

Teste de aderência

Professor
Julio Cezar



AULA DE HOJE

- Teste de aderência.

TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

- Teste de aderência que verifica se a distribuição de um conjunto de valores observados (amostra) está de acordo com uma distribuição teórica.
 - Pode ser utilizado para qualquer distribuição teórica (principalmente para pequenas amostras)
 - Uniforme, Exponencial, Poisson, Gama, **Normal**...
- **Hipótese testada:**
 - H_0 : dados amostrais = distribuição teórica
 - H_1 : dados amostrais \neq distribuição teórica

TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

Criado pelos russos:

Kolmogorov e Smirnov em 1933

Andrei Nikolaevich Kolmogorov: foi um matemático soviético, que fez contribuições significativas em teoria da probabilidade, topologia, lógica intuicionista, turbulência, mecânica clássica, teoria algorítmica da informação e análise de algoritmos.

Nikolai Vasilyevich Smirnov: foi um matemático russo soviético conhecido por seu trabalho em vários campos, incluindo teoria da probabilidade e estatística.

· https://pt.wikipedia.org/wiki/Teste_Kolmogorov-Smirnov; [https://en.wikipedia.org/wiki/Nikolai_Smirnov_\(mathematician\)](https://en.wikipedia.org/wiki/Nikolai_Smirnov_(mathematician));
https://www.wikiwand.com/pt/Andrei_Kolmogorov

1903-1987



1900-1966



Teste Kolmogorov-Smirnov

- **Estatística do Teste:**
 - **Compara a distribuição de frequências acumuladas, sob H_0 (dados teóricos), com a distribuição acumulada dos dados amostrais (valores empíricos). Então, calcula-se uma distância máxima das frequências acumuladas, ou seja, distância entre os dados teóricos e os dados empíricos:**

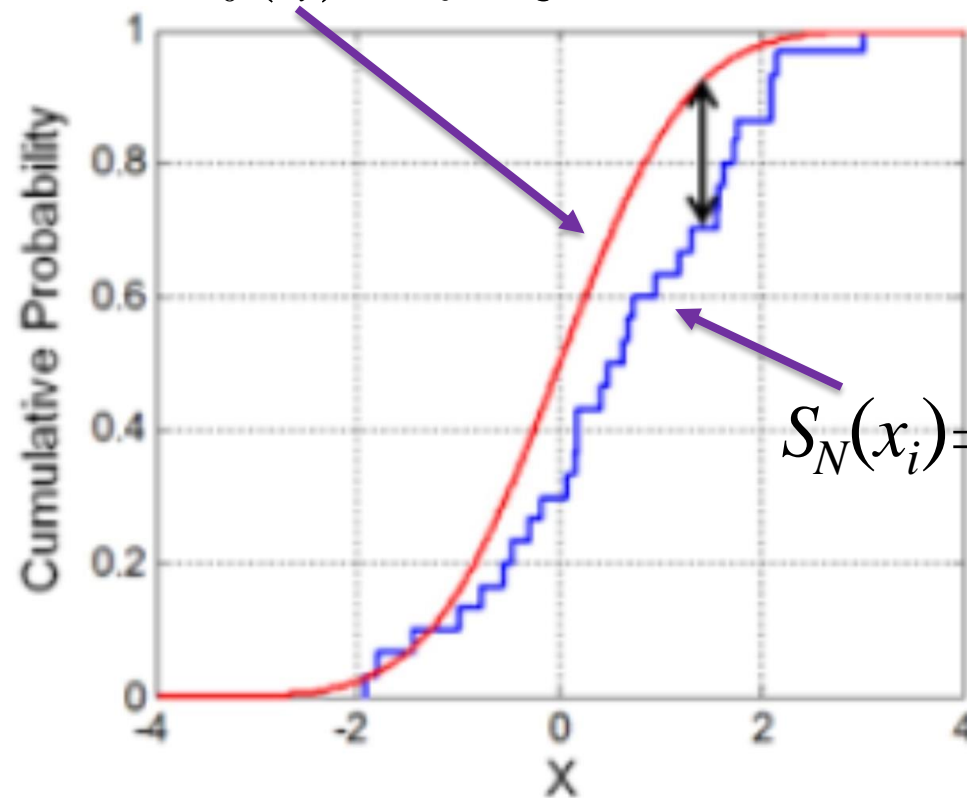
$$D = \max |F_0(x_i) - S_N(x_i)|$$

onde: $\begin{cases} F_0(x_i) = & \text{função acumulada dist. teórica} \\ S_N(x_i) = & \text{função acumulada amostra} \end{cases}$

