

Estatística Aplicada a Recursos Hídricos

Docente: Rachid Muleia

(rachid.muleia@uem.mz)

Mestrado em Gestão de Recursos Hídricos - DGEO/UEM

Tema: Distribuições Teóricas de Probabilidade de v.a.'s contínuas

Ano lectivo: 2023

Distribuições de Probabilidade para v.a.'s contínuas

- Para caracterizar completamente uma v.a. contínua, precisamos fornecer sua f.d.p.
- **Distribuições de Probabilidade para v.a.'s contínuas mais importantes:**
 - Distribuição Uniforme
 - Distribuição Exponencial
 - Distribuições Normal

Distribuição Uniforme

Definição: Dados os números reais a e b ($a < b$), uma v.a. X tem distribuição uniforme no intervalo $[a, b]$, se sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrário.} \end{cases}$$

- A **Esperança** e **Variância** de X são definidas como:

$$E(X) = \frac{(a+b)}{2} \quad \text{e} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

- E a **Função de distribuição** é:

$$F(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x-a}{b-a}, & \text{se } a < x < b. \\ 1, & \text{se } x \geq b. \end{cases}$$

- Para indicar que uma v.a. X segue distribuição Uniforme contínua no intervalo $[a, b]$, denota-se $X \sim U(a, b)$

Distribuição Uniforme

Exemplo 1: Quando números reais são arredondados para os inteiros que lhes são mais próximos, os erros de arredondamento assumem valores entre $-0,5$ e $+0,5$. Podemos então considerar que X se comporta como uma v.a. uniforme contínua no intervalo $[-0,5; 0,5]$ cuja função de densidade é

$$f(x) = \begin{cases} 1, & \text{se } -0,5 \leq x \leq 0,5 \\ 0, & \text{caso contrário.} \end{cases}$$

Calcule $P(-0,2 \leq X \leq 0)$ e $P(0,1 \leq X \leq 0,3)$.

Distribuição Uniforme

Exemplo 1: Quando números reais são arredondados para os inteiros que lhes são mais próximos, os erros de arredondamento assumem valores entre $-0,5$ e $+0,5$. Podemos então considerar que X se comporta como uma v.a. uniforme contínua no intervalo $[-0,5; 0,5]$ cuja função de densidade é

$$f(x) = \begin{cases} 1, & \text{se } -0,5 \leq x \leq 0,5 \\ 0, & \text{caso contrário.} \end{cases}$$

Calcule $P(-0,2 \leq X \leq 0)$ e $P(0,1 \leq X \leq 0,3)$.

$$P(-0,2 \leq X \leq 0) = \int_{-0,2}^0 1 dx = x \Big|_{-0,2}^0 = 0 + 0,2 = 0,2$$

$$P(0,1 \leq X \leq 0,3) = \int_{0,1}^{0,3} 1 dx = x \Big|_{0,1}^{0,3} = 0,3 - 0,1 = 0,2$$

Distribuição Uniforme

Exemplo 2: Helena é estudante de Informática e todos os dias vai de TPM até sua faculdade. Há somente um TPM que lhe serve, e ele costuma passar pela paragem em qualquer instante entre 7h e 7h30. Se num certo dia Helena chega à paragem às 7h24, qual a probabilidade de que ela consiga pegar o TPM?

Resposta: Seja X o instante da chegada do TPM na paragem. Então $X \sim U(0, 30)$, onde X é medida em minutos e 0 (zero) corresponde às 7h. Pede-se $P(X > 24)$:

Distribuição Uniforme

Exemplo 2: Helena é estudante de Informática e todos os dias vai de TPM até sua faculdade. Há somente um TPM que lhe serve, e ele costuma passar pela paragem em qualquer instante entre 7h e 7h30. Se num certo dia Helena chega à paragem às 7h24, qual a probabilidade de que ela consiga pegar o TPM?

