Entendendo Estatística Divertidamente

Profa. Adriana Silva

Seja bem vindX!!!
Câmera ligada e
Microfone mutado sempre
que não estiver falando

entendendo
estatística
divertidamente

Distribuições de Probabilidade

Variáveis Contínuas

- Normal

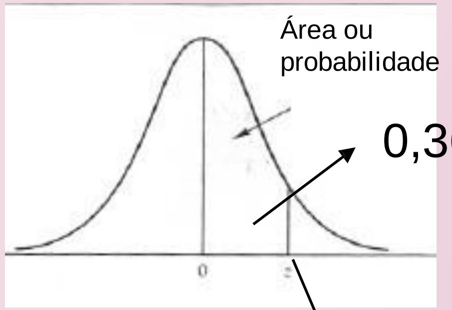
- Exponencial

- t de Student

entendendo
estatística
divertidamente

Distribuições de Probabilidade

Normal



0,3643

1,1

P(0 < Z < 1,1) = 0,3643

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2.3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2.4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2.5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2.6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2.7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2.8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2.9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3.0	0,4986	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

entendendo
estatística
divertidamente

Distribuições de Probabilidade

Normal – Exercício

- Supondo que Z segue uma distribuição normal padrão [$Z \sim N(0,1)$]. Desenhe a curva e calcule:

1) $P(0 < Z < 1,96)$
2) $P(-1,64 < Z < 0)$
3) $P(Z > 1,62)$
4) $P(Z < -0,84)$
5) $P(-0,68 < Z < 1,85)$

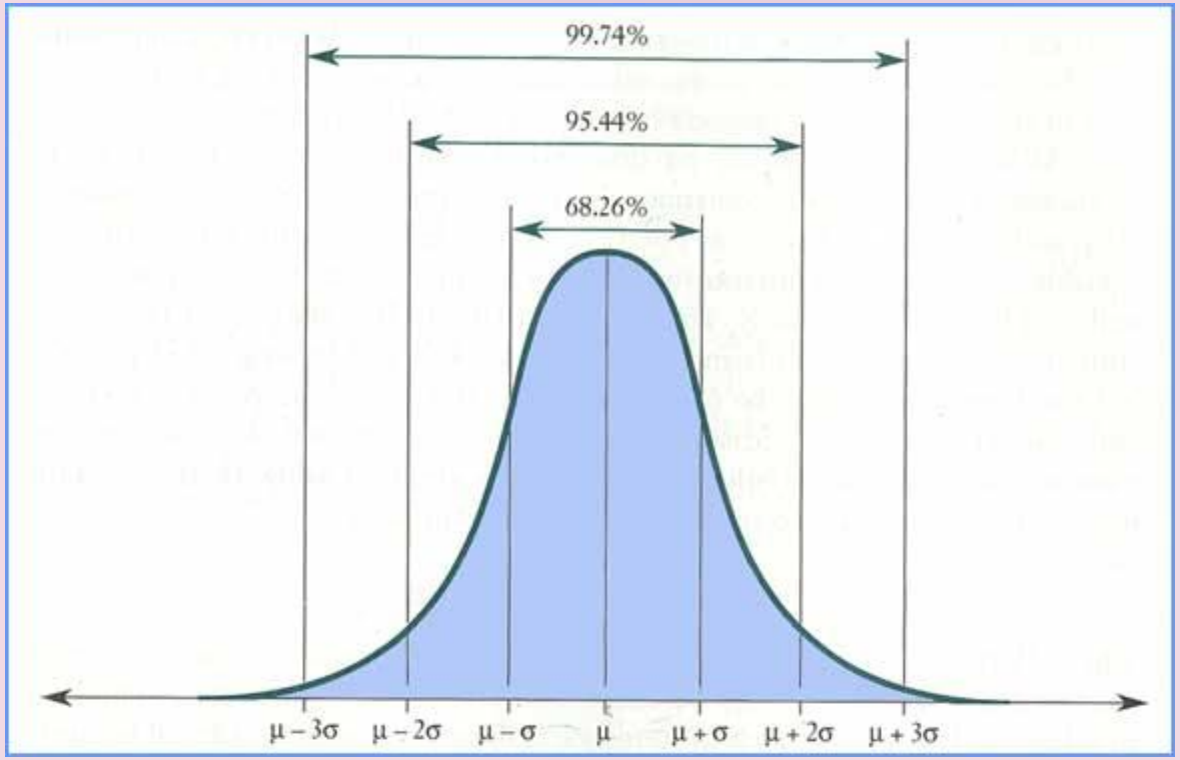
1) $P(-1 < Z < 1)$
2) $P(-2 < Z < 2)$
3) $P(-3 < Z < 3)$