TESTE DE ADERÊNCIA

Teste de Kolmogorov-Smirnov

$F_0(x_i) =$ função acumulada dist. teórica



$S_N(x_i) =$ função acumulada amostra

https://pt.wikipedia.org/wiki/Teste_Kolmogorov-Smirnov

TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

- **Tabela auxiliar para realização do teste:**

$$D = \max |F_0(x_i) - S_N(x_i)|$$

- Sugestão para Variáveis Quantitativas e aderência à **distribuição Normal**:

Variável Quantitativa Contínua – Dist. Normal						
dados	fi	fri	SN(x)=Fri	z	Fo(x)	Fo(x) - SN(x)
x1						
x2						
x3						
x4						
...						

Obs: como se trata da dist. Normal, z é obtida a partir da transformação →
e $F_0(x_i)$, a partir da Tabela Z (**usar a acumulada** = área à esquerda da curva)

$$\frac{x_i - \mu}{\sigma}$$

Teste de Kolmogorov-Smirnov

- **Resumo**

- 1) Elabore as Hipóteses Nula e Alternativa
- 2) Determine as frequências acumuladas da amostra e da distribuição teórica
- 3) Calcule a Estatística **D**
- 4) Procure o valor crítico na **tabela KS1**, de acordo com o 'n' e o nível de significância α
- 5) Tome a decisão e elabore a conclusão
 - Se $D \geq \text{Valor crítico} \rightarrow \text{Rejeita-se } H_0$
 - Se $D < \text{Valor crítico} \rightarrow \text{Não rejeita-se } H_0$

Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**

- O diâmetro de um parafuso produzido por determinada indústria, segue uma distribuição normal com média = 302 mm e $\sigma = 2$ mm.
- Foi coletada uma amostra de 10 parafusos e os resultados estão a seguir:

301	297	304	299	295	298	300	302	294	303
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- Considerando um nível de significância $\alpha = 5\%$ verifique se as peças seguem a especificação da indústria.

Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294						
295						
297						
298						
299						
300						
301						
302						
303						
304						

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)

Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294	1					
295	1					
297	1					
298	1					
299	1					
300	1					
301	1					
302	1					
303	1					
304	1					

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)

Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294	1	0,1				
295	1	0,1				
297	1	0,1				
298	1	0,1				
299	1	0,1				
300	1	0,1				
301	1	0,1				
302	1	0,1				
303	1	0,1				
304	1	0,1				

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)



Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294	1	0,1	0,1			
295	1	0,1	0,2			
297	1	0,1	0,3			
298	1	0,1	0,4			
299	1	0,1	0,5			
300	1	0,1	0,6			
301	1	0,1	0,7			
302	1	0,1	0,8			
303	1	0,1	0,9			
304	1	0,1	1,0			

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)



Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294	1	0,1	0,1	-4		
295	1	0,1	0,2	-3,5		
297	1	0,1	0,3	-2,5		
298	1	0,1	0,4	-2		
299	1	0,1	0,5	-1,5		
300	1	0,1	0,6	-1		
301	1	0,1	0,7	-0,5		
302	1	0,1	0,8	0		
303	1	0,1	0,9	0,5		
304	1	0,1	1,0	1		

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)



TESTE DE ADERÊNCIA

Kolmogorov-Smirnov

t. Normal):

tabela auxiliar:

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)

$P(Z < z)$

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

xi	Sn(x)	z	Fo(x)	Fo(x) - Sn(x)
0,1	0,1	-4	0,0000	
0,2	0,2	-3,5	0,0002	
0,3	0,3	-2,5	0,0062	
0,4	0,4	-2	0,0228	
0,5	0,5	-1,5	0,0668	
0,6	0,6	-1	0,1587	
0,7	0,7	-0,5	0,3085	
0,8	0,8	0	0,5000	
0,9	0,9	0,5	0,6915	
1,0	1,0	1	0,8413	

Teste Kolmogorov-Smirnov

- **Exemplo 1 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

padronização

$$\frac{x_i - 302}{2}$$

Área à esquerda (Tabela da Normal Z)

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
294	1	0,1	0,1	-4	0,0000	0,0999
295	1	0,1	0,2	-3,5	0,0002	0,1998
297	1	0,1	0,3	-2,5	0,0062	0,2938
298	1	0,1	0,4	-2	0,0228	0,3773
299	1	0,1	0,5	-1,5	0,0668	0,4332
300	1	0,1	0,6	-1	0,1587	0,4413
301	1	0,1	0,7	-0,5	0,3085	0,3915
302	1	0,1	0,8	0	0,5000	0,3000
303	1	0,1	0,9	0,5	0,6915	0,2085
304	1	0,1	1,0	1	0,8413	0,1587

Teste Kolmogorov-Smirnov

- **Exemplo 1**

$$D = \max |F_0(x_i) - S_N(x_i)| = \mathbf{0,44134}$$

- → Tabela KS1: Valor crítico = 0,410

- **Decisão:**

- Como $D > \text{Valor crítico}$ → Rejeita-se H_0 , ou seja, os dados observados não possuem distribuição normal com média = 302mm e dp = 2mm.

