Métodos Estatísticos – L.EIC

Semana 5

Aula 5

4 de abril de 2022

Métodos Estatísticos – L.EIC

Aula 5

Distribuição por Amostragem

Populações Normais

Populações Não Normais

Teorema do Limite Central

Proporção Amostral

Aproximação Binomial $\to$ Normal

Teorema de Moivre-Laplace

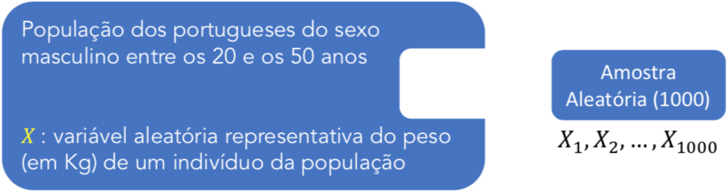
Correção de Continuidade

4

DISTRIBUIÇÃO por AMOSTRAGEM

Distribuição por Amostragem
Conceito e Considerações Gerais

Suponhamos que pretendemos estudar a **característica 'Peso'** na população constituída pelos portugueses do sexo masculino entre os 20 e os 50 anos.



$X_1, X_2, \ldots, X_{1000}$ – **variáveis aleatórias independentes e identicamente distribuídas**

Exemplo de uma amostra: $x_1, x_2, \ldots, x_{1000} \equiv 71, 82, \ldots, 63$

Distribuição por Amostragem
Conceito e Considerações Gerais

Este exemplo serve para ilustrar como proceder quando pretendemos conhecer certas **quantidades numéricas** acerca da população (parâmetros). Suponhamos que queríamos responder à questão: "Qual a **média do peso** dos portugueses do sexo masculino entre os 20 e os 50 anos"?

- Como é impraticável obter o peso de todos e calcular a média, recorre-se a uma **amostra** para estimar essa média.
- Uma vez que é possível recolher muitas amostras distintas da mesma dimensão, obtemos várias estimativas diferentes.
- Essas estimativas são os valores possíveis de uma função dos elementos da amostra, a que se dá o nome de estimador.

Distribuição por Amostragem
Conceito e Considerações Gerais

Estudo do Peso (X) de uma certa população, pretendendo-se **conhecer a média** μ, de X.

População dos portugueses do sexo masculino entre os 20 e os 50 anos

X : variável aleatória representativa do peso (em Kg) de um indivíduo da população

Amostra Aleatória (10)

$X_1, X_2, \ldots, X_{10}$

Estimador de μ: $\overline{X} = \frac{1}{10}(X_1 + X_2 + \cdots + X_{10})$

Para a amostra (71,82,79,90,63,50,55,46,65,60) **o estimador tem o valor 66.1**, e assim **66.1 é uma estimativa** de μ

Distribuição por Amostragem
Conceito e Considerações Gerais

A **variabilidade** entre diversas amostras aleatórias retiradas da mesma população pode ser caracterizada por uma **distribuição de probabilidade**.

Uma distribuição deste tipo denomina-se **distribuição por amostragem**.

Normalmente, uma amostra aleatória assemelha-se à população da qual foi selecionada, mas **é de esperar certas diferenças entre a amostra e a população**.

Uma distribuição por amostragem permite-nos tirar conclusões probabilísticas acerca de possíveis amostras.

Distribuição por Amostragem
Conceito e Considerações Gerais

Por exemplo podemos querer conhecer a probabilidade de numa amostra de dimensão 10, selecionada aleatoriamente de uma população, obtermos uma **média que não difira da média da população em mais do que** 10%?

Uma distribuição por amostragem é uma distribuição de probabilidade de uma variável aleatória baseada numa amostra genérica e cuja variabilidade resulta da aleatoriedade da amostragem.

Uma variável aleatória que é função de uma amostra aleatória genérica é designada de **estatística**.