entendendo estatística divertidamente

Distribuições de Probabilidade

Normal

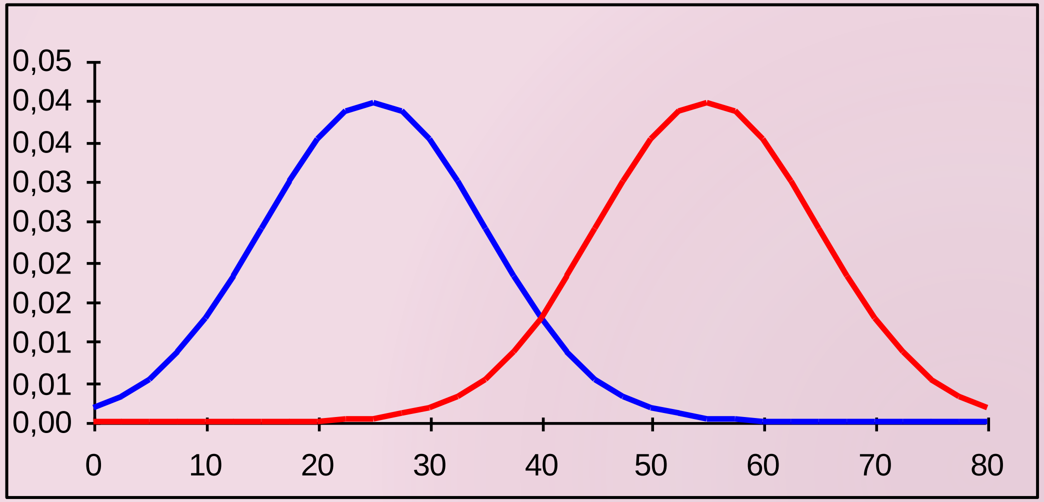
- As porcentagens dos valores de alguns intervalos em relação a uma variável aleatória Normal são comumente utilizados:



entendendo estatística divertidamente

Distribuições de Probabilidade

Normal

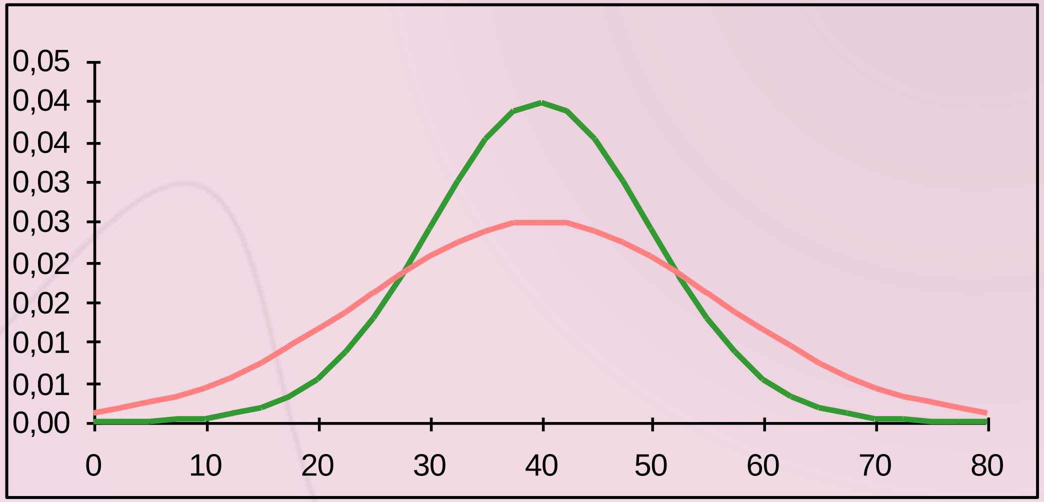


X~N(25,100)

X~N(55,100)

Média de X

Variância de X



X~N(40,250)

X~N(40,900)

entendendo estatística divertidamente

Distribuições de Probabilidade

Normal

- Para converter qualquer variável aleatória com distribuição Normal de média μ e variância σ^2 na distribuição Normal Padrão (média 0 e variância 1) é necessário fazer a seguinte transformação:

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

onde

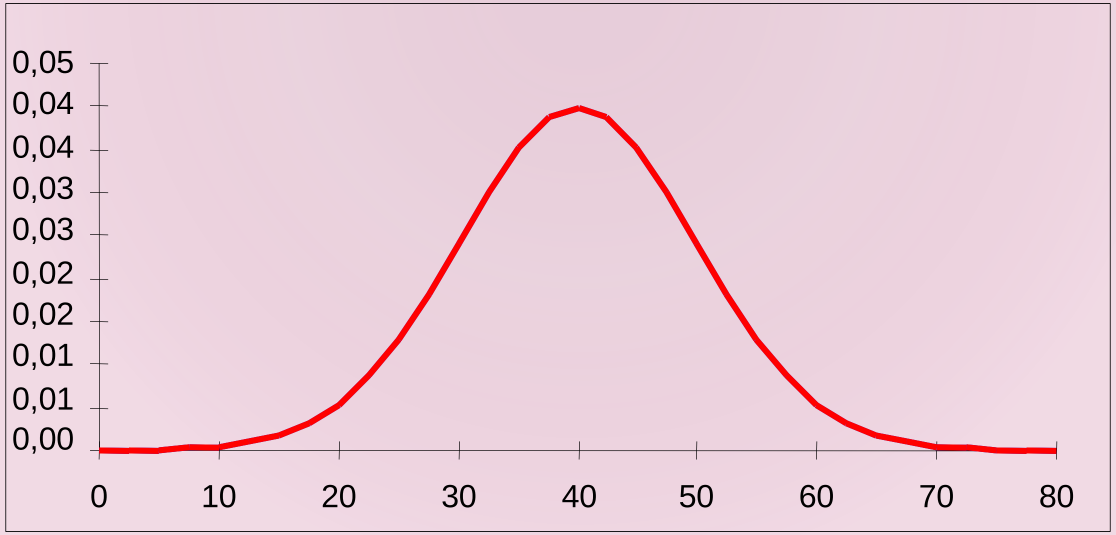
X é uma variável aleatória com distribuição Normal de média μ e variância σ^2 [X ~ N(μ, σ^2)]

Z é uma variável aleatória com distribuição Normal Padrão [Z ~ N(0,1)]

Distribuições de Probabilidade

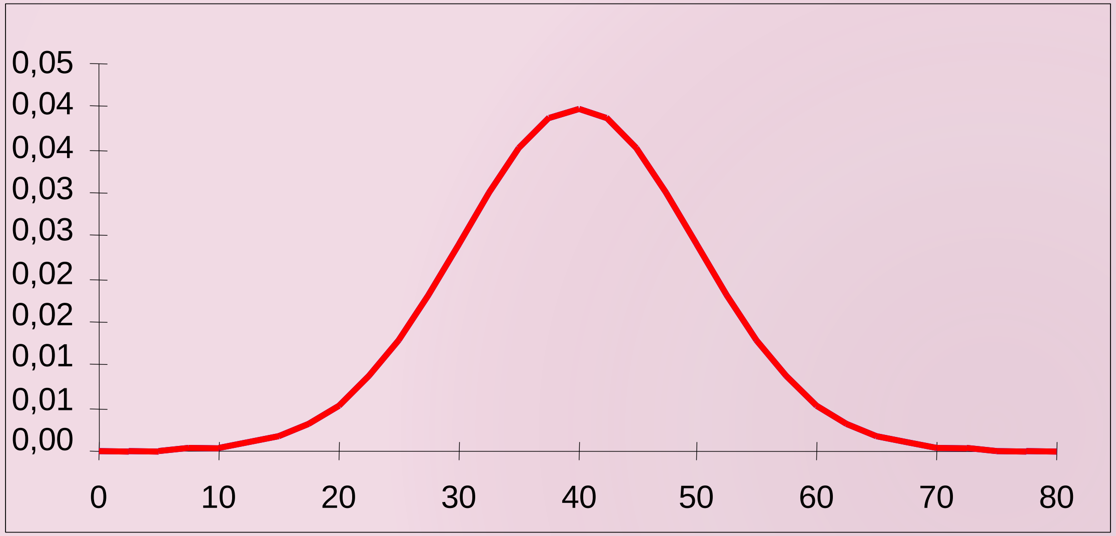
Normal – Exercício

- Uma empresa possui um faturamento mensal médio de 40 milhões de reais com um desvio padrão de 10 milhões de reais.
  - Qual a probabilidade de que em um determinado mês o faturamento esteja entre 30 a 50 milhões de reais?



