ANÁLISE DE DADOS

REVISÃO II

Prof. Dr. Daniel Caetano

2020 - 1

Variáveis Aleatórias e Funções de Probabilidade

Variáveis Aleatórias

- Definidas como a função que atribui um valor para cada elemento do espaço amostral

- Espaço amostral qualitativo
  - Variáveis aleatórias discretas
- Espaço amostral numérico (quantitativo)
  - Variáveis aleatórias discretas ou contínuas

Função de Probabilidade

- Uma função que atribui uma probabilidade para cada valor da variável aleatória
- Se X é a variável aleatória
  - P(X) vai indicar a função de probabilidade de X
  - P(X=n) é a probabilidade de X igual ao valor **n**
- Ex.: ao lançar uma moeda, X é a variável que anota a se ocorre coroa. Qual a P(X)?

	X	P(X)
CA	0	0,5
CO	1	0,5
	TOTAL	1

Frequência Relativa

X

Frequência Absoluta

Esperança, Variância e Desvio Padrão

Média, Variância e Desvio Padrão

- Média ou esperança populacional: μ

$$\mu = \frac{x_1.f_1 + x_2.f_2 + \cdots + x_n.f_n}{f_1 + f_2 + \cdots + f_n}$$

$$\boldsymbol{\mu = x_1.p(x_1) + x_2.p(x_2) + \cdots + x_n.p(x_n)}$$

- Variância: σ^2

$$\sigma^2 = \frac{(x_1 - \mu)^2.f_1 + (x_2 - \mu)^2.f_2 + \cdots + (x_n - \mu)^2.f_n}{n}$$

$$\boldsymbol{\sigma^2 = (x_1 - \mu)^2.p(x_1) + (x_2 - \mu)^2.p(x_2) + \cdots + (x_n - \mu)^2.p(x_n)}$$

- Desvio Padrão: σ

$$\boldsymbol{\sigma = \sqrt{\sigma^2}}$$

Exemplo

- Lançamento de duas moedas, X = caras obtidas. Determine a média e o desvio padrão.

X	Freq.	P(X)	X.P(X)	$(X-\mu)^2.P(X)$
0	1	0,25	0	0,25
1	2	0,5	0,5	0
2	1	0,25	0,5	0,25
Total	4	1	**1**	**0,5**

Média: 1

Variância: 0,5

Desvio Padrão: 0,71

$$\sigma = \sqrt{\sigma^2}$$

Modelos Discretos de Probabilidade

Modelos Discretos de Probabilidade

- São modelos que descrevem a probabilidade de eventos representados por números inteiros
  - Distribuição de Bernoulli
  - Distribuição Binomial
  - Distribuição de Poisson

Distribuição de Bernoulli

- Há apenas dois resultados possíveis para o experimento: sucesso ou fracasso

- Equações Prontas para Bernoulli

$P(X = 1) = \mathbf{p}$

$P(X = 0) = 1\text{-}\mathbf{p}$

$\mu(X) = \mathbf{p}$

$\sigma^2(X) = \mathbf{p}.(1\text{-}\mathbf{p})$

$\sigma(X) = \sqrt{\sigma^2}$

Distribuição de Bernoulli

- Exemplo: ao entrar na UTI, a probabilidade de óbito de um paciente é de 20%. Seja X uma variável de bernoulli indicativa de óbito, calcule a distribuição de probabilidade, a média, a variância e o desvio padrão.