Resposta: Seja X o instante da chegada do TPM na paragem. Então $X \sim U(0, 30)$, onde X é medida em minutos e 0 (zero) corresponde às 7h. Pede-se $P(X > 24)$:

$$P(X > 24) = P(24 < X \leq 30) = \int_{24}^{30} \frac{1}{30} dx = \frac{1}{30} x \Big|_{24}^{30} = \frac{6}{30} = 0,2$$

Distribuição Exponencial

- A distribuição exponencial é usada para medir o intervalo de **tempos entre a ocorrência de eventos sucessivos**. Isto é, são usadas para modelar fenômenos como os seguintes:
 - tempo entre duas chegadas consecutivas de navios a um porto;
 - tempo entre duas chegadas de clientes a um supermercado;
 - tempo entre as chegadas de dois e-mails consecutivos à caixa de entrada de uma pessoa.
- Por isso é que a distribuição exponencial está fortemente relacionada à distribuição de Poisson:

Distribuição Exponencial

Definição: Diz-se que X tem uma distribuição exponencial com parâmetro λ ($\lambda > 0$) se a *f.d.p* de X for dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases}$$

- A **Esperança** e **Variância** de X são definidas como:

$$E(X) = \frac{1}{\lambda} \quad \text{e} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

- E a **Função de distribuição** é:

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1 - e^{-\lambda x}, & \text{se } x \geq 0. \end{cases}$$

- Para indicar que uma v.a. X segue distribuição exponencial com parâmetro λ , denota-se $X \sim \text{Exp}(\lambda)$

Distribuição Exponencial

Exemplo 3: Uma v.a. contínua X tem *f.d.p* dada por

$$f(x) = \begin{cases} \frac{k}{2}e^{-x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases}$$

i) Calcule o valor de k ; ii) Determine $F(x)$

Distribuição Exponencial

Exemplo 3: Uma v.a. contínua X tem *f.d.p* dada por

$$f(x) = \begin{cases} \frac{k}{2}e^{-x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases}$$

i) Calcule o valor de k ; ii) Determine $F(x)$

Resolução:

i) Obtenção do valor de k :

$$\int_0^{\infty} \frac{k}{2}e^{-x} dx = 1 \Leftrightarrow \frac{k}{2}(-e^{-x}) \Big|_0^{\infty} = 1 \rightarrow k = 2$$

ii) Obtenção de $F(x)$

$$F(x) = \int_0^x e^{-t} dt = (-e^{-t}) \Big|_0^x = 1 - e^{-x}$$

Logo

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1 - e^{-x}, & \text{se } x \geq 0. \end{cases}$$

Distribuição Exponencial

Exemplo 4: Suponha que o tempo de resposta X em um terminal de computador on-line específico (o tempo entre o final de uma consulta de um usuário e o começo da resposta do sistema para essa consulta) tenha distribuição exponencial com tempo de resposta esperado igual a 5 segundos. Obtenha a probabilidade de o tempo de resposta ser no máximo 10 segundos; e estar entre 5 e 10 segundos.

Distribuição Exponencial

Exemplo 4: Suponha que o tempo de resposta X em um terminal de computador on-line específico (o tempo entre o final de uma consulta de um usuário e o começo da resposta do sistema para essa consulta) tenha distribuição exponencial com tempo de resposta esperado igual a 5 segundos. Obtenha a probabilidade de o tempo de resposta ser no máximo 10 segundos; e estar entre 5 e 10 segundos.

Resolução: Seja X : “tempo de resposta...”. Tem-se que $E(X) = 5$, portanto, $\frac{1}{\lambda} = 5 \rightarrow \lambda = \frac{1}{5}$, ou seja,

$$f(x) = \frac{1}{5} e^{-\frac{1}{5}x}$$

$$F(x) = 1 - e^{-\frac{1}{5}x}$$

$$P(X \leq 10) = F(10) = 1 - e^{-2} = 1 - 0,135 = 0,865$$

$$\begin{aligned} P(5 \leq X \leq 10) &= F(10) - F(5) = (1 - e^{-2}) - (1 - e^{-1}) \\ &= e^{-1} - e^{-2} = 0,233 \end{aligned}$$

Relações dist. Exponencial e Poisson

- Existe uma interessante relação entre a distribuição exponencial e a distribuição de Poisson:
- **Preposição:** Suponha que o número de eventos que ocorrem em um intervalo de tempo de duração t tenha distribuição de Poisson com parâmetro λt (onde λ , a taxa do processo do evento, é o número esperado de eventos que ocorrem em uma unidade de tempo) e que os números das ocorrências em intervalos não-sobrepostos sejam independentes um do outro. Então a distribuição do tempo decorrido entre a ocorrência de dois eventos sucessivos é exponencial com parâmetro λ
- O parâmetro λ da exponencial trata-se da **frequência média de ocorrências do fenômeno considerado por unidade de tempo**
- Por outro lado, $\tau = \frac{1}{\lambda}$ é o valor médio desse intervalo de tempo, à luz da distribuição exponencial.