TESTE DE ADERÊNCIA

RODANDO O BANCO



- *Teste de Kolmogorov 1 amostra*
 - *Instalar pacote 'dgof' e depois carregar*
 - *Teste de aderência à curva normal com média igual a 302 e desvio padrão igual a 2.*

```
x=c(301,297,304,299,295,298,300,302,294,303)  
ks.test(x,"pnorm",302,2)
```

One-sample Kolmogorov-Smirnov test

```
data: x  
D = 0.44134, p-value = 0.02722  
alternative hypothesis: two-sided
```

Esta variável NÃO possui distribuição normal

H_0 : dados são
normalmente distribuídos

Teste Kolmogorov-Smirnov

- **Exemplo 2 (Dist. Normal):**

- A seguir estão os pesos de 16 estudantes submetidos à um tratamento de ganho de massa muscular:

57	58	58	59	59	59	62	62
63	63	63	63	65	68	68	70

- Teste a hipótese, considerando um $\alpha = 5\%$ que os dados são normalmente distribuídos com média igual a 62 kg e $dp=6$ kg.

TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

- **Exemplo 2 (Dist. Normal):**
 - Construção da tabela auxiliar:

padronização

$$\frac{x_i - 62}{6}$$

Área à esquerda

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)	F _o (x) - S _N (x)
57	1	0,0625	0,0625	-0,8333	0,2023	0,1398
58	2	0,1250	0,1875	-0,6667	0,2525	0,0649
59	3	0,1875	0,3750	-0,5000	0,3085	0,0665
62	2	0,1250	0,5000	0,0000	0,5000	0
63	4	0,2500	0,7500	0,1667	0,5662	0,1838
65	1	0,0625	0,8125	0,5000	0,6915	0,1210
68	2	0,1250	0,9375	1,0000	0,8413	0,0962
70	1	0,0625	1,0000	1,3333	0,9088	0,0912

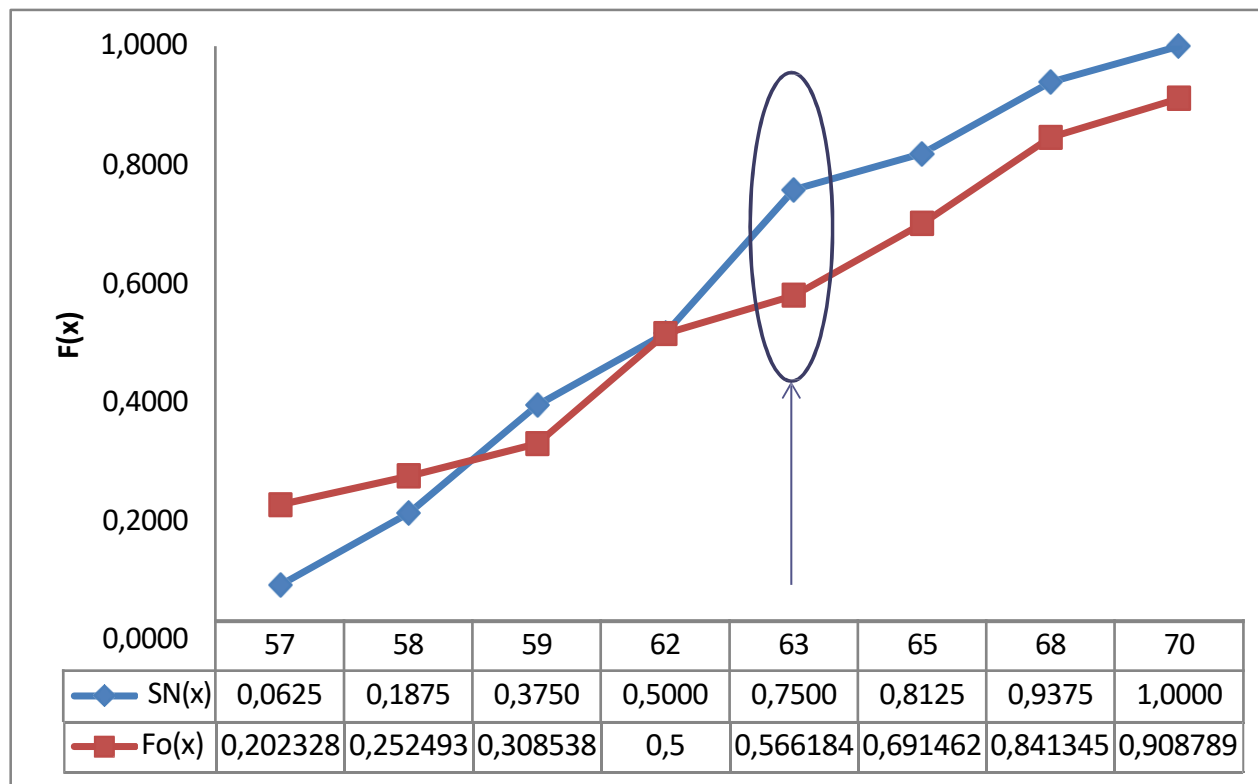
TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