$P(X = 1) = \mathbf{p}$
$P(X = 0) = 1-\mathbf{p}$
$\mu(X) = \mathbf{p}$
$\sigma^2(X) = \mathbf{p}.(1-\mathbf{p})$
$\sigma(X) = \sqrt{\sigma^2}$

$P(X=1) = 0{,}2$
$P(X=0) = 0{,}8$
Média: 0,2
Variância: 0,16
Desvio: 0,4

Distribuição Binomial

- Repetições de experimentos independentes de Bernoulli, X indica o total de sucessos
- Equações Prontas para Binomial

$$p(X = x) = C_{n,x}.p^{x}.(1 - p)^{(n-x)}$$

$$\boxed{C_{n,x} = \frac{n!}{x!\,(n - x)!}}$$

$$\mu(X) = n.p$$

$$\sigma^2(X) = n.p.(1 - p)$$

$$\sigma(X) = \sqrt{\sigma^2(X)}$$

n: número de repetições independentes

p: probabilidade de sucesso em cada repetição

x: número de sucessos desejados

Distribuição Binomial

- Exemplo: O gerente de uma loja estima que, de 10 vendas, 3 são microcomputadores e 7 são outros produtos. Qual a probabilidade de 1 das próximas 4 vendas seja um computador? Qual a média e o desvio padrão?

- n: **4**
- p: **0,3**
- x: **1**

$$p(X = x) = C_{n,x} . p^{x} . (1 - p)^{(n-x)}$$

$$C_{n,x} = \frac{n!}{x!\,(n-x)!} \qquad C_{4,1} = \frac{4!}{1!\,(4-1)!} = 4$$

$$p(X = 1) = 4 . 0{,}3^{1} . (1 - 0{,}3)^{(4-1)} = 1{,}2 . 0{,}7^{3} = 0{,}4116$$

41,2%

Distribuição Binomial

- Exemplo: O gerente de uma loja estima que, de 10 vendas, 3 são microcomputadores e 7 são outros produtos. Qual a probabilidade de 1 das próximas 4 vendas seja um computador? Qual a média e o desvio padrão?
- n: **4**
- p: **0,3**
- x: **1**

$$p(X = 1) = \mathbf{0,4116}$$

$$\mu(X) = n.p \qquad \mu(X) = 4.0,3 = \mathbf{1,2}$$

$$\sigma^2(X) = n.p.(1-p) \qquad \sigma^2(X) = 4.0,3.(1-0,3) = 0,84$$

$$\sigma(X) = \mathbf{0,92}$$

Distribuição de Poisson

- Experimento mede ocorrências em intervalo de tempo/espaço/volume na variável X
- Equações Prontas para Poisson

$$p(X = k) = \frac{e^{-\lambda}.\lambda^{k}}{k!}$$

$$\mu(X) = \lambda$$

$$\sigma^2(X) = \lambda$$

$$\sigma(X) = \sqrt{\sigma^2(X)}$$

e ~ 2,71828182846

$$\lambda = n.p$$

λ: taxa média de ocorrências no intervalo

k: número de ocorrências desejado

Distribuição de Poisson

- Exemplo: probabilidade de ocorrer 3 defeitos em uma hora, com média de 4 defeitos por hora. Calcule a média e o desvio padrão
- k: **3**
- λ: **4**

$$p(X = k) = \frac{e^{-\lambda}.\lambda^{k}}{k!}$$

$$p(X = 3) = \frac{e^{-4}.4^{3}}{3!} = \frac{0{,}018.64}{6} = \mathbf{0{,}192}$$

$$\mu(X) = \lambda \qquad \mu(X) = \mathbf{4}$$

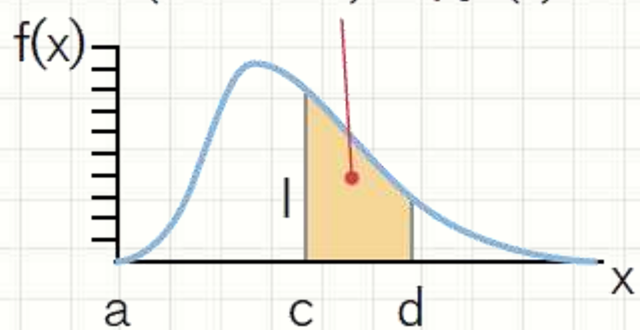
$$\sigma(X) = \sqrt{\lambda} \qquad \sigma(X) = \mathbf{2}$$

Modelos Contínuos de Probabilidade

Modelos Contínuos de Probabilidade

- São modelos que descrevem a probabilidade de eventos representados por números reais
  – Função Densidade de Probabilidade (FDP)

$$P(c \le x \le d) = \int_c^d f(x)dx$$



Tabelas!