Distribuição Normal

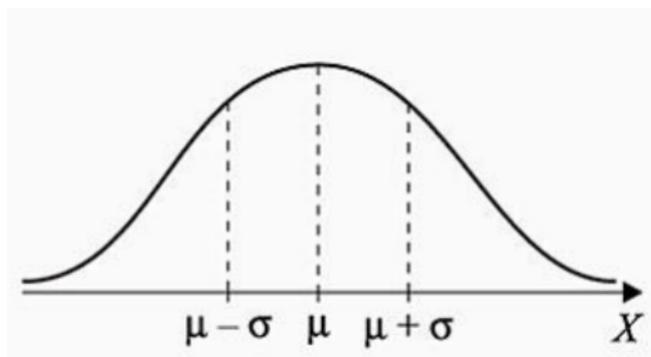
- A distribuição normal é a mais importante de todas em cálculo de probabilidade e estatística.
- Muitas populações numéricas possuem distribuições que podem ser ajustadas aproximadamente por uma curva normal ou gaussiana apropriada;
- Os dois parâmetros que a caracterizam são μ , que especifica o seu valor central, e σ^2 , que define a sua variabilidade.
- **Definição:** Diz-se que uma v.a. contínua X possui uma **distribuição normal** com parâmetros μ e σ^2 (onde $-\infty < \mu < \infty$ e $\sigma > 0$), simbolicamente denotado por $X \sim N(\mu, \sigma^2)$, se a f.d.p. de X é dada por

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

- Novamente, e denota a base do sistema de logaritmos naturais e é aproximadamente igual a 2,71828; e π representa a constante matemática familiar com valor aproximado 3,14159

Distribuição Normal

O gráfico de $f(x)$ (f.d.p. de X) é:



As principais características dessa função são:

- i) O ponto máximo de $f(x)$ é o ponto $X = \mu$.
- ii) Os pontos de inflexão da função são: $X = \mu + \sigma$ e $X = \mu - \sigma$
- iii) A curva é simétrica com relação a μ
- iv) $f(x) \rightarrow 0$ quando $x \rightarrow \pm\infty$
- v) A **esperança** de X é $E(X) = \mu$ e a **variância** é $Var(X) = \sigma^2$.

Distribuição Normal Padrão

Para calcular $P(a \leq X \leq b)$ quando $X \sim N(\mu, \sigma^2)$, faz-se:

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx.$$

- Nenhuma das técnicas de integração-padrão podem ser usadas para calcular a integral acima (só por métodos numéricos)
- Por essa razão, as probabilidades para a distribuição normal são calculadas com auxílio de tabelas, quando $\mu = 0$ e $\sigma = 1$;
- Seja $X \sim N(\mu, \sigma^2)$, define-se uma v.a. Z , $z = \frac{x-\mu}{\sigma}$
- Demonstra-se que Z tem distribuição normal com $\mu = 0$ e $\sigma = 1$. Nesse caso, diz-se que Z **tem distribuição normal padrão** ou **distribuição normal reduzida**, $Z \sim N(0, 1)$. A f.d.p. de Z é

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

- A $F(z) = P(Z \leq z) = \int_{-\infty}^z f(t)dt$, denotada por $\Phi(z)$.

Distribuição Normal Padrão

- A tabela de $Z \sim N(0, 1)$ fornece valores de $\Phi(z) = P(Z \leq z)$, área abaixo do gráfico da f.d.p. normal padrão à esquerda de z .