Distribuições de Probabilidade

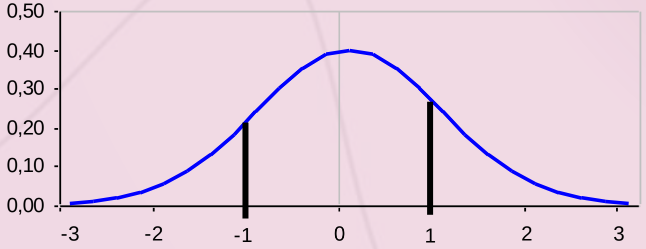
Normal – Exercício



$$X \sim N(40{,}100)$$

$$Z = \frac{X - \mu}{\sigma} \qquad Z \sim N(0{,}1)$$

$$P(30 < X < 50) = P\left(\frac{30 - 40}{10} < Z < \frac{50 - 40}{10}\right) = P(-1 < Z < 1)$$

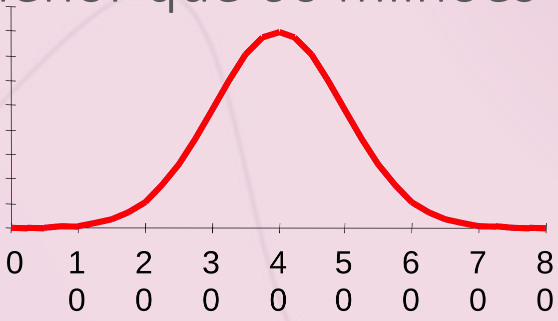


$P(-1 < Z < 1) = 2 * P\,(0 < Z < 1) = 2 * 0{,}34 = 0{,}68$

Distribuições de Probabilidade

Normal – Exercício

- Uma empresa possui um faturamento mensal médio de 40 milhões de reais com um desvio padrão de 10 milhões de reais.
  - Qual a probabilidade de que em um determinado mês o faturamento esteja entre 20 a 50 milhões de reais?
  - Qual a probabilidade de que em um determinado mês o faturamento seja maior que 30 milhões de reais?
  - Qual a probabilidade de que em um determinado mês o faturamento seja menor que 60 milhões de reais?



$$X \sim N(40{,}100) \qquad Z = \frac{X - \mu}{\sigma} \qquad Z \sim N(0{,}1)$$

Distribuições de Probabilidade

Variáveis Contínuas

Normal

A distribuição normal é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais. Isso se deve ao fato de que um grande número de fenômenos naturais apresenta sua distribuição de probabilidade tão proximamente normal. A distribuição normal também é chamada distribuição gaussiana.

Exponencial

A distribuição exponencial é um tipo de distribuição contínua de probabilidade, representada por um parâmetro.

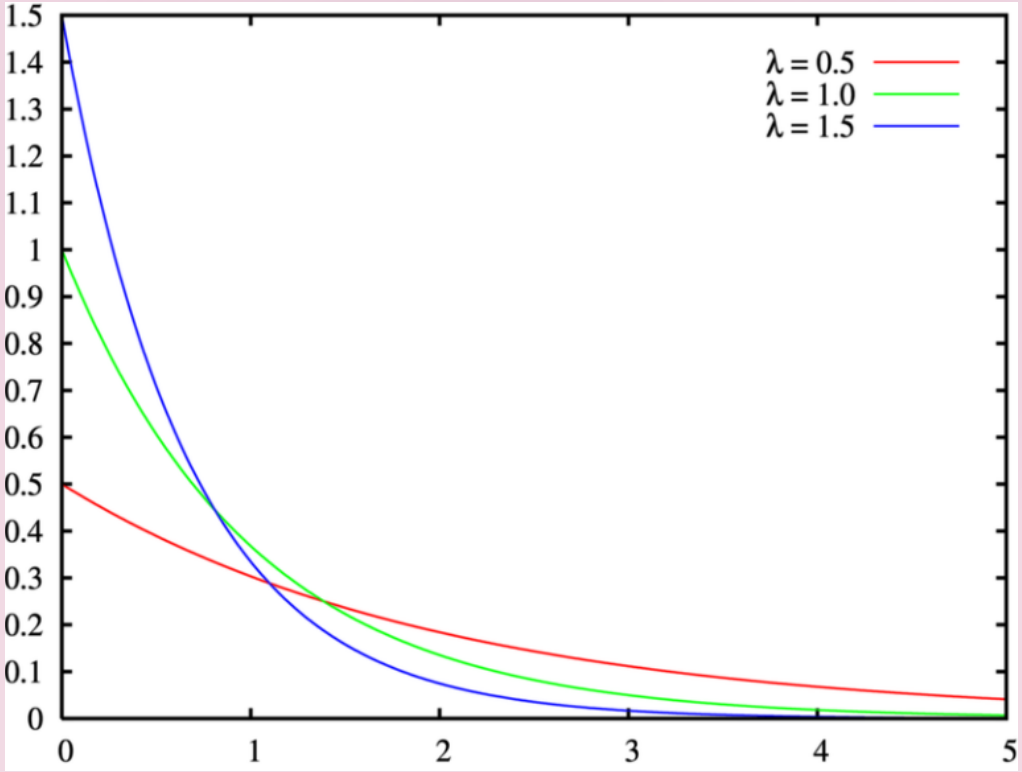
t de Student

entendendo estatística divertidamente

Distribuições de Probabilidade

Exponencial

- X: duração do atendimento prestado pela central de atendimento 24 horas de uma seguradora (em minutos).