- **Exemplo 2**

Lembrando que: S_N : função acumulada amostra e F_0 : função acumulada distribuição teórica.

▫ → Comparação entre as Funções Acumuladas



TESTE DE ADERÊNCIA

Teste Kolmogorov-Smirnov

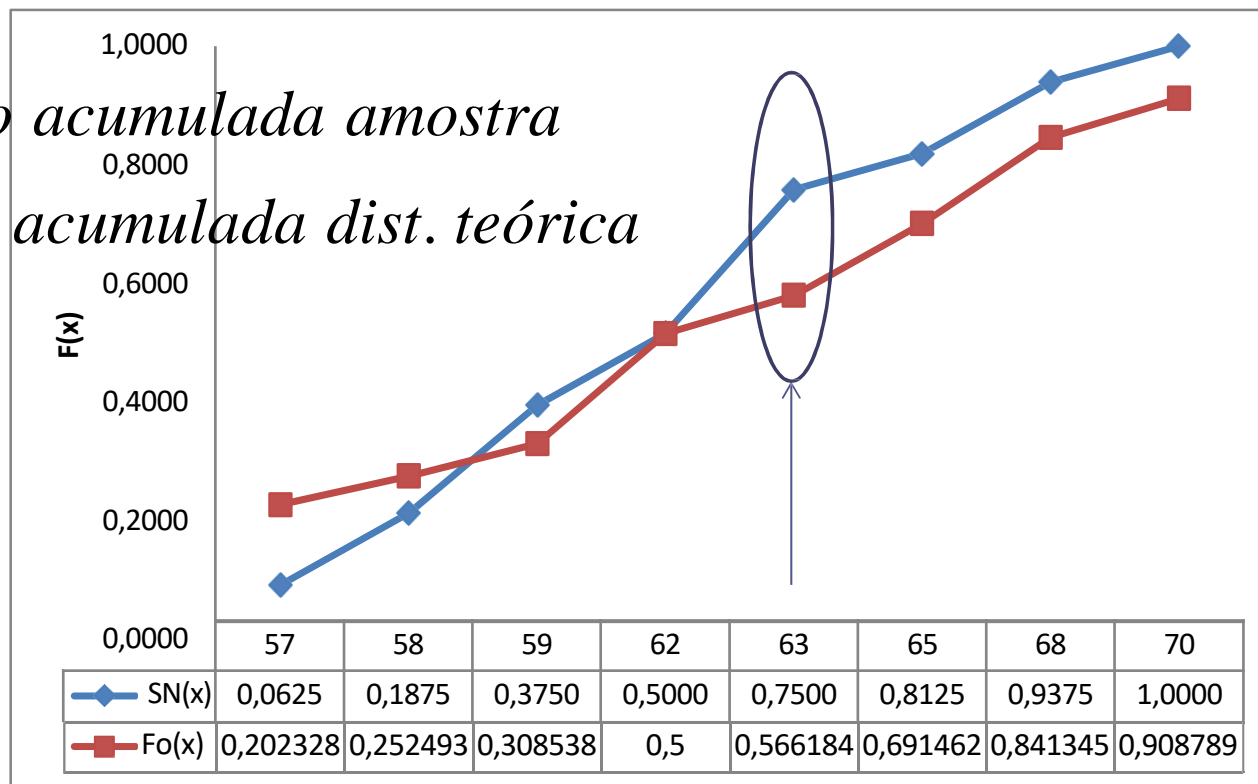
- Exemplo 2

Lembrando que: S_N : função acumulada amostra e F_0 : função acumulada distribuição teórica.