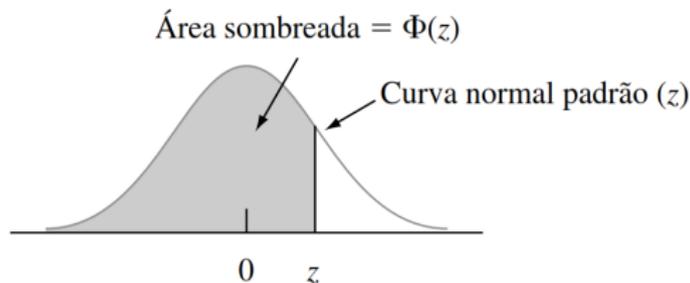


Tabela A.3 Área sob a Curva Normal Padronizada (cont.)

z	$\Phi(z) = P(Z \leq z)$									
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3482
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170

Distribuição Normal Padrão

Exemplo 5: Calcule as seguintes probabilidades normais padrão:

i) $P(Z \leq 1,25)$, ii) $P(Z > 1,25)$, iii) $P(Z \geq -1,25)$ e iv) $P(-0,38 \leq Z \leq 1,25)$.

Distribuição Normal Padrão

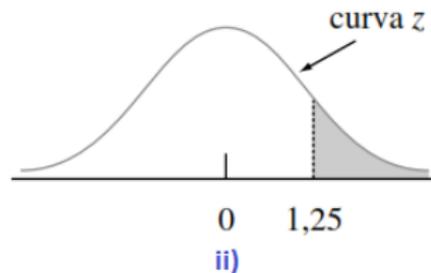
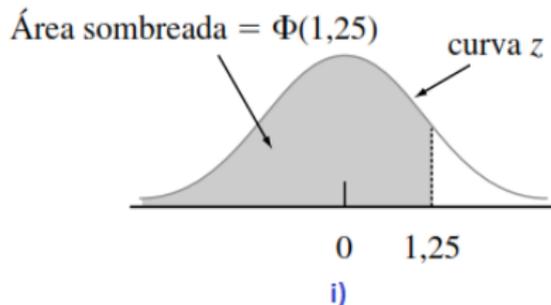
Exemplo 5: Calcule as seguintes probabilidades normais padrão:

i) $P(Z \leq 1,25)$, ii) $P(Z > 1,25)$, iii) $P(Z \geq -1,25)$ e iv) $P(-0,38 \leq Z \leq 1,25)$.

Resolução:

i) $P(Z \leq 1,25) = \Phi(1,25)$. Usando a tabela de $Z \sim N(0,1)$, tem-se que $P(Z \leq 1,25) = 0,8944$

ii) $P(Z > 1,25) = 1 - P(Z \leq 1,25) = 1 - \Phi(1,25) = 0,1056$



Distribuição Normal Padrão

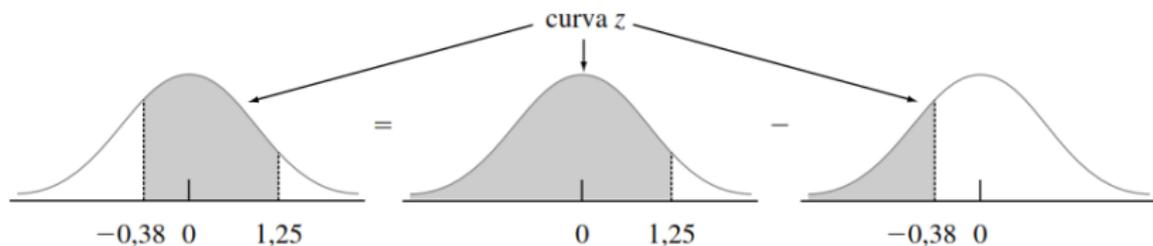
Resolução (Cont.):

iii) $P(Z \leq -1,25) = \Phi(-1,25) = 0,1056$.

iv) $P(-0,38 \leq Z \leq 1,25)$?

Lembre-se que se X é uma v.a. contínua com $f(x)$ e $F(x)$, então $P(a \leq X \leq b) = F(b) - F(a)$. Usando esse resultado tem-se

$$\begin{aligned}P(-0,38 \leq Z \leq 1,25) &= P(Z \leq 1,25) - P(Z \leq -0,38) \\ &= \Phi(1,25) - \Phi(-0,38) \\ &= 0,8944 - 0,3520 = 0,5424\end{aligned}$$



Distribuição Normal Não Padrão

- Quando $X \sim N(\mu, \sigma^2)$, as probabilidades que envolvem X são calculadas por “padronização”
- **Proposição:** Se X tem distribuição normal com média μ e desvio padrão σ , então a v.a.