Distribuição por Amostragem
Conceito e Considerações Gerais

Uma estatística é então uma função da amostra aleatória, sem parâmetros desconhecidos:

$$Y = f\left(X_1, X_2, \ldots, X_n\right)$$

Seja, por exemplo $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população com média desconhecida μ.

Exemplos de estatísticas:

- Média amostral
- Variância amostral
- Extremos da amostra
- Mediana amostral

Distribuição por Amostragem
Conceito e Considerações Gerais

Num estudo, consideramos os dados como uma amostra aleatória retirada de uma população.

Geralmente, apenas **dispomos de uma única amostra** selecionada aleatoriamente de uma população muito grande.

No entanto, para ter em linha de conta a variabilidade devido à amostragem, devemos considerar um quadro de referência mais amplo, de modo a incluir não apenas uma amostra, mas **todas as amostras possíveis** de serem extraídas da população.

Distribuição por Amostragem
Conceito e Considerações Gerais

População

X – variável aleatória que representa uma característica em estudo

Amostra aleatória (é uma amostra genérica)

$X_1, X_2, \ldots, X_n$ são variáveis aleatórias associadas às n observações

Amostra (é uma amostra específica)

$x_1, x_2, \ldots, x_n$ são os valores das variáveis aleatórias X_i

Notar que antes de se proceder à seleção da amostra não sabemos quais os valores que iremos obter. **Cada elemento da "futura" amostra é portanto representado por uma variável aleatória**.

Distribuição por Amostragem
Conceito e Considerações Gerais

População

X – variável aleatória que representa uma característica em estudo

Amostra aleatória (é uma amostra genérica)

$X_1, X_2, \ldots, X_n$ são variáveis aleatórias associadas às n observações

Amostra (é uma amostra específica)

$x_1, x_2, \ldots, x_n$ são os valores das variáveis aleatórias X_i

Se a amostragem for **aleatória simples com reposição**, então as variáveis aleatórias X_i são **independentes** e têm **distribuição igual à de X**.

Distribuição por Amostragem
Média Amostral

Considere-se a **média de uma amostra aleatória** genérica:

$$(X_1, X_2, \ldots, X_n)$$

Essa **média é também uma variável aleatória**, chamada média amostral, a qual é dada por:

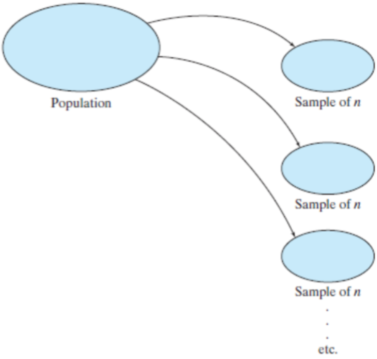
$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

A **distribuição de probabilidade de $\overline{X}$ é uma distribuição por amostragem**.

Como a determinamos?

Distribuição por Amostragem
Média Amostral

A média de uma amostra ($\overline{x}$), pode ser usada para caracterizar os dados amostrais, e para estimar a média da população (μ).

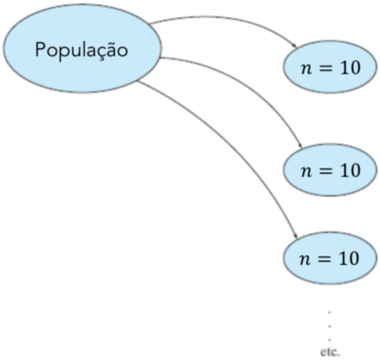


Estudo:

Obter diversas amostras de dimensão n, e o respetivo valor da média amostral $\overline{x}$, para uma dada população com média μ, e desvio padrão σ.

Distribuição por Amostragem
Média Amostral

A **variação do valor de** $(\overline{x})$ para cada uma das amostras é especificada pela **distribuição de probabilidade de** $(\overline{X})$, e é então a distribuição por amostragem (ou distribuição amostral) de $(\overline{X})$.



