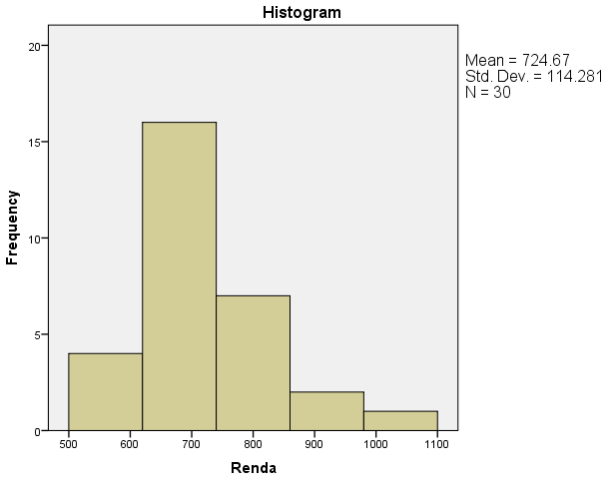
k = 7



Exemplo 3.
Os dados seguintes referem-se à renda de casa (em euros) mensal, de 30 estudantes (fora de residências para estudantes):

730	730	730	930	700	570	690	1030	740	620
720	670	560	740	650	660	850	930	600	620
760	690	710	500	730	800	820	840	720	700

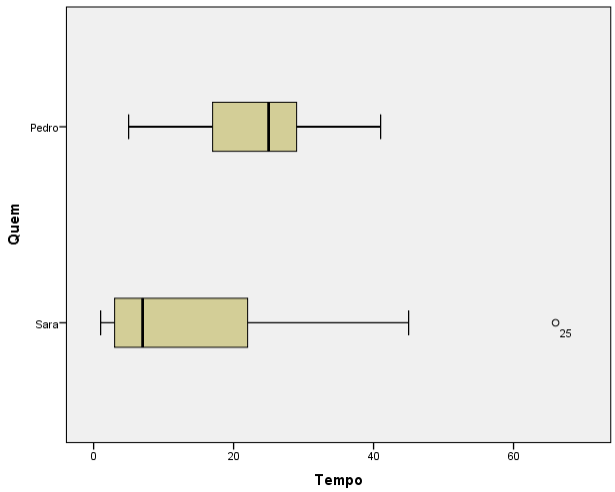




Exemplo 4.
A Sara e o Pedro dividem um plano de 1000 minutos pelos seus telemóveis. Para comparar os tempos das chamadas de cada um deles recolheram os tempos (em segundos) das chamadas durante um dia, tendo obtido:

Sara:	1	1	1	1	2	3	3	3	5
	5	6	6	7	8	8	12	14	14
	22	23	29	33	38	45	66		
Pedro:	5	8	9	14	17	21	23	23	24
	26	27	27	28	29	31	33	39	41

Se pretendermos apenas comparar os tempos gastos em chamadas pela Sara e pelo Pedro, podemos construir *Boxplots* paralelos:



Exemplo 5.

Num laboratório que contém equipamento polarográfico recolheu-se 6 amostras de pó a várias distâncias do polarógrafo e registou-se a concentração de mercúrio (ConcHg) em cada amostra, tendo-se obtido:

Distância	ConcHg
1,4	2,4
3,8	2,5
7,5	1,3
10,2	1,3
11,7	0,7
15	1,2

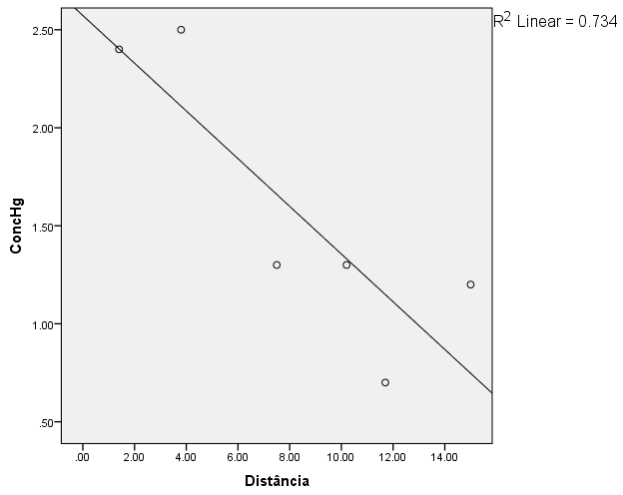
Pretende averiguar-se a possibilidade de a contaminação de mercúrio resultar do polarógrafo.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.573	.347		7.419	.002
	Distância	-.122	.037	-.857	-3.325	.029

a. Dependent Variable: ConcHg

Recta dos MQ: $\text{ConcHg} = -0.122 \times \text{Distância} + 2.573$



Exemplo 6.
Admita que a mediana da nota de Matemática dos alunos do 10ºano de determinada escola no ano 2008/2009 (numa escala de 0 a 100) foi de 35. Com base na amostra seguinte de notas de alunos do mesmo ano e da mesma escola em 2010/2011 verifique se há razões para suspeitar que os alunos do 10ºano dessa escola têm tendência a ser mais fracos na disciplina de Matemática.