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

tem distribuição normal padrão. Assim

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \qquad P(X \geq b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

- Significa que qualquer probabilidade que envolva X é expressa como uma probabilidade que envolve uma va normal padrão Z .

Distribuição Normal Não Padrão

Exemplo 5: Doentes, sofrendo de uma doença grave, são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma distribuição Normal, de média 15 e desvio padrão 2 (em dias).

- i) Que proporção desses pacientes demora mais de 17 dias para se recuperar?
- ii) Qual é a probabilidade de um paciente, escolhido ao acaso, apresentar tempo de cura inferior a 20 dias?
- iii) Qual é o tempo máximo necessário para que a recuperação de 25% dos pacientes?
- iv) Se 100 pacientes são escolhidos aleatoriamente, qual é o número esperado de curados em menos de 11 dias?

Distribuição Normal Não Padrão

Exemplo 5 (Resolução): Seja X “O tempo de cura”, temos $X \sim N(\mu = 15, \sigma^2 = 4)$. Portanto,

$$Z = \frac{X - 15}{\sqrt{4}}$$

i) Proporção de pacientes que demoram mais de 17 dias para se recuperar.

$$\begin{aligned} P(X > 17) &= P\left(Z > \frac{17 - 15}{2}\right) = P(Z > 1) \\ &= 1 - P(Z \leq 1) = 1 - \Phi(1) = 1 - 0,8413 = 0,1587. \end{aligned}$$

Ou seja, cerca de 16% dos pacientes demoram mais de 17 dias para se recuperar.

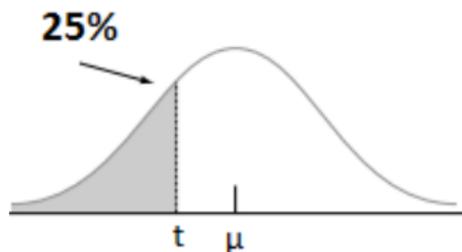
ii) A probabilidade de um paciente, escolhido ao acaso, apresentar tempo de cura inferior a 20 dias:

$$\begin{aligned} P(X < 20) &= P\left(Z < \frac{20 - 15}{2}\right) = P(Z < 2,5) \\ &= \Phi(2,5) = 0,9938. \end{aligned}$$

Distribuição Normal Não Padrão

Resolução (Cont.):

- iii) O tempo máximo necessário para que a recuperação de 25% dos pacientes. Significa que $P(X \leq t) = 0,25$



$$P(X \leq t) = P\left(Z \leq \frac{t - 15}{2}\right) = 0,25$$

Com o uso da tabela (e alguma malandragem) obtemos

$$\frac{t - 15}{2} = -0,67 \Rightarrow t = 13,66$$

Conclui-se, então, que 25% dos pacientes ficarão curados antes de, aproximadamente, 14 dias.

Distribuição Normal Não Padrão

Resolução (Cont.):

- iv) Se 100 pacientes são escolhidos aleatoriamente, qual é o número esperado de curados em menos de 11 dias?

Inicialmente, obtemos a probabilidade de um paciente qualquer ser curado em menos de 11 dias. Em seguida, essa probabilidade é interpretada como proporção de pacientes curados em menos de 11 dias e é multiplicada por 100 para obter a resposta.

$$\begin{aligned}P(X < 11) &= P\left(Z \leq \frac{11 - 15}{2}\right) = P(Z < -2) \\ &= \Phi(-2) = 0,0228 \quad (\text{para um paciente})\end{aligned}$$

Então, para 100 paciente, o número esperado com tempo de cura inferior a 11 dias será de $100 \times 0,0220 \simeq 2$ pacientes.

Distribuição Normal Não Padrão

Exemplo 6: Uma fábrica de carros sabe que os motores de sua fabricação têm duração normal com média de 150.000 *km* e desvio padrão de 5.000 *km*. Qual a probabilidade de que um carro, escolhido ao acaso, dos fabricados por essa firma, tenha um motor que dure:.