Distribuições de Probabilidade

Exponencial

- X: duração do atendimento prestado pela central de atendimento 24 horas de uma seguradora (em minutos).
- λ: duração média do atendimento na central 24 horas (em minutos)
- $F(x)$: função de densidade de probabilidade acumulada

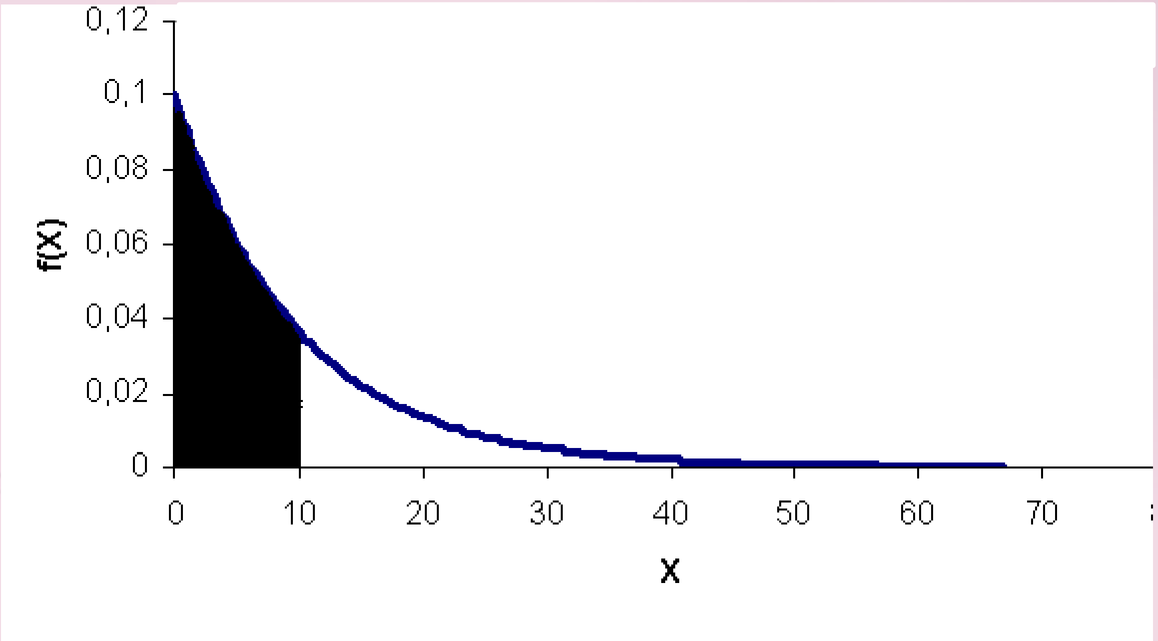
$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\frac{x}{\lambda}}, se\ x \geq 0 \\ 0 \qquad\qquad , se\ x < 0 \end{cases}$$

entendendo estatística divertidamente

Distribuições de Probabilidade

Exponencial – Exercício 11 em sala

- X: duração do atendimento prestado pela central de atendimento 24 horas de uma seguradora (em minutos).
- Qual probabilidade do atendimento ser inferior a 10 minutos? Sabendo que a média no tempo de atendimento é de 12 minutos.



$$P(X \leq x) = \begin{cases} 1 - e^{-\frac{x}{\lambda}}, se\ x \geq 0 \\ 0 \qquad\quad , se\ x < 0 \end{cases}$$

$$P(X \leq 10) = 1 - e^{-\frac{x}{\lambda}} = 1 - e^{-\frac{10}{12}} = 1 - e^{-0,833} = 0,5654$$

Distribuições de Probabilidade

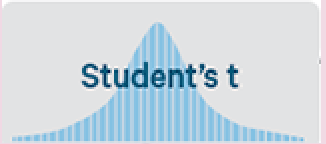
Variáveis Contínuas

Normal

A distribuição normal é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais. Isso se deve ao fato de que um grande número de fenômenos naturais apresenta sua distribuição de probabilidade tão proximamente normal. A distribuição normal também é chamada distribuição gaussiana.