23	43	36	17	25	47	39	26	50	63
27	45	48	29	6	38	37	62	10	51

R:

1. Formular H_0 e H_1
2. Indicar a Estatística de teste (E.T.) e a sua distribuição sob a validade de H_0
3. Encontrar o valor observado (V.O.) da E.T.
4. Determinar a região de rejeição (R.R.) ao nível α ($\mathcal{R}_\alpha$) ou o valor-p e decidir
5. Concluir

p – proporção de indivíduos com nota ≤ 35

Pretende testar-se: $H_0 : p = 0.5$ *vs* $H_1 : p > 0.5$

E.T.: X – v.a. que rep. o nº de indivíduos, em 20, com nota ≤ 35; H_0 verd. => $X \cap Bi(20, 0.5)$
V.O.: $r_0 = 8$
R.R.: Rejeita-se H_0 para valores "grandes" de X, ou seja, rejeita-se H_0 ao nível α se $X \geq r_\alpha$ =>
encontrar o menor r_α tal que $P(X \geq r_\alpha \mid H_0 \text{ verd.}) \leq \alpha$
Consultando a tabela da Binomial vem:
$P(Bi(20, 0.5) \geq 13) = 1 - P(Bi(20, 0.5) \leq 12) = 1 - 0.8684 = 0.1316$
$P(Bi(20, 0.5) \geq 14) = 1 - P(Bi(20, 0.5) \leq 13) = 1 - 0.9423 = 0.0577$
$P(Bi(20, 0.5) \geq 15) = 1 - P(Bi(20, 0.5) \leq 14) = 1 - 0.9793 = 0.0207$
$P(Bi(20, 0.5) \geq 16) = 1 - P(Bi(20, 0.5) \leq 15) = 1 - 0.9941 = 0.0059$
$\alpha = 5\%$: $\mathcal{R}_{0.05} = \{r: r \geq 15\} = \{15, 16, \ldots, 20\}$; $8 \notin \mathcal{R}_{0.05}$ => Não se rejeita H_0 a 5% =>
Não se rejeita H_0 a 1%
$\alpha = 10\%$: $\mathcal{R}_{0.1} = \{r: r \geq 14\} = \{14, 15, \ldots, 20\}$; $8 \notin \mathcal{R}_{0.1}$ => Não se rejeita H_0 a 10%
$\therefore$ Não se rejeita H_0 aos níveis usuais

ou

Valor-p: p-*value* = $P(Bi(20, 0.5) \geq 8) = 1 - P(Bi(20, 0.5) \leq 7) = 1 - 0.1316 = 0.8684$
p-*value* > α, para todo o α usual
$\therefore$ Não se rejeita H_0 aos níveis usuais

Conclusão: Não há evidência para afirmar que os novos alunos sejam mais fracos.

Exemplo 7.
Um estudo envolvendo um determinado modelo automóvel de gama média-alta permitiu concluir que, metade dos automóveis desse modelo no 3º ano de vida (para veículos que percorrem cerca de 15000 Km/ano) tinham um custo de manutenção de 295 euros. Numa amostra constituída por 130 automóveis da mesma gama que se encontravam nas condições referidas, mas de outra marca, encontraram-se 53 cujo custo de manutenção no 3º ano foi superior a 295 euros. Será possível concluir, que os custos de manutenção são diferentes nas duas marcas?

R:
p – proporção de veículos da "outra" marca cujo custo de manutenção é superior a 295€

$H_0 : p = 0.5$ ($p \geq 0.5$) *vs* $H_1 : p < 0.5$

E.T.: X – v.a. que rep. o nº de veículos, em 130, com custo de manutenção superior a 295€;
H_0 verd. => $X \cap Bi(130, 0.5)$
V.O.: $r_0 = 53$

Valor-p: p-*value* = $P(Bi(130, 0.5) \leq 53) \underset{TLC}{\approx} \Phi\left(\frac{53 - 65}{\sqrt{32.5}}\right) = \Phi(-2.105) = 1 - \underset{\approx 2.11}{\Phi(2.105)} = 0.01743$

Rejeita-se H_0 para níveis $\alpha \geq$ p-*value*
0.01 < p-*value* => Não se rejeita H_0 a 1%
0.05 > p-*value* => Rejeita-se H_0 a 5% e a 10%

Conclusão: Para α = 1%, não há evidência para afirmar os veículos da "outra" marca tenham custos de manutenção mais baratos.

Para α = 5% e α = 10%, há evidência de que os veículos da "outra" marca tenham custos de manutenção mais baratos.

Exemplo 8.
Em 2009 realizou-se uma sondagem sobre as preferências radiofónicas dos ouvintes do período da manhã (7h-10h) e concluiu-se que, na zona de Lisboa, as preferências eram as seguintes:

Rádio Comercial	TSF	Antena 1	RFM	Antena 2
35%	25%	20%	15%	5%

No início deste mês de Setembro recolheu-se uma amostra de 300 indivíduos da zona de Lisboa e ouvintes no período da manhã, tendo-se obtido:

Rádio Comercial	TSF	Antena 1	RFM	Antena 2
98	80	55	40	27

Verifique se as preferências radiofónicas se mantiveram.