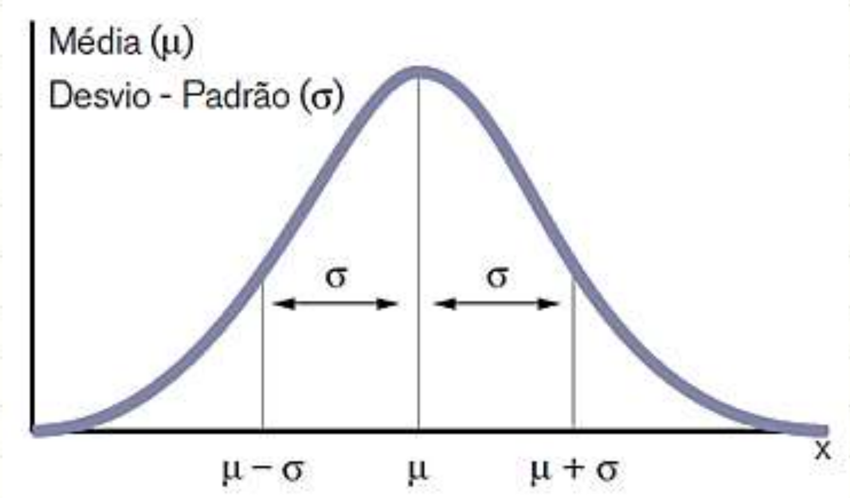
  – Probabilidade: área sob a curva entre os limites
  – A área total sob a curva deve valer 1!

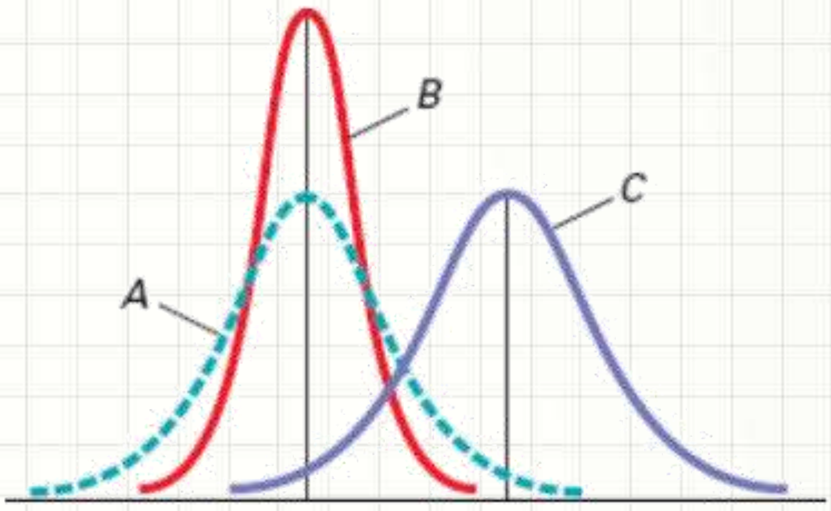
Distribuição Normal

- É a distribuição mais usada na estatística

$$f(x) = \frac{1}{\sqrt{2.\pi.\sigma^2}}.e^{-\frac{1}{2}.\left(\frac{x-\mu}{\sigma}\right)^2}$$

