EST 105

INICIAÇÃO À ESTATÍSTICA

RESUMO

Distribuições de probabilidade

Departamento de Estatística – UFV
Av. Peter Henry Rolfs, s/n
Campus Universitário
36570.977 – Viçosa, MG
http://www.det.ufv.br/

Distribuição Normal

É a mais importante das distribuições contínuas de probabilidades, sendo aplicada para a modelagem de inúmeros fenômenos e constantemente utilizada para o desenvolvimento teórico da inferência estatística. É também conhecida como distribuição de Gauss, Laplace ou Laplace-Gauss.

i. Características:

Seja X uma variável aleatória contínua:

$$X \sim Normal(\mu, \sigma^2)$$

Lê-se: A variável X segue distribuição normal com parâmetros μ e σ^2 .

em que:

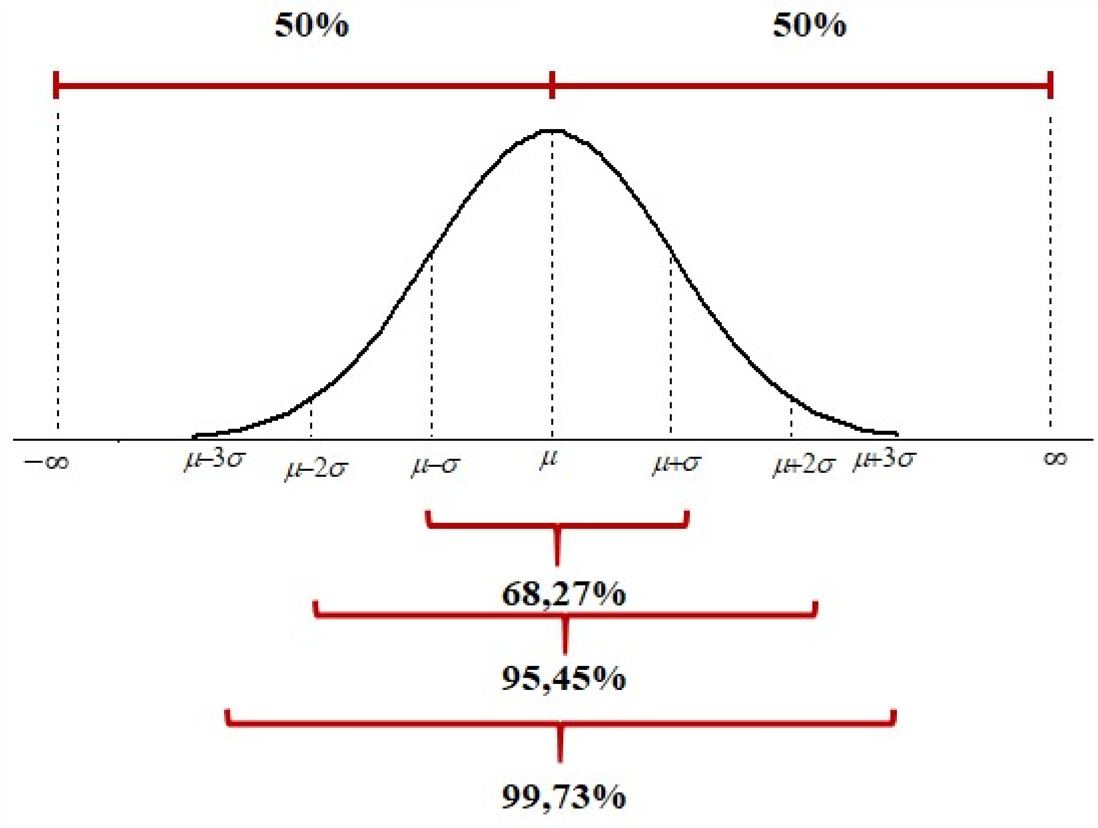
- $-\infty < x < +\infty$
- μ e σ^2 são parâmetros;
- $\mu = E(X)$ ou **média** da v. a. c. X, tal que $-\infty < \mu < \infty$;
- $\sigma^2 = V(X)$ ou **variância** da v. a. c. X, tal que $\sigma^2 > 0$.

ii. Função densidade de probabilidade ou f. d. p.