R:
p_1 – proporção de indivíduos da zona de Lisboa que no período da manhã preferem a RC
p_2 – " " " " " " " " " " " " preferem a TSF
p_3 – " " " " " " " " " " " " preferem a A1
p_4 – " " " " " " " " " " " " preferem a RFM
p_5 – " " " " " " " " " " " " preferem a A2

$H_0 : p_1 = 0.35, p_2 = 0.25, p_3 = 0.20, p_4 = 0.15, p_5 = 0.05$
vs
$H_1 : p_1 \neq 0.35 \vee p_2 \neq 0.25 \vee p_3 \neq 0.20 \vee p_4 \neq 0.15 \vee p_5 \neq 0.05$

E.T.: $X^2 = \sum_{i=1}^{5} \frac{(o_i - e_i)^2}{e_i}$; H_0 verd. => $X^2 \sim \chi^2_{(4)}$

V.O.:

o_i	e_i
98	105
80	75
55	60
40	45
27	15

$x_0^2 = \frac{7^2}{105} + \frac{5^2}{75} + \frac{5^2}{60} + \frac{5^2}{45} + \frac{12^2}{15} \approx 11.372$

Valor-p: p-*value* = $P(\chi^2_{(4)} \geq 11.372) = 1 - P(\chi^2_{(4)} \leq 11.372)$

Pela tabela: $0.975 < P(\chi^2_{(4)} \leq 11.372) < 0.99$ => 0.01 < p-*value* < 0.025

0.01 < p-*value* => Não se rejeita H_0 a 1%
0.05 > p-*value* => Rejeita-se H_0 a 5% e a 10%

Conclusão: Para α = 1%, não há alteração nas preferências radiofónicas.

Para α = 5% e α = 10%, há alteração nas preferências radiofónicas.

Exemplo 9.
O médico responsável pelo gabinete médico de uma fábrica registou o nº de acidentes, por mês, verificados nessa fábrica durante os últimos 10 anos:

Nº acidentes/mês:	0	1	2	3	4	5	6	7	≥8
Nº de meses:	2	10	15	30	28	15	10	6	4

Pensa-se que o número de acidentes por mês nessa fábrica é uma variável aleatória com distribuição de *Poisson* de valor médio 4. Verifique se existem razões que nos levem a duvidar desta suposição.

R:
Y – v.a. que rep. o número de acidentes num mês (escolhido ao acaso)

$H_0 : Y \cap Poi(4)$ *vs* $H_1 : Y \not\cap Poi(4)$

$n = 120$
$p_1 = P(Poi(4) = 0) = 0.0183 \Rightarrow e_1 = 2.196^{(*)}$
$p_2 = P(Poi(4) = 1) = 0.0733 \Rightarrow e_2 = 8.796$
$p_3 = P(Poi(4) = 2) = 0.1465 \Rightarrow e_3 = 17.58$
$p_4 = P(Poi(4) = 3) = 0.1954 \Rightarrow e_4 = 23.448$
$p_5 = P(Poi(4) = 4) = 0.1954 \Rightarrow e_5 = 23.448$
$p_6 = P(Poi(4) = 5) = 0.1563 \Rightarrow e_6 = 18.756$
$p_7 = P(Poi(4) = 6) = 0.1042 \Rightarrow e_7 = 12.504$
$p_8 = P(Poi(4) = 7) = 0.0595 \Rightarrow e_8 = 7.14$
$p_9 = P(Poi(4) \geq 8) = 1 - P(Poi(4) \leq 7) = 1 - 0.9489 = 0.0511 \Rightarrow e_9 = 6.132$
$^{(*)}$ $e_1 < 5 \Rightarrow$ juntar as classes 1 e 2

E.T.: $X^2 = \sum_{i=1}^{8} \frac{(o_i - e_i)^2}{e_i}$; H_0 verd. $\Rightarrow X^2 \sim \chi^2_{(7)}$

V.O.:

o_i	12	15	30	28	15	10	6	4
e_i	10.992	17.58	23.448	23.448	18.756	12.504	7.14	6.132

$$x_0^2 = \frac{(12-10.992)^2}{10.992} + \cdots + \frac{(4-6.132)^2}{6.132} \approx 5.362$$

Valor-p: p-*value* = $P(\chi^2_{(7)} \geq 5.362)$

$0.3 < P(\chi^2_{(7)} \leq 5.362) < 0.4 \Rightarrow 0.6 <$ p-*value* < 0.7

p-*value* $> \alpha$, para todo o α usual $\Rightarrow$ Não se rejeita H_0 para todo o α usual

Conclusão: Não existem razões que levem a duvidar da suposição.

Exemplo 10.
Verifique se podemos duvidar que o seguinte conjunto de dados provém de uma população com distribuição: *F(x) = 1-exp(-(x-1.03)/0.789)*.