- i) menos de 170.000 *km*?
- ii) entre 140.000 *km* e 165.000 *km*?
- iii) Se a fábrica substitui o motor que apresenta duração inferior à garantia, qual deve ser esta garantia para que a percentagem de motores substituídos seja inferior a 0,2%?

Distribuição Normal Não Padrão

Exemplo 6 (Resolução): Seja X : “Duração de motor em km ”:
 $X \sim N(150.000, 5.000^2)$. Portanto

$$Z = \frac{X - 150.000}{5000}$$

i) Probabilidade de o motor durar menos de 170.000 km :

$$\begin{aligned} P(X < 170.000) &= P\left(Z < \frac{170000 - 150000}{5000}\right) = P(Z < 4) \\ &= \Phi(4) = 0,9998. \end{aligned}$$

ii) Probabilidade de o motor durar entre 140.000 km e 165.000 km , ou seja,
 $P(14000 < X < 165.000) = ?$

$$\begin{aligned} P\left(\frac{140000 - 150000}{5000} < Z < \frac{165000 - 150000}{5000}\right) &= \\ = P(-2 \leq Z \leq 3) &= \Phi(3) - \Phi(-2) = 0,9987 - 0,0228 = 0,9759. \end{aligned}$$

Distribuição Normal Não Padrão

Resolução (Cont.):

iii) Significa que $P(X \leq t) = 0,002$

$$P(X \leq t) = P\left(Z \leq \frac{t - 150000}{5000}\right) = 0,002$$

Com o uso da tabela (e alguma reflexão) obtemos

$$\frac{t - 150000}{5000} = -2,88 \Rightarrow t = 135.650$$

A garantia deve ser de 135.650 *km*.

Aproximação Normal para a Distribuição Binomial

- Lembre-se que o valor médio e o desvio padrão de uma variável aleatória $X \sim \text{Bin}(n, p)$ são, respectivamente, $E(X) = np$ e $\sigma_X = \sqrt{np(1-p)}$
- **Proposição:** Seja X uma v.a. binomial com base em n tentativas com probabilidade de sucesso p , ou seja, $X \sim \text{Bin}(n, p)$. Se $np \geq 10$ e $n(1-p) \geq 10$, (algumas pessoas menos conservadoras acreditam que para $np \geq 5$ e $n(1-p) \geq 5$ é suficiente) então X terá uma distribuição aproximadamente normal com $\mu = np$ e $\sigma = \sqrt{np(1-p)}$. Em particular, para x valor possível de X

$$P(X \leq x) \approx \Phi \left(\frac{x + 0,5 - np}{\sqrt{np(1-p)}} \right)$$

Exemplo 7: Suponha que 25% de todos os motoristas habilitados de uma determinada região não tenham seguro. Represente por X o número de motoristas sem seguro em uma amostra aleatória de tamanho 50 (o sucesso é o motorista não ter seguro), de forma que $p = 0,25$. Determine i) $P(X \leq 10)$ e ii) $P(5 \leq X \leq 15)$.

Aproximação Normal para a Distribuição Binomial

Exemplo 7 (Resolução): Observe que $X \sim \text{Bin}(50, 0, 25)$. Além disso, $E(X) = np = 50(0, 25) = 12,5 \geq 10$ e $n(1 - p) = 50(0, 75) = 37,5 \geq 10$, a aproximação pode ser aplicada com segurança, ou seja, $X \sim N[12, 5, (3, 06)^2]$.

i) $P(X \leq 10) = ?$

$$\begin{aligned} P(X \leq 10) &= P\left(Z \leq \frac{10 + 0,5 - 12,5}{3,06}\right) \\ &= P(Z \leq -0,65) = 0,2578 \end{aligned}$$

ii) $P(5 \leq X \leq 15) = ?$

$$\begin{aligned} &\approx P\left(Z \leq \frac{15 + 0,5 - 12,5}{3,06}\right) - P\left(Z \leq \frac{5 - 0,5 - 12,5}{3,06}\right) = \\ &= \Phi\left(\frac{15,5 - 12,5}{3,06}\right) - \Phi\left(\frac{4,5 - 12,5}{3,06}\right) \\ &= 0,8320 \end{aligned}$$