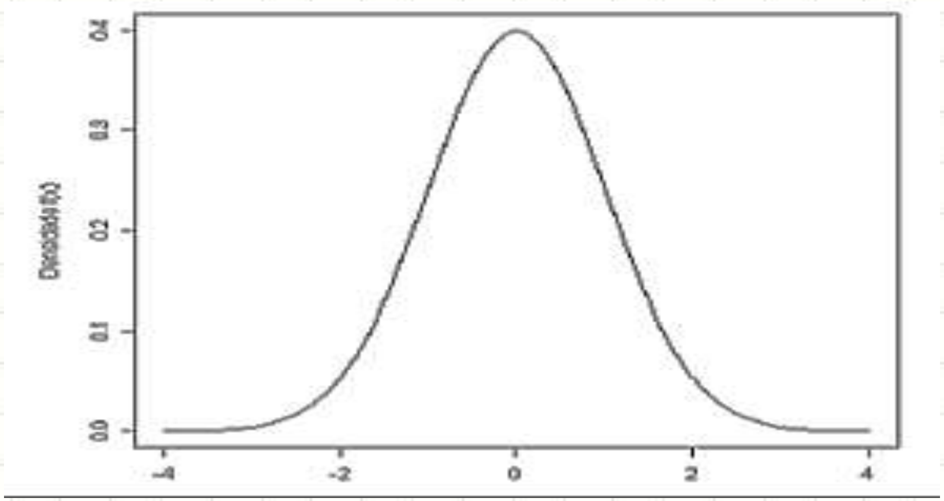
- Definida por μ e σ (ou $\boldsymbol{\overline{x}}$ e $\boldsymbol{s}$)





Distribuição Normal Padrão

- Distribuição normal de $\mu=0$ e $\sigma=1$...



  – “Padronizar” uma variável: $$Z = \frac{x - \mu}{\sigma}$$

Distribuição Normal

- Exemplo: Em uma fábrica, verificou-se que a vida útil das lâmpadas é de, em média, 2000 horas, com um desvio padrão de 200 horas. Determine a probabilidade de uma lâmpada durar mais que 2400 horas.

$$Z = \frac{x - \mu}{\sigma} \qquad Z = \frac{2400 - 2000}{200} = \ 2{,}00$$

$$\boldsymbol{P(X \geq 2400) =}$$

$$\boldsymbol{0,5 - 0,47725}$$

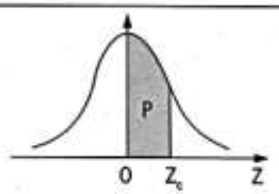
$$\boldsymbol{P(X \geq 2400) = 0,02275}$$

$$\boldsymbol{\approx 2,3\%}$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$

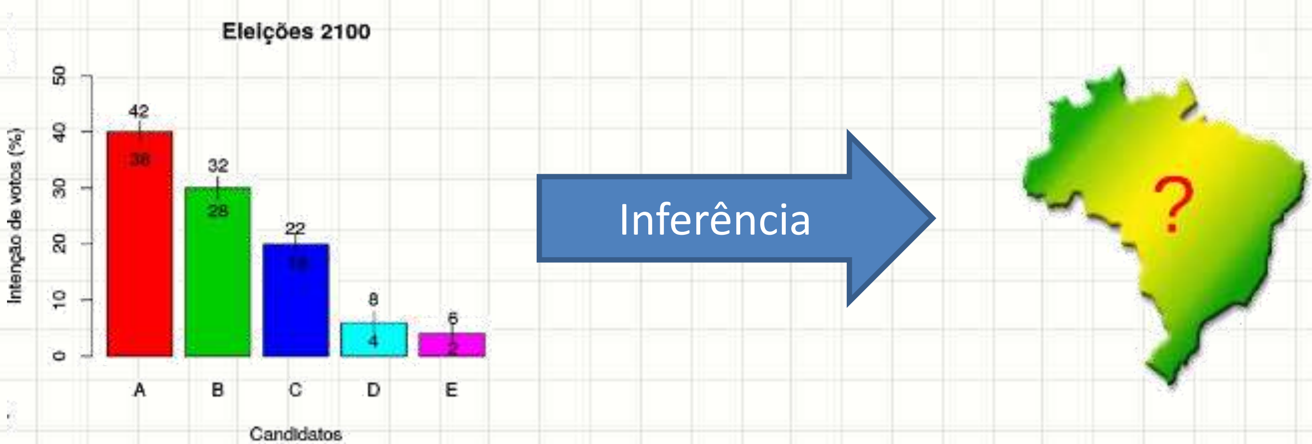


parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Distribuições de Amostragens da Média e da Proporção