1.10	1.62	2.46	1.55	3.08	1.05	1.13	2.25	2.04	1.23
2.83	3.09	1.9	1.96	1.27	2.06	1.57	1.19	1.03	1.97

R:
X – população de onde provêm os dados

$H_0 : X \cap F$ *vs* $H_1 : X \not\cap F$, $F(x) = 1-\exp(-(x-1.03)/0.789)$, $x \geq 1.03$

Caso 1: Utilizando as classes construídas para o histograma (*h* = 0.42)

	o_i		o_i
$C_1 = [1.03, 1.45)$	7	$C_4 = [2.29, 2.71)$	1
$C_2 = [1.45, 1.87)$	3	~~$C_5 = [2.71, 3.13)$~~	~~3~~
$C_3 = [1.87, 2.29)$	6	$C_5 = [2.71, +\infty)$	3

$n = 20$

$p_1 = P(X \in C_1) = F(1.45) - F(1.03) = 0.4128 \Rightarrow e_1 = 8.256$
$p_2 = P(X \in C_2) = F(1.87) - F(1.45) = 0.2424 \Rightarrow e_2 = 4.848^{(*)}$
$p_3 = P(X \in C_3) = F(2.29) - F(1.87) = 0.1423 \Rightarrow e_3 = 2.846^{(**)}$
$p_4 = P(X \in C_4) = F(2.71) - F(2.29) = 0.0836 \Rightarrow e_4 = 1.672^{(**)}$
$p_5 = 1 - p_1 - p_2 - p_3 - p_4 = 0.1189 \Rightarrow e_5 = 2.378^{(**)}$
$^{(*)}$ $e_2 < 5$ mas $e_2 \approx 5 \Rightarrow$ não se junta a classe 2 à anterior
$^{(*)}$ $e_5 < 5 \Rightarrow$ juntar as classes 5 e 4; $e_5 + e_4 < 5 \Rightarrow$ juntar as classes 5, 4 e 3

E.T.: $X^2 = \sum_{i=1}^{3} \frac{(o_i - e_i)^2}{e_i}$; H_0 verd. $\Rightarrow X^2 \sim \chi^2_{(2)}$

V.O.:

o_i	7	3	10
e_i	8.256	4.848	6.896

$$x_0^2 = \frac{1.256^2}{8.256} + \frac{1.848^2}{4.848} + \frac{3.104^2}{6.896} \approx 2.293$$

Valor-p: p-*value* = $P(\chi^2_{(2)} \geq 2.293)$

$0.6 < P(\chi^2_{(2)} \leq 2.293) < 0.7 \Rightarrow 0.3 < \text{p-}value < 0.4$

p-*value* > α, para todo o α usual => Não se rejeita H_0 para todo o α usual

Conclusão: Não existem razões que levem a duvidar de que F seja a f.d. de X.

Caso 2: Utilizando classes de igual probabilidade (sob a validade de H_0)

k classes A_i: $P(X \in A_i \mid H_0)=1/k$ e k: $n/k \geq 5 \Rightarrow k = [n/5] = 4$
A_i: $P(X \in A_i \mid H_0)=0.25$
$A_1 = [1.03, a_1)$; $P(1.03 \leq X \leq a_1 \mid H_0) = 0.25 <=> F(a_1) - F(1.03) = 0.25 <=> F(a_1) = 0.25$
$A_2 = [a_1, a_2)$; $P(a_1 \leq X \leq a_2 \mid H_0) = 0.25 <=> F(a_2) - F(a_1) = 0.25 <=> F(a_2) = 0.50$
$A_3 = [a_2, a_3)$; $P(a_2 \leq X \leq a_3 \mid H_0) = 0.25 <=> F(a_3) - F(a_2) = 0.25 <=> F(a_3) = 0.75$
$A_4 = [a_3, +\infty)$
a_1: $F(a_1) = 0.25 <=> 1-\exp(-(a_1-1.03)/0.789) = 0.25 <=> a_1 = 1.257$
a_2: $F(a_2) = 0.50 <=> 1-\exp(-(a_2-1.03)/0.789) = 0.50 <=> a_2 = 1.577$
a_3: $F(a_3) = 0.75 <=> 1-\exp(-(a_3-1.03)/0.789) = 0.75 <=> a_2 = 2.124$

	o_i		o_i
$A_1 = [1.03, 1.257)$	6	$A_3 = [1.577, 2.124)$	6
$A_2 = [1.257, 1.577)$	3	$A_4 = [2.124, +\infty)$	5

E.T.: $X^2 = \frac{4}{20}\sum_{i=1}^{4} o_i^2 - n$; H_0 verd. $\Rightarrow X^2 \sim \chi^2_{(3)}$

V.O.: $x_0^2 = \frac{1}{5}(6^2 + 3^2 + 6^2 + 5^2) - 20 = 1.2$

Valor-p: p-*value* = $P(\chi^2_{(3)} \geq 1.2)$

$0.2 < P(\chi^2_{(3)} \leq 1.2) < 0.3 \Rightarrow 0.7 < \text{p-}value < 0.8$

p-*value* > α, para todo o α usual => Não se rejeita H_0 para todo o α usual

Conclusão: Não existem razões que levem a duvidar de que F seja a f.d. de X.

Exemplo 11.
Selecionaram-se aleatoriamente 10 transístores de uma determinada marca e registaram-se os seus tempos de vida (em horas):

28.9 15.2 28.7 72.5 48.6 52.4 37.6 49.5 62.1 54.5

Verifique, para α=1%, se podemos duvidar que a distribuição do tempo de vida seja exponencial com valor médio 45 horas.