Exponencial

A distribuição exponencial é um tipo de distribuição contínua de probabilidade, representada por um parâmetro.



t de Student

A distribuição t é uma distribuição de probabilidade teórica. É simétrica e semelhante à curva normal padrão, porém com caudas mais largas, ou seja, uma simulação da t de Student pode gerar valores mais extremos que uma simulação da normal.

entendendo estatística divertidamente

Distribuições de Probabilidade

t de Student

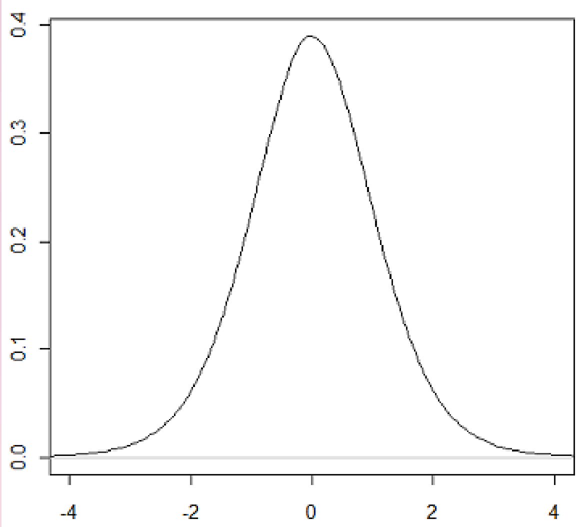
- Uma variável aleatória contínua X tem distribuição t de Student com ν graus de liberdade se sua função densidade de probabilidade é dada por

$$f(x) = \frac{\Gamma\left(\frac{\upsilon + 1}{2}\right)}{\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left(1 + \frac{x^2}{\upsilon}\right)^{-\left(\frac{\upsilon+1}{2}\right)}$$

Distribuições de Probabilidade

t de Student

- A função densidade da distribuição t de Student tem a mesma forma em sino da distribuição Normal, mas reflete a maior variabilidade (com curvas mais alargadas) que é de se esperar em amostras pequenas.
- Quanto maior o grau de liberdade, mais a distribuição t de Student se aproxima da distribuição Normal.
- Abaixo temos um gráfico da função densidade de um t de Student com 10 graus de liberdade.

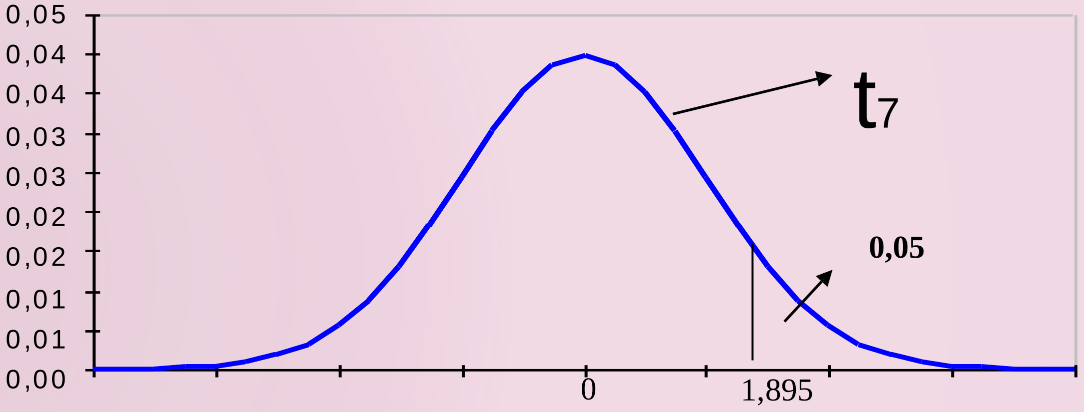


entendendo estatística divertidamente

Distribuições de Probabilidade

t de Student

Graus de Liberdade	Área da Extremidade Superior				
	0,10	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576



$P(X > 1{,}895) = 0{,}05$

- De onde vem o poder extraordinário de generalização?
- TLC = fonte do poder quando se usa uma amostra para inferir

Teorema do Limite Central

Festival da Salsicha

- Maratona na cidade
- Ônibus quebra => peso médio = 95kg

Teorema do Limite Central

Festival da Salsicha

TLC nos permite fazer as seguintes inferências:

1. Se tivermos informações detalhadas sobre alguma população, então podemos fazer inferências poderosas sobre qualquer amostra adequadamente extraída dessa população;
2. Se tivermos informações detalhadas sobre uma amostra extraída de modo adequado, podemos fazer inferências surpreendentementes acuradas sobre a população da qual a amostra foi retirada;
3. Se tivermos dados que descrevem uma amostra particular e dados sobre uma população particular, podemos inferir se a amostra é consistente ou não com uma amostra com probabilidade de ter sido tirada dessa população;
4. Por fim, se soubermos características subjacentes de duas amostras, podemos inferir se ambas foram provavelmente extraídas ou não de uma mesma população;

Teorema do Limite Central

Festival da Salsicha

- Maratona na cidade
- Ônibus quebra => peso médio = 95kg
- Outro ônibus quebra => peso médio = 72kg

Teorema do Limite Central

Festival da Salsicha

De acordo com TLC, as médias das amostras para qualquer população estarão distribuídas aproximadamente como uma distribuição normal em torno da média da população;

- População dos Maratonistas

ônibus	ônibus	ônibus	ônibus	ônibus	ônibus
$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$

Teorema do Limite Central

Festival da Salsicha

ônibus	ônibus	ônibus	ônibus	ônibus	ônibus
$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$

A maioria destas médias estarão próximas a da população e poucas serão ou mais baixas , ou mais altas.

- O TLC nos diz que as médias das amostras estarão distribuídas em torno da média da população aproximadamente numa distribuição normal;
- Isso é verdadeiro independentemente do aspecto da distribuição da população em questão. A população da qual as amostras estão sendo tiradas não precisa ter uma distribuição normal para que as médias das amostras esteja distribuídas normalmente;

Como regra prática, o tamanho da amostra deve ser de pelo menos 30 para validar o TLC

Teorema do Limite Central

Festival da Salsicha

ônibus	ônibus	ônibus	ônibus	ônibus	ônibus
$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$

O desvio padrão das médias é conhecido como ERRO PADRÃO

- O desvio padrão mede a dispersão da população;
  - Quanto os dados estão dispersos da média da população
- O erro padrão mede a dispersão das médias das amostras;
  - Quanto as médias das amostras estão dispersas da média delas

Teorema do Limite Central

Festival da Salsicha

ERRO PADRÃO é o desvio padrão das médias das amostras!!

$$EP = \frac{\sigma}{\sqrt{n}}$$

- Exemplo do ônibus: são 62 pessoas com média 95kg no ônibus
- Sabendo que a população de maratonistas tem média 74kg com desvio 16kg
- Temos que $EP = 16/\sqrt{62} = 2{,}03$

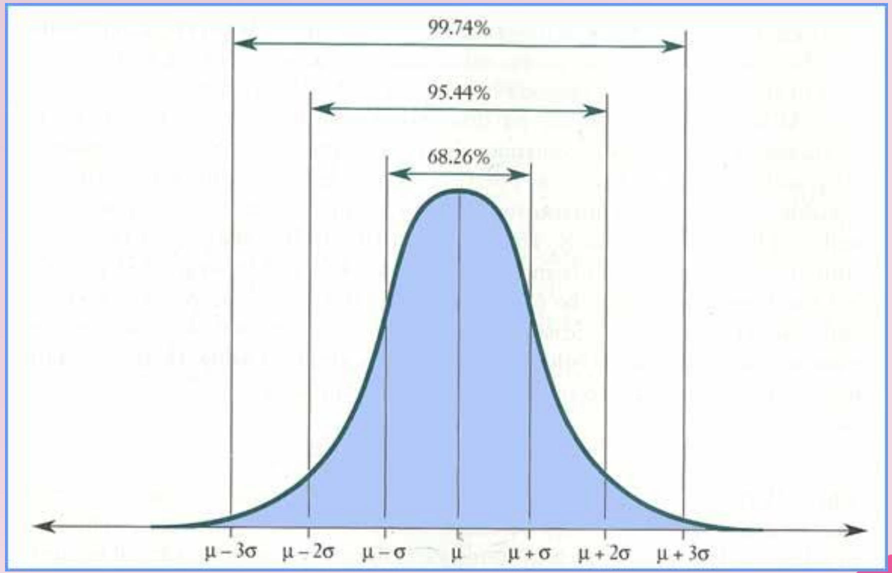
entendendo estatística divertidamente

Teorema do Limite Central

Festival da Salsicha

- Exemplo do ônibus: são 62 pessoas com média 95kg no ônibus
- Sabendo que a população de maratonistas tem média 74kg com desvio 16kg
- Temos que $EP = 16/\sqrt{62} = 2{,}03$
- Lembrando da normal:

Podemos dizer que com 99,74% de confiança o busão perdido (média 95kg) NÃO é dos maratonistas!