Da Amostra à População

- Sabendo a média ou proporção da amostra
  - Podemos querer inferir o parâmetro (população)
  - Qual é o “erro-padrão” ao se fazer isso?
- Exemplos:
  - Média de altura do brasileiro...
  - Pesquisa eleitoral...



D.A. da Média Aritmética

- Média da amostra: estima a da população
- Erro-padrão: desvio para todas as amostras

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Probabilidade de amostra ter média $\bar{x}$:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

D.A. da Média Aritmética

- Exemplo: Considerando que a média de conteúdo de uma caixa é 368g com distribuição normal e desvio padrão 15g, qual a probabilidade de uma amostra de 25 caixas possuir menos que 365g? Qual o erro-padrão dessa amostra?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{365 - 368}{\frac{15}{\sqrt{25}}} = \mathbf{-1,00}$$

$$P(C < 365) = 0{,}5 - 0{,}34134$$

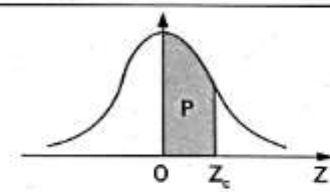
$$P(C < 365) = 0{,}15866$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3$$

Tabela III — Distrib

$Z \sim$ 1

Corpo da tabela dá a probabili



parte inteira e primeira decimal de Z_c	0	1	2	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0						
1,0	34134	34375	34614	35769	35993	36214	1,0
1,1	36433	36650	36864	37900	38100	38298	1,1
1,2	38493	38686	38877	39796	39973	40147	1,2
1,3	40320	40490	40658	41466	41621	41774	1,3
1,4	41924	42073	42220	42922	43056	43189	1,4
1,5	43319	43448	43574	44179	44295	44408	1,5
1,6	44520	44630	44738	45254	45352	45449	1,6
1,7	45543	45637	45728	46164	46246	46327	1,7
1,8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

D.A. da Proporção

- Proporção da amostra: estima a da população
- Erro-padrão: desvio para todas as amostras

$$\sigma_p = \sqrt{\frac{\Pi.(1-\Pi)}{n}}$$

- Probabilidade de amostra ter média $\bar{x}$:

$$Z = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}}$$

D.A. da Proporção

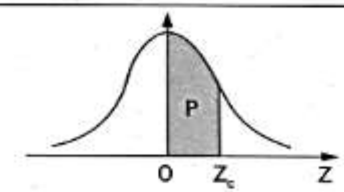
- Exemplo: Considerando que 77% dos adultos em férias desejam acesso à internet, qual a probabilidade de, em uma amostra de 200 adultos em férias, mais de 80% desejar internet.

$$Z = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}} = \frac{0{,}8 - 0{,}77}{\sqrt{\frac{0{,}77.0{,}23}{200}}} = \mathbf{1{,}01}$$

$$P(I > 80\%) = 0{,}5 - 0{,}34375$$

$$P(I > 80\%) = 0{,}15625$$

Tabela III — Distrib
$Z \sim$ 1
Corpo da tabela dá a probabil



parte inteira e primeira decimal de Z_c	0	1	2	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0						
1,0	34134	34375	34614	35769	35993	36214	1,0
1,1	36433	36650	36864	37900	38100	38298	1,1
1,2	38493	38686	38877	39796	39973	40147	1,2
1,3	40320	40490	40658	41466	41621	41774	1,3
1,4	41924	42073	42220	42922	43056	43189	1,4
1,5	43319	43448	43574	44179	44295	44408	1,5
1,6	44520	44630	44738	45254	45352	45449	1,6
1,7	45543	45637	45728	46164	46246	46327	1,7

Exercícios

Exercício 1

O espaço amostral de um experimento é $S = \{1, 2, 7, 10\}$. Sabendo que a probabilidade de ocorrência desses valores é diretamente proporcional aos mesmos, determine a função de probabilidade, a esperança e o desvio padrão.