R:
X – v.a. que rep. o tempo de vida de um transístor (escolhido ao acaso)

$H_0 : X \cap Exp(1/45)$ *vs* $H_1 : X \not\cap Exp(1/45)$, $F(x) = 1 - e^{-x/45} = 1 - 0.978^x$, $x \geq 0$

Teste de ajustamento de *Kolmogorov/Smirnov*

x_i	$F(x_i)$	i	i-1	i/n	$(i-1)/n$	$i/n - F(x_i)$	$F(x_i) - (i-1)/n$
15.2	0.2866	1	0	0.1	0	-0.1866	0.2866
28.7	0.4715	2	1	0.2	0.1	-0.2715	0.3715
28.9	0.4739	3	2	0.3	0.2	-0.1739	0.2739
37.6	0.5664	4	3	0.4	0.3	-0.1664	0.2664
48.6	0.6604	5	4	0.5	0.4	-0.1604	0.2604
49.5	0.6671	6	5	0.6	0.5	-0.0671	0.1671
52.4	0.6879	7	6	0.7	0.6	0.0121	0.0879
54.5	0.7021	8	7	0.8	0.7	0.0979	0.0021
62.1	0.7484	9	8	0.9	0.8	0.1516	-0.0516
72.5	0.8003	10	9	1	0.9	0.1997	-0.0997
					$d_{10}^+ =$	0.1997	
						$d_{10}^- =$	0.3715

$d_{10} = 0.3715$
Para $\alpha = 1\%$, $d_{10,0.01} = 0.489$
Como $0.3715 < 0.489$, não se rejeita H_0 a 1%

Conclusão: Não existem razões que levem a duvidar de que a distribuição do tempo de vida de um transístor seja exponencial com valor esperado igual a 45 horas

Exemplo 12.
Num inquérito realizado a 1500 adultos da zona de Coimbra e a 1000 da zona de Lisboa, verificou-se que a percentagem de fumadores foi, respetivamente, 15,2% e 18,5%. Há evidência suficiente para afirmar que a proporção de fumadores é menor em Coimbra do que em Lisboa?

R:
p_1 – proporção de fumadores em Lisboa
p_2 – proporção de fumadores em Coimbra

$H_0 : p_1 \leq p_2$ *vs* $H_1 : p_1 > p_2$

E.T.: $Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$; H_0 verd. => $Z \sim N(0, 1)$

V.O.: $\hat{p}_1 = 0.185$; $\hat{p}_2 = 0.152$; $\bar{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \dfrac{185 + 228}{2500} = 0.1652$

$$z_0 = \frac{0.185 - 0.152}{\sqrt{0.1652 \times 0.8348\left(\frac{1}{1000} + \frac{1}{1500}\right)}} \approx 2.18$$

Valor-p: p-*value* = $P(N(0, 1) \geq 2.18) = 1 - \Phi(2.18) = 1 - 0.98537 = 0.01463$
Rejeita-se H_0 para níveis $\alpha \geq$ p-*value*
$0.01 <$ p-*value* => Não se rejeita H_0 a 1%
$0.05 >$ p-*value* => Rejeita-se H_0 a 5% e a 10%

Conclusão: Para $\alpha = 1\%$, não há evidência para afirmar que a proporção de fumadores seja menor em Coimbra do que em Lisboa.

Para $\alpha = 5\%$ e $\alpha = 10\%$, há evidência de que a proporção de fumadores seja menor em Coimbra do que em Lisboa.

Exemplo 13.
O gerente de uma dependência de determinado banco afirma que o número mediano de clientes dessa dependência não excede os 750 clientes. O Diretor dessa dependência duvida da afirmação do gerente e selecionou 16 dias, ao acaso, em que registou o número de clientes da citada dependência, tendo obtido:

775	754	745	756	765	753	750	760
801	739	777	782	742	751	769	789

Verifique se há razões para afirmar que o Diretor tem razão.

R:

$H_0 : \chi_{0.5} \leq 750$ *vs* $H_1 : \chi_{0.5} > 750$

E.T.: $S_n = \#\{i: x_i > 750\}$ (nº de observações > 750); H_0 verd. => $S_n \cap Bi(n, 0.5)$
V.O.: $n = 16 - 1 = 15$ (existe uma observações igual a 750)
$s_{15} = 12$
Valor-p: p-*value* = $P(Bi(15, 0.5) \geq 12) = 1 - P(Bi(15, 0.5) \leq 11) = 1 - 0.9824 = 0.0176$
Rejeita-se H_0 para níveis $\alpha \geq$ p-*value*
$0.01 <$ p-*value* => Não se rejeita H_0 a 1%
$0.05 >$ p-*value* => Rejeita-se H_0 a 5% e a 10%

Conclusão: Para $\alpha = 1\%$, não há evidência para afirmar que o Diretor tenha razão.

Para $\alpha = 5\%$ e $\alpha = 10\%$, há evidência de que o Diretor tenha razão.