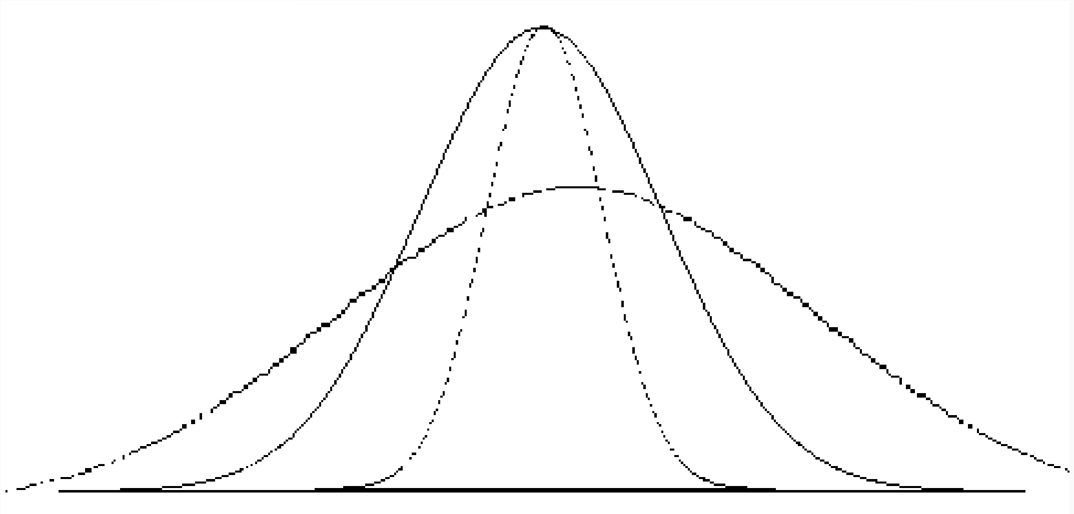
A função densidade de probabilidade é dada por:

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

iii. Representação gráfica:



É um gráfico em forma de sino. O seu posicionamento em relação ao eixo das ordenadas e seu achatamento vai ser determinado pelos parâmetros μ e σ^2 , respectivamente.



iv. Propriedades:

P1. $f(x)$ possui um ponto de máximo para $x = \mu$;

P2. $f(x)$ tem dois pontos de inflexão cujas abscissas valem $\mu - \sigma$ e $\mu + \sigma$;

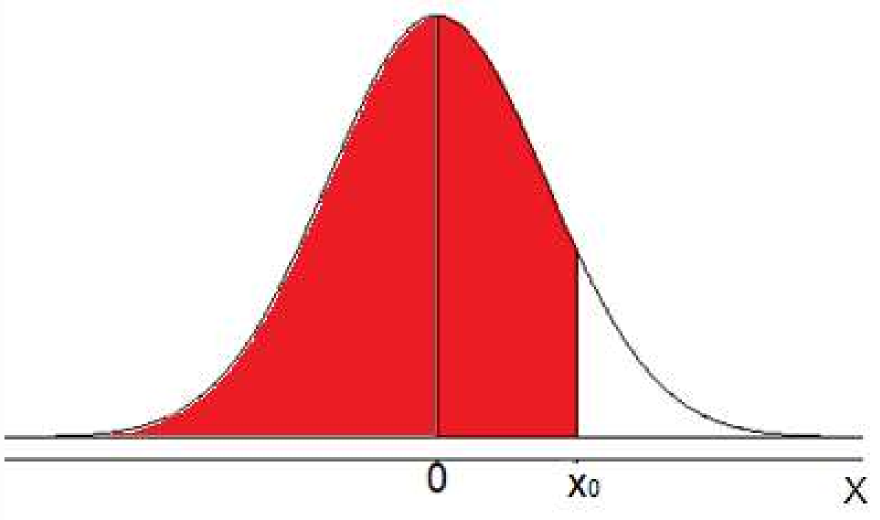
P3. $f(x)$ é simétrica em relação a $x = \mu$, ou seja, $\mu = Mo = Md$;

P4. $f(x)$ tende a zero quando x tende para $\pm\infty$ (assintótica em relação ao eixo x).

v. Cálculo de probabilidades

Por exemplo:

$$P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x_0} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



Para o cálculo de probabilidade, no caso de uma v. a. c. normal, surgem dois problemas:

1º) Integração de $f(x)$, pois para o seu cálculo é necessário o desenvolvimento em séries de potência;

2º) Elaboração de tabelas que já forneçam resultados do tipo $P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$, para todo $-\infty < x < \infty$. Como a f. d. p. depende de dois parâmetros, μ e σ^2, e ainda dos valores de x, essa tarefa se torna inviável para todas as infinitas combinações possíveis de valores.