→ Comparação entre as Funções Acumuladas

$S_N(x_i) =$ função acumulada amostra

$F_0(x_i) =$ função acumulada dist. teórica



Teste Kolmogorov-Smirnov

- **Exemplo 2**

$$D = \max |F_0(x_i) - S_N(x_i)| = 0,1838$$

- → Tabela KS1: Valor crítico = 0,328

- **Decisão:**

- Como $D < \text{Valor crítico}$ → Não rejeita-se H_0 , ou seja, os dados observados são normalmente distribuídos.

TESTE DE ADERÊNCIA

RODANDO O BANCO



- *Teste de Kolmogorov 1 amostra*

- *Teste de aderência à curva normal com média igual a 62 e desvio padrão igual a 6.*

```
x=c(57,58,58,59,59,59,62,62,63,63,63,63,65,68,68,70)
ks.test(x,"pnorm",62,6)
```

```
data: x
D = 0,1838, p-value = 0.5291
alternative hypothesis: two-sided
```

Warning message:

```
In ks.test(x, "pnorm", 62, 6) :
ties should not be present for the Kolmogorov-Smirnov test
```

Esta variável possui distribuição normal

H_0 : dados são
normalmente distribuídos.

Teste Kolmogorov-Smirnov

- **Exemplo 3 (Dist. Normal):**

- A seguir estão os pesos de 7 adolescentes submetidos à um tratamento de emagrecimento:

67	68	68,2	69
69,9	71	73,4	

- Teste a hipótese, considerando um $\alpha = 1\%$ que os dados são normalmente distribuídos

Teste Kolmogorov-Smirnov

- **Exemplo 3 (Dist. Normal):**
 - **Construção da tabela auxiliar:**

Padronização:
Média=69,50
Dp=2,1626

dados	f _i	f _{ri}	S _N (x)	z	F _o (x)*	F _o (x) - S _N (x)
67,0	1	0,1429	0,1429	-1,1560	0,1238	0,0191
68,0	1	0,1429	0,2858	-0,6936	0,2439	0,0418
68,2	1	0,1429	0,4286	-0,6011	0,2739	0,1547
69,0	1	0,1429	0,5714	-0,2312	0,4086	0,1628
69,9	1	0,1429	0,7143	0,1849	0,5734	0,1409
71,0	1	0,1429	0,8571	0,6936	0,7560	0,1011
73,4	1	0,1429	1	1,8034	0,9643	0,0357

Teste Kolmogorov-Smirnov

- Exemplo 3

$$D = \max |F_0(x_i) - S_N(x_i)| = 0,1628$$

- ▫ → Tabela KS1: Valor crítico = 0,577

- Decisão:

- ▫ Como $D < \text{Valor crítico}$ → Aceita-se H_0 , ou seja, os dados observados são normalmente distribuídos.

TESTE DE ADERÊNCIA

RODANDO O BANCO



- *Testando a normalidade dos dados*
 - *Teste de KS (utilizando os dados da variável)*

```
x=c(67,68,68.2,69,69.9,71,73.4)
ks.test(x,"pnorm",mean(x),sd(x))
```

One-sample Kolmogorov-Smirnov test

```
data: x
D = 0.16285, p-value = 0.9766
alternative hypothesis: two-sided
```

Esta variável possui distribuição normal

H_0 : dados são
normalmente distribuídos



Referências

Arango HG. **Bioestatística Teórica e Computacional**. Guanabara Koogan. 2ª ed. Rio de Janeiro, 2005.

Devore, JL & Cordeiro, MTA. **Probabilidade e estatística: para engenharia e ciências**. Cengage Learning Edições Ltda., 2014.

Latorre MRDO. Estatística Não - Paramétrica (Apostila Pós-graduação) FSP/USP, 2009.

Mendonça J & Segri NJ. Estatística Não - Paramétrica (Apostila Pós-Graduação) UFMT, 2014.

Milone G. **Estatística Geral e Aplicada**. Thomson. São Paulo, 2006.

Siegel S & Castellan, Jr NJ. **Estatística Não Paramétrica para Ciências do Comportamento. Bookman**. 2ª edição. São Paulo, 2006.

Siqueira AL & Tibúrcio JD. **Estatística na Área da Saúde**. Coopmed. Belo Horizonte, 2011.

Vieira S. **Bioestatística - Tópicos Avançados**. ELSEVIER. 3ª ed. Rio de Janeiro, 2010.

Class finished