Exemplo 14.
Pretende-se saber se há evidência para afirmar que há diferença entre os níveis de colesterol de dia (D) e de noite (N). Para tal, efetuaram-se medições em 8 indivíduos durante o dia (d_i) e durante a noite (n_i), tendo-se obtido os seguintes resultados:

d_i:	17.6	15.2	13.7	12.1	31.6	25.7	14.8	15.3
n_i:	18.2	13.6	15.6	11.6	30.7	27.8	17.9	16.6

R:
Não há diferença entre o nível de colesterol de dia (D) e de noite (N) se
$P(D > N) = P(D < N) = ½ \iff P(D - N > 0) = P(D - N < 0) = ½ \iff \chi_{0.5} (D - N) = 0$

$H_0 : \chi_{0.5} (D - N) = 0$ *vs* $H_1 : \chi_{0.5} (D - N) \neq 0$

E.T.: $S_n = \#\{i: d_i - n_i > 0\}$ (nº de diferenças positivas); H_0 verd. => $S_n \cap Bi(n, 0.5)$
V.O.: $n = 8$ (não existem "empates"); $s_8 = 3$
Valor-p: p-*value* = $2 \min \{ P(Bi(8, 0.5) \leq 3), P(Bi(8, 0.5) \geq 3) \} =$
$= 2 \min \{ P(Bi(8, 0.5) \leq 3), 1 - P(Bi(8, 0.5) \leq 2) \} = 2 \min \{ 0.3633, 1 - 0.1445 \} = 0.7266$
Rejeita-se H_0 para níveis $\alpha \geq$ p-*value*
p-*value* > α, para todo o α usual => Não se rejeita H_0 para todo o α usual

Conclusão: Não há evidência de que os níveis de colesterol de dia e de noite sejam diferentes.

Exemplo 15.
Foi administrado a 13 indivíduos um remédio com o objetivo de diminuir a pressão sistólica do sangue, tendo-se obtido os seguintes resultados:

Antes:	115	135	140	130	135	150	122
Depois:	120	128	142	112	111	150	110
Antes:	135	138	190	180	99	110	
Depois:	135	125	180	160	103	108	

Será que o remédio é eficaz?

R:
O remédio é eficaz se
$P(A > D) > 0.5 \iff P(A - D > 0) > 0.5 \iff \chi_{0.5}(A - D) > 0$

$H_0 : \chi_{0.5}(A - D) = 0$ *vs* $H_1 : \chi_{0.5}(A - D) > 0$

Admitindo que a distribuição de A – D é simétrica, vamos utilizar o teste de *Wilcoxon*.

$n = 13 - 2 = 11$ (existem 2 diferenças nulas)

$X_i = A_i - D_i$	$Y_{i:n}$	R_i
–5	(–) 2	1.5
7	2	1.5
–2	(–) 4	3
18	(–) 5	4
24	7	5
12	10	6
13	12	7
10	13	8
20	18	9
–4	20	10
2	24	11

E.T.: T_{11}^+ é igual à soma das ordens correspondentes aos sinais positivos
V.O.: $t_{11}^+ = 1.5 + 5 + \ldots + 11 = 57.5$
R.R.: Rejeita-se H_0 para valores "grandes" de T_{11}^+, ou seja, rejeita-se H_0 ao nível α se $t_{11}^+ \geq c_\alpha$, onde c_α é tal que $P(T_{11}^+ \geq c_\alpha) \leq \alpha$.
$P(T_{11}^+ \geq c_\alpha) = P(T_{11}^+ > c_\alpha - 1)$
Consultando a tabela dos quantis da estatística de *Wilcoxon* vem:

$$c_\alpha - 1 = \begin{cases} 66-8 & \alpha = 0.01 \\ 66-14 & \alpha = 0.05 \\ 66-18 & \alpha = 0.1 \end{cases} \Rightarrow c_\alpha = \begin{cases} 59 & \alpha = 0.01 \\ 53 & \alpha = 0.05 \\ 49 & \alpha = 0.1 \end{cases}$$

Então, rejeita-se H_0 se:
$T_{11}^+ \geq 59$ para $\alpha = 1\%$
$T_{11}^+ \geq 53$ para $\alpha = 5\%$
$T_{11}^+ \geq 49$ para $\alpha = 10\%$
Como $t_{11}^+ = 57.5$, rejeita-se H_0 a 5% e a 10% e não se rejeita a 1%

Conclusão: A 1% não há evidência de que o remédio seja eficaz, mas a 5% e a 10% já há essa evidência.

Exemplo 14 Revisitado

Retomemos o Exemplo 14 e admitamos que a distribuição das diferenças é simétrica.
Vamos testar

$H_0 : \chi_{0.5}\,(D - N) = 0 \quad vs \quad H_1 : \chi_{0.5}\,(D - N) \neq 0$

Utilizando o teste de *Wilcoxon*.
$n = 8$

$X_i = D_i - N_i$	$Y_i = \lvert X_i \rvert$	$Y_{i:n}$	R_i
–0.6	0.6	0.5	1
1.6	1.6	(–) 0.6	2
–1.9	1.9	0.9	3
0.5	0.5	(–) 1.3	4
0.9	0.9	1.6	5
–2.1	2.1	(–) 1.9	6
–3.1	3.1	(–) 2.1	7
–1.3	1.3	(–) 3.1	8