Estes problemas são resolvidos pela padronização dos valores, obtendo-se assim a distribuição normal padronizada ou reduzida.

vi. Variável Normal Padronizada (Z)

É obtida por meio de uma transformação linear da variável aleatória normal X, obtendo-se assim uma nova variável aleatória Z, na qual a média será sempre nula ou igual a zero, e o desvio padrão será sempre igual a um.

$$X \sim Normal(\mu, \sigma^2) \qquad \Rightarrow \qquad Z \sim Normal(0,1)$$

$$\boldsymbol{Z = \frac{X - \mu}{\sqrt{\sigma^2}} = \frac{X - \mu}{\sigma}}$$

- **Média:**

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma}[E(X) - E(\mu)] = \frac{1}{\sigma}[\mu - \mu] = 0$$

- **Variância**

$$V(Z) = V\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X - \mu) = \frac{1}{\sigma^2} V(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

Conclusão: Se $X \sim Normal(\mu, \sigma^2)$ e $Z = \frac{X-\mu}{\sigma}$ então $Z \sim Normal(0,1)$, para quaisquer valores de x, μ e σ^2. Portanto, será possível tabelar as probabilidades, $P(X \leq x) = P(Z \leq z)$, em função dos valores de Z.

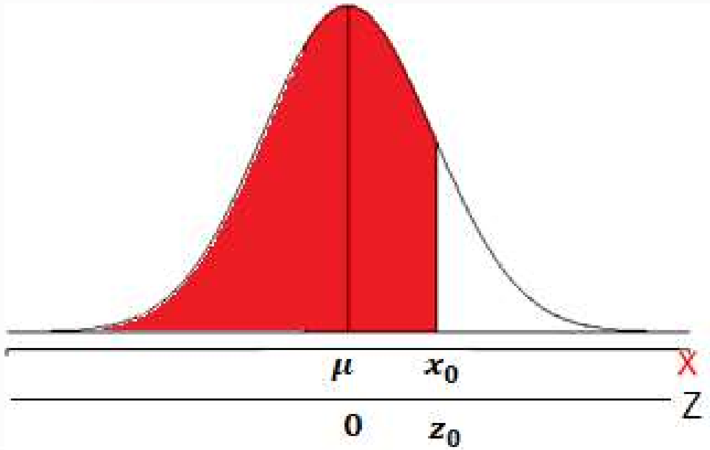
A função densidade de probabilidade ou f. d. p de Z é dada por:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

para $-\infty < z < +\infty$.

Isto implica que:

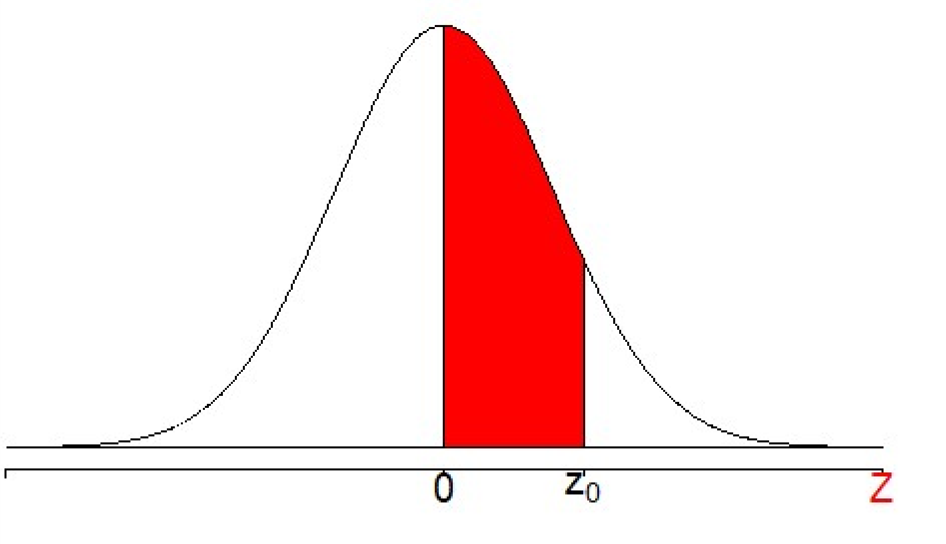
$P(X \leq x_0) = P(Z \leq z_0)$



vii. Tabela da Distribuição Normal Padrão

Há vários tipos de tabelas que nos fornecem as probabilidades sob a curva normal. A tabela que vamos utilizar é aquela que fornece a probabilidade da variável Z assumir um valor entre zero e um particular z_0, ou seja:

Área hachurada:



$$P(0 \leq Z \leq z_0) = \int_0^{z_0} \varphi(z)dz = \frac{1}{\sqrt{2\pi}} \int_0^{z_0} e^{-\frac{z^2}{2}} dz$$

Tabela 1 - Áreas de uma distribuição normal padrão. Cada casa na tabela dá a proporção sob a curva inteira entre **z**=0 e um valor positivo de **z**. As áreas para os valores de **z** negativos são obtidas por simetria,

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	0,0754
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	0,2258	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	0,2996	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	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Esta tabela foi adaptada do livro Estatística. Costa Neto, P.L.O. Ed. Edgard Blucher.

Exemplo 1

Seja Z uma v.a. tal que Z ~ Normal (0,1). Pede-se:

a. $P(0 \leq Z \leq 0,12)$

b. $P(-0,12 \leq Z \leq 0)$

c. $P(Z \geq 1,13)$

d. $P(Z \leq 1,13)$

e. $P(Z \geq -0,63)$

f. $P(-0,32 \leq Z \leq 0,63)$

g. $\mathrm{P}(0,32 \leq Z \leq 0,63)$

UFV

Exemplo 2

Suponha que o tempo necessário para atendimento de clientes em uma central telefônica (X) apresente distribuição normal com média de 8 minutos e desvio-padrão de 2 minutos. Pede-se:

a) Calcule a probabilidade de ocorrer um valor de X maior que 11 minutos?

b) Calcule a probabilidade de ocorrer um valor de X entre 7 e 11 minutos?

c) Em 200 chamadas telefônicas, quantas terão tempo de atendimento entre 5,5 e 7 minutos?

d) 30% das chamadas telefônicas requerem tempo de atendimento de 8 a x_0 minutos. Encontre o valor de x_0 .

e) 75% das chamadas telefônicas requerem no mínimo x_0 minutos. Encontre o valor de x_0 .

Teorema da combinação linear

Sejam $\boldsymbol{X}$ **e** $\boldsymbol{Y}$ **variáveis aleatórias independentes** cujas distribuições de probabilidade sejam dadas por $X \sim Normal(\mu_X, \sigma_X^2)$ e $Y \sim Normal(\mu_Y, \sigma_Y^2)$. Seja W uma nova variável aleatória definida pela seguinte combinação linear de X e Y:

$$W = aX + bY + c$$

em que a, b e c são constantes. Assim, W possui distribuição normal com média e variância dadas, respectivamente, por:

$$E(W) = E(aX + bY + c) = E(aX) + E(bY) + E(c) = aE(X) + bE(Y) + c$$
$$= a\mu_X + b\mu_Y + c$$

$$V(W) = V(aX + bY + c) = V(aX + bY) = V(aX) + V(bY) + 2Cov(aX, bY)$$
$$= a^2V(X) + b^2V(Y) + 2abCov(X, Y)$$

No entanto, como X e Y são independentes então $Cov(X,Y) = 0$

$$V(W) = a^2V(X) + b^2V(Y) = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

Portanto,

$$W \sim Normal(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

Exemplo 3: Sejam $\boldsymbol{X}$ **e** $\boldsymbol{Y}$ **variáveis aleatórias independentes** tais que $X \sim N(60, 15)$ e $Y \sim N(20,\ 10{,}25)$. Calcule a $P(W \leq 159{,}6)$ quando W é uma variável aleatória dada pela seguinte combinação linear $W = 3X - 2Y - 6$.

Atividade Proposta

Resolver os exercícios do Roteiro de Aulas abaixo relacionados:

- Exercício 19 – pág. 153
- Exercício 22 – pág. 153
- Exercício 25 – pág. 153

Campus Viçosa:
Avenida Peter Henry Rolfs, s/n
CEP 36570-900
Viçosa - MG - Brasil | + 55 31 3899-2200

Campus Florestal:
Rodovia LMG 818, km 6
CEP 35690-000
Florestal - MG - Brasil | + 55 31 3536-3300

Campus Rio Paranaíba:
Rodovia MG-230, Km 8
CEP 38810-000
Rio Paranaíba- MG - Brasil | + 55 34 3855-9300

www.ufv.br

UFV
Universidade Federal de Viçosa