Exercício 1

S = {1, 2, 7, 10}. Sabendo que a probabilidade de ocorrência desses valores é diretamente proporcional aos mesmos, determine a função de probabilidade, a esperança e o desvio padrão.

X	Frq.	P(X)	X.P(X)	$(X-\mu)^2.P(X)$
1	1	0,05	0,05	2,2445
2	2	0,1	0,2	3,2490
7	7	0,35	2,45	0,1715
10	10	0,5	5	2,645
Tot.	**20**	**1**	**7,7**	**8,31**

Média: 7,70
Variância: 8,31
Desvio Padrão: 2,88

Exercício 2

- A taxa de devoluções de pedidos de uma loja é de 13%. Considerando que 20 consumidores compraram, qual a probabilidade de que não mais que algum consumidor faça devoluções?

Exercício 2 **Binomial!**

- Taxa de devoluções: 13%. Considerando que 20 compras, probabilidade alguma devolução
- X é o número de devoluções: P (X > 0)

$$p(X = x) = C_{n,x} . p^{x} . (1 - p)^{(n-x)}$$

$$C_{n,x} = \frac{n!}{x!\,(n - x)!} \qquad C_{20,0} = \frac{20!}{0!\,(20 - 0)!} = 1$$

$$p(X = 0) = 1 . 0{,}13^{0} . (1 - 0{,}13)^{(20-0)}$$

$$p(X = 0) = 0{,}062$$

$$p(X > 0) = 1 - 0{,}062 = 0{,}938 \qquad \approx \mathbf{93{,}8\%}$$

Exercício 3

- As pizzas produzidas em uma fábrica recebem, em média, 8 azeitonas. Qual a probabilidade de uma pizza ter exatamente 3 azeitonas?

Exercício 3 **Poisson**

- As pizzas produzidas em uma fábrica recebem, em média, 8 azeitonas. Qual a probabilidade de uma pizza ter exatamente 3 azeitonas?

$$p(X = k) = \frac{e^{-\lambda}.\lambda^{k}}{k!}$$

$$p(X = 3) = \frac{e^{-8}.8^{3}}{3!} = \quad 0{,}0286\ldots$$

$$\approx \mathbf{2{,}9\%}$$

Exercício 4

- A duração de uma tela de LCD em uso contínuo é normalmente distribuída com média de 1700 dias e desvio padrão de 90 dias. Calcule a probabilidade do componente durar menos que 1500 dias

Exercício 4

- Duração média: 1700 dias e desvio: 90 dias. Probabilidade de durar menos que 1500 dias

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{1500 - 1700}{90} = -2{,}22$$

$$P(Z \leq -2{,}22) =$$

$$0{,}5 - 0{,}48679$$

$$P(Z \leq -2{,}22) = 0{,}01321$$

$$\approx \mathbf{1{,}3\%}$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$

parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c							
	0	1	2	3	4	5	6	7
	p = 0							
2,0	47725	47778	47831	47882	47932	47982	48030	4807
2,1	48214	48257	48300	48341	48382	48422	48461	4850
2,2	48610	48645	48679	48713	48745	48778	48809	4884
2,3	48928	48956	48983	49010	49036	49061	49086	4911
2,4	49180	49202	49224	49245	49266	49286	49305	4932
2,5	49379	49396	49413	49430	49446	49461	49477	4949
2,6	49534	49547	49560	49573	49585	49598	49609	4962
2,7	49653	49664	49674	49683	49693	49702	49711	4972
2,8	49744	49752	49760	49767	49774	49781	49788	4979
2,9	49813	49819	49825	49831	49836	49841	49846	4985
3,0	49865	49869	49874	49878	49882	49886	49889	4989
3,1	49903	49906	49910	49913	49916	49918	49921	4992
3,2	49931	49934	49936	49938	49940	49942	49944	4994
3,3	49952	49953	49955	49957	49958	49960	49961	4996
3,4	49966	49968	49969	49970	49971	49972	49973	4997
3,5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Exercício 5

Para uma amostra de 50 domicílios, foi feita a pergunta: “Você tem um dispositivo Apple?”. Dos 50 respondentes, 20 afirmaram que sim. Determine a proporção da amostra e o erro-padrão da proporção.

Exercício 5

Para uma amostra de 50 domicílios, foi feita a pergunta: “Você tem um dispositivo Apple?”. Dos 50 respondentes, 20 afirmaram que sim. Determine a proporção da amostra e o erro-padrão da proporção determinada.

$$p = \frac{X}{n} = \frac{20}{50} = 0{,}4$$

$$\sigma_p = \sqrt{\frac{0{,}4.\,(1 - 0{,}4)}{50}} = 0{,}11$$

Exercício 6

Um atendente de pedágio faz um atendimento, em média, $\mu=2{,}50$ minutos e $\sigma = 0{,}60$ minutos. Em uma amostra de 9 carros, qual a probabilidade de que a média por carro seja de, pelo menos, 3 minutos?

Exercício 6

Atendimento: μ=2,50   0,60

Amostra: 9 carros... P($\bar{T}$>3,00)?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3 - 2{,}5}{\frac{0{,}6}{\sqrt{9}}} = \frac{0{,}5}{0{,}2} = 2{,}50$$

$$P(\bar{T} > 3) = 0{,}5 - 0{,}49379$$

$$P(\bar{T} > 3) = 0{,}00621$$

$$\approx \mathbf{0{,}6\%}$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z$

parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6
	p = 0						
2,0	47725	47778	47831	47882	47932	47982	4803[illegible]
2,1	48214	48257	48300	48341	48382	48422	48461
2,2	48610	48645	48679	48713	48745	48778	4880[illegible]
2,3	48928	48956	48983	49010	49036	49061	4908[illegible]
2,4	49180	49202	49224	49245	49266	49286	4930[illegible]
2,5	49379	49396	49413	49430	49446	49461	4947[illegible]
2,6	49534	49547	49560	49573	49585	49598	4960[illegible]
2,7	49653	49664	49674	49683	49693	49702	4971[illegible]
2,8	49744	49752	49760	49767	49774	49781	4978[illegible]
2,9	49813	49819	49825	49831	49836	49841	[illegible]

Exercício 7

Um instituto fez uma pesquisa eleitoral com duas candidatas. Supondo que na amostra uma delas receber pelo menos 55% dos votos, essa candidata terá o prognóstico de vencedora. Se selecionar uma amostra de 100 eleitores, qual a probabilidade de ela ter prognóstico de vencedora se o porcentual de votos pra ela na população for 50,1%?

$\Pi =$ 50,1% $\quad$ n = 100 eleitores $\quad$ p = 55%

Exercício 7

4. Π = 50,1%; n = 100 eleitores; p = 55%

$$Z = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}} = \frac{0{,}55 - 0{,}501}{\sqrt{\frac{0{,}501.0{,}499}{100}}} = \mathbf{0,98}$$

$$P(x > 50{,}1\%) = 0{,}5 - 0{,}33646 \qquad P(x > 50{,}1\%) = 0{,}16354$$

decimal de Z_c	0	1	2	3	4	5	6	7	8	9	decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Perguntas?