E.T.: T_8^+ é igual à soma das ordens correspondentes aos sinais positivos
V.O.: $t_8^+ = 1+3+5 = 8$
R.R.: Rejeita-se H_0 para valores "grandes" ou "pequenos" de T_8^+, ou seja, rejeita-se H_0 ao nível α se $t_8^+ \leq c_1$ ou $t_8^+ \geq c_2$, sendo c_1 e c_2 tais que $P(T_8^+ \leq c_1) \leq \alpha/2$ e $P(T_8^+ \geq c_2) \leq \alpha/2$; equivalentemente, $P(T_8^+ < c_1+1) \leq \alpha/2$ e $P(T_8^+ > c_2-1) \leq \alpha/2$
Consultando a tabela dos quantis da estatística de *Wilcoxon* vem:

$$c_1 + 1 = \begin{cases} 1 & \alpha = 0.01 \\ 4 & \alpha = 0.05 \\ 6 & \alpha = 0.1 \end{cases} \Rightarrow c_1 = \begin{cases} 0 & \alpha = 0.01 \\ 3 & \alpha = 0.05 \\ 5 & \alpha = 0.1 \end{cases}$$

$$c_2 - 1 = \begin{cases} 36-1 & \alpha = 0.01 \\ 36-4 & \alpha = 0.05 \\ 36-6 & \alpha = 0.1 \end{cases} \Rightarrow c_2 = \begin{cases} 36 & \alpha = 0.01 \\ 33 & \alpha = 0.05 \\ 31 & \alpha = 0.1 \end{cases}$$

Então, rejeita-se H_0 se:
$T_8^+ \in \{0, 36\}$ para $\alpha = 1\%$
$T_8^+ \in \{0, 1, 2, 33, 34, 35, 36\}$ para $\alpha = 5\%$
$T_8^+ \in \{0, \ldots, 5, 31, \ldots, 36\}$ para $\alpha = 10\%$
Como $t_8^+ = 8$, que não pertence a qualquer das regiões de rejeição, não se rejeita H_0 aos níveis usuais

Conclusão: Não há evidência de que os níveis de colesterol de dia e de noite sejam diferentes.

Exemplo 16.

Os dados seguintes apresentam a quantidade de espigas de determinado cereal, produzidas por dois tipos de solo:

Solo tipo arenoso:	18	20	16	21	22	18	19	17	21
Solo tipo argiloso:	21	23	14	24	21	20	19	21	23

Pretende-se testar se os dois tipos de solos são equivalentes para a produção de cereal.

R:

X – v.a. que rep. o nº de espigas do cereal no solo argiloso
Y – v.a. que rep. o nº de espigas do cereal no solo arenoso
$m = n = 9$

$H_0 : X = Y \quad vs \quad H_1 : X \neq Y$ (em rigor, $H_0 : F_X(a) = G_Y(a), \forall a \in \mathbb{R} \quad vs \quad H_1 : \exists a \in \mathbb{R}: F_X(a) \neq G_Y(a)$)

Temos 2 amostras independentes, pelo que vamos utilizar o teste de *Mann-Whitney-Wilcoxon*.

E.T.: $W_{9,9}$ é igual à soma do nº de Y_j's maiores do que cada X_i

V.O.: $w_{9,9} = 1 + 0 + 9 + 0 + 1 + 3 + 4 + 1 + 0 = 19$

R.R.: Rejeita-se H_0 para valores "grandes" ou "pequenos" de $W_{9,9}$: rejeita-se H_0 ao nível α se $w_{9,9} \leq c_1$ ou $w_{9,9} \geq c_2$, sendo c_1 e c_2 tais que $P(W_{9,9} \leq c_1) \leq \alpha/2$ e $P(W_{9,9} \geq c_2) \leq \alpha/2$; equivalentemente, $P(W_{9,9} < c_1+1) \leq \alpha/2$ e $P(W_{9,9} > c_2-1) \leq \alpha/2$

Consultando a tabela dos quantis da estatística de *Mann-Whitney-Wilcoxon* vem:

$$c_1 + 1 = \begin{cases} 81-63 & \alpha = 0.05 \\ 81-59 & \alpha = 0.10 \end{cases} \Rightarrow c_1 = \begin{cases} 17 & \alpha = 0.05 \\ 21 & \alpha = 0.10 \end{cases}$$

$$c_2 - 1 = \begin{cases} 63 & \alpha = 0.05 \\ 59 & \alpha = 0.10 \end{cases} \Rightarrow c_2 = \begin{cases} 64 & \alpha = 0.05 \\ 60 & \alpha = 0.10 \end{cases}$$

Então, rejeita-se H_0 se:

$W_{9,9} \in \{0, ..., 17, 64, \ldots, 81\}$ para $\alpha = 5\%$

$W_{9,9} \in \{0, ..., 21, 60, \ldots, 81\}$ para $\alpha = 10\%$

Como $w_{9,9} = 19$, não se rejeita H_0 a 5% mas rejeita-se a 10%.

Conclusão: A 5% não há evidência de que os solos não sejam equivalentes, mas a 10% essa evidência já existe.

Exemplo 17.

Vinte indivíduos com cancro do mesmo tipo e na mesma fase de desenvolvimento, receberam 1 de 4 tipos de tratamento. Os indivíduos foram escolhidos aleatoriamente para formar 4 grupos de 5 doentes cada um, tendo-se administrado o tratamento a cada um destes grupos. O tempo de sobrevivência, em anos, foi o seguinte:

Tipo de tratamento	Tempo de sobrevivência (anos)				
A	14.2	10.6	9.4	5.6	2.4
B	12.8	12.3	6.4	6.1	1.6
C	11.5	10.1	5.1	5.0	4.8
D	14.9	13.7	8.5	7.7	5.9

Verifique se há evidência para afirmar que algum dos tratamentos é melhor.