Teorema do Limite Central

FATOS

- Se você obteve amostras grandes, aleatórias, de qualquer população, as médias dessas amostras serão distribuídas normalmente em torno da média da população dela (foda-se a distribuição da população original);
- A maioria das médias de amostras estará razoavelmente perto da média da população; o ERRO PADRÃO quem define o tão perto;
- O TLC nos diz a probabilidade de que a média de uma amostra se situe dentro de certa distância da média da população. É relativamente improvável que uma média de amostra se situe a mais de dois erros padrões em relação à média da população e MUITO improvável que se situe à 3 EP´s;

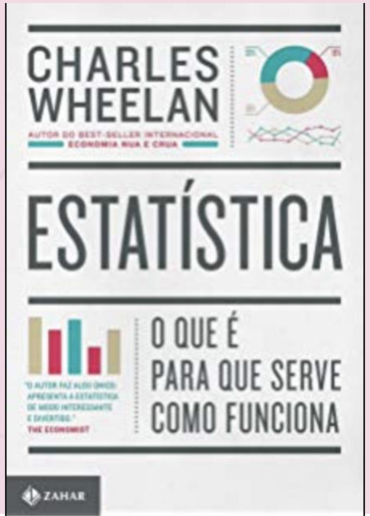
entendendo estatística divertidamente

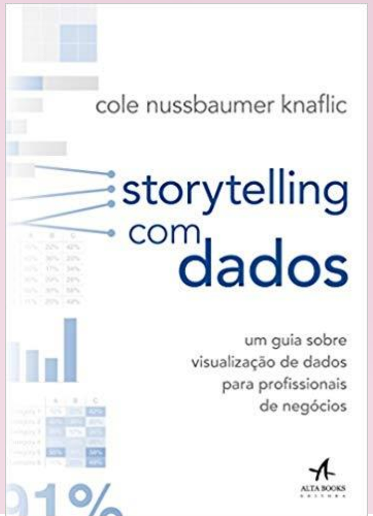
Não esqueça de deixar seu feedback!

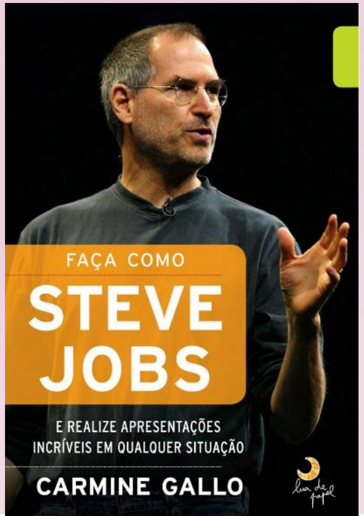
=]

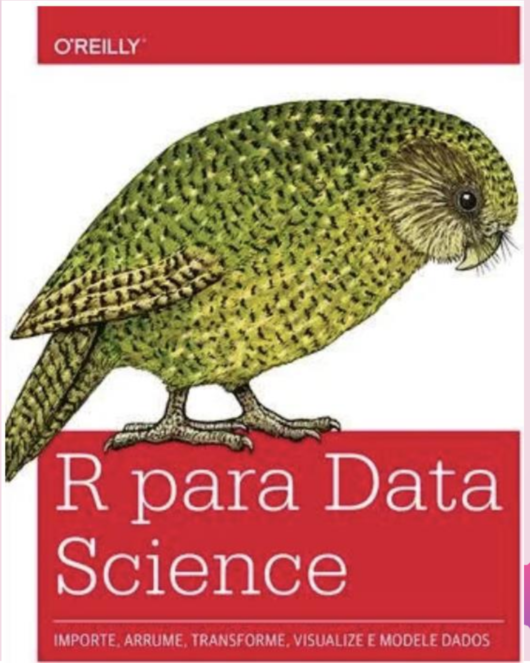
Referência

- Moore, D., McCabe, G., Duckworth, W., Sclove, S. *A prática da Estatística Empresarial*. LTC, Rio de Janeiro, 2006.
- Anderson, D., Sweeney, D., Williams, T. *Estatística Aplicada à Administração e Economia*. Segunda Edição. Cengage Learning, São Paulo, 2011.
- www.asn.rocks
- www.curso-r.com









entendendo estatística divertidamente

It´s kind of fun to do the IMPOSSIBLE

dri@asn.rocks

/in/adrianamms

/in/asn.rocks

asn.rocks

www.asn.rocks