R:

X – v.a. que rep. o nº de espigas do cereal no solo argiloso

Y – v.a. que rep. o nº de espigas do cereal no solo arenoso

$m = n = 9$

H_0 : Os tempos de sobrevivência são iguais para os 4 tratamentos

vs

H_1 : Existe pelo menos um tratamento com tempo de sobrevivência maior do que os restantes

(Formalmente:

em rigor, $H_0 : F_A(x) = F_B(x) = F_C(x) = F_D(x), \forall x \in \mathbb{R}$ *vs* $H_1 : \exists x \in \mathbb{R}: F_A(x) \neq F_B(x)$ ou $F_A(x) \neq F_C(x)$ ou …)

Temos 4 amostras independentes, pelo que vamos utilizar o teste de *Kruskall-Wallis*.

E.T.: $K_{5,5,5,5} = \frac{1}{S^2}\left\{\sum_{i=1}^{4}\frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right\}$, onde $S^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{4}\sum_{j=1}^{n_i}\tilde{R}^2(X_{ij}) - \frac{N(N+1)^2}{4}\right\}$

H_0 verd. => $K_{5,5,5,5} \sim \chi^2_{(3)}$

Rejeita-se H0 para valores "grandes" da E.T.

V.O.:

A_j<-X_{1j}	B_j<-X_{2j}	C_j<-X_{3j}	D_j<-X_{4j}	$\tilde{R}(X_{1j})$	$\tilde{R}(X_{2j})$	$\tilde{R}(X_{3j})$	$\tilde{R}(X_{4j})$
	1.6				1		
2.4				2			
		4.8				3	
		5.0				4	
		5.1				5	
5.6				6			
			5.9				7
	6.1				8		
	6.4				9		
			7.7				10
			8.5				11
9.4				12			
		10.1				13	
10.6				14			
		11.5				15	
	12.3				16		
	12.8				17		
			13.7				18
14.2				19			
			14.9				20
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n_4 = 5$	$R_1 = 53$	$R_2 = 51$	$R_3 = 40$	$R_4 = 66$
			$\sum \tilde{R}^2(X_{ij})$:	741	691	444	994

$N = 20$

$$s^2 = \frac{1}{19}\left(741 + 691 + 444 + 994 - \frac{20 \times 21^2}{4}\right) = 35$$

$$k_{5,5,5,5} = \frac{1}{35}\left\{\frac{53^2 + 51^2 + 40^2 + 66^2}{5} - \frac{20 \times 21^2}{4}\right\} \approx 1.95$$

Valor-p: p-*value* = $P(\chi^2_{(3)} \geq 1.95)$

$0.4 < P(\chi^2_{(3)} \leq 1.95) < 0.5 \Rightarrow 0.5 <$ p-*value* < 0.6

p-*value* > α, para todo o α usual => Não se rejeita H_0 para todo o α usual

Conclusão: Não existe evidência para afirmar que algum dos tratamentos seja melhor do que os restantes.

Exemplo 18.

A 150 alunos escolhidos aleatoriamente foi dado um teste de Matemática. A classificação podia ser uma de três: A, B ou C. Depois do teste realizado e corrigido, verificou-se que 40 crianças obtiveram a classificação A, 80 a classificação B e as restantes, C. Das que obtiveram A, 35% eram canhotas; das que obtiveram B, 25% eram canhotas e nas que obtiveram C a percentagem de canhotas era 10%. Tendo em consideração os dados anteriores, poderá concluir entre a existência de associação entre a nota de Matemática e o facto de a criança ser, ou não, canhota?

Exemplo 19.

Pediu-se a 100 homens e a 100 mulheres que usassem uma nova pasta dentífrica dizendo se gostavam, ou não, do seu sabor. 32 homens e 26 mulheres afirmaram que gostavam. Acha que este facto indica uma diferença de preferência entre homens e mulheres?

Exemplo 20.
Os dados seguintes referem-se a classificações dadas por dois professores a um grupo de estudantes que exprimiram os seus conhecimentos sobre 30 conceitos de geomorfologia. Pretende-se investigar se as classificações dos dois professores são idênticas isto é, estão relacionadas, no sentido em que quanto maior é uma, tanto maior é a outra:

Professor I	Professor II	Professor I	Professor II
79	62	64	45
71	66	52	39
108	96	72	76
35	33	86	97
114	116	41	41
69	55	56	117
63	39	92	84
121	133	62	73
61	44	99	104
120	103	107	129
118	137	78	78
102	122	24	12
75	125	109	111
37	31	67	57
117	119	90	93

Exemplo 21.
A televisão chegou à Austrália em 1957. Nos 5 anos seguintes os números de novas licenças de rádio (em milhares) e de novas licenças de televisão (em milhares) foram os seguintes:

Ano	Novas licenças de rádio	Novas licenças de TV
1957	171	74
1958	178	224
1959	251	300
1960	160	404
1961	155	323

Indicarão os resultados anteriores, que com a chegada da televisão o número de licenças de rádio diminuiu significativamente?