Exemplo – Peso (X):

Obter várias amostras de dimensão $n = 10$, da mesma população, e calcular o respetivo valor da média amostral $(\overline{x})$

Distribuição por Amostragem
Média Amostral

Exemplo

Consideremos a população seguinte:

$$\{1, 2, 3, 4, 6\}$$

e seja (X_1, X_2) uma amostra aleatória simples, com reposição, desta população.

Vejamos como determinar a distribuição por amostragem da **média amostral**:

$$\overline{X} = \frac{X_1 + X_2}{2}$$

O par (X_1, X_2) **pode tomar 25 valores** e cada um deles dá origem a um valor para a média, que **pode ser** 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 ou 6.

Notar que a média da população é 3.2

Distribuição por Amostragem
Média Amostral

Exemplo

A distribuição por amostragem de $\overline{X}$, é:

$\bar{x}$	1	1.5	2	2.5	3	3.5	4	4.5	5	6
$f(\bar{x})$	1/25	2/25	3/25	4/25	3/25	4/25	3/25	2/25	2/25	1/25

$$f(\overline{x}) = P(\overline{X} = \overline{x})$$

Qual a **probabilidade** de numa amostra de tamanho 2, selecionada aleatoriamente da população dada, **obtermos uma média superior à média da população**?

$$P(\overline{X} > 3.2) = \frac{12}{25} = 48\%$$

Distribuição por Amostragem

Média Amostral

Exercício

O tempo de vida, em anos, dos indivíduos de uma determinada espécie, segue uma distribuição normal de média 10 anos e desvio padrão 3.7 anos.

Qual a probabilidade de num grupo de 100 indivíduos escolhidos aleatoriamente a média do tempo de vida ser superior a 11 anos?

X_i – v.a. representativa do tempo de vida (em anos) do i-ésimo indivíduo (em 100)

$$X_i \sim N\left(10, 3.7^2\right)$$

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \sim N\left(10, \frac{3.7^2}{100}\right)$$

$$P(\overline{X} > 11) = P\left(\frac{\overline{X} - 10}{0.37} > \frac{11 - 10}{0.37}\right) \approx 1 - \phi(2.7)$$

Distribuição por Amostragem
Média Amostral

Mas em geral o que podemos afirmar acerca da variável aleatória representativa da média amostral

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

baseada numa amostra de tamanho n, selecionada aleatoriamente de uma população com média μ e desvio padrão σ?

Distribuição por Amostragem
Média Amostral

Média da distribuição

A média de $\overline{X}$ é igual à média da população:

$$\mu_{\overline{X}} = E(\overline{X}) = \mu_X$$

Desvio Padrão da distribuição

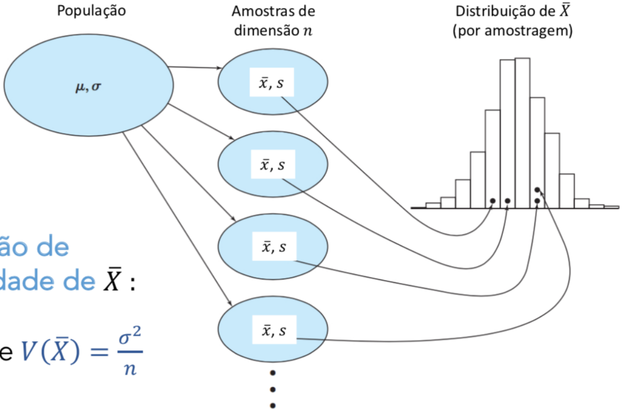
O desvio padrão de $\overline{X}$ é igual ao desvio padrão da população dividido pela raiz quadrada do tamanho da amostra:

$$\sigma_{\overline{X}} = \sqrt{V(\overline{X})} = \frac{\sigma_X}{\sqrt{n}}$$

Forma da distribuição

Se a população tiver distribuição normal então $\overline{X}$ tem também distribuição normal.

Distribuição por Amostragem
Média Amostral



Distribuição de Probabilidade de $\bar{X}$:

$$E(\bar{X}) = \mu \text{ e } V(\bar{X}) = \frac{\sigma^2}{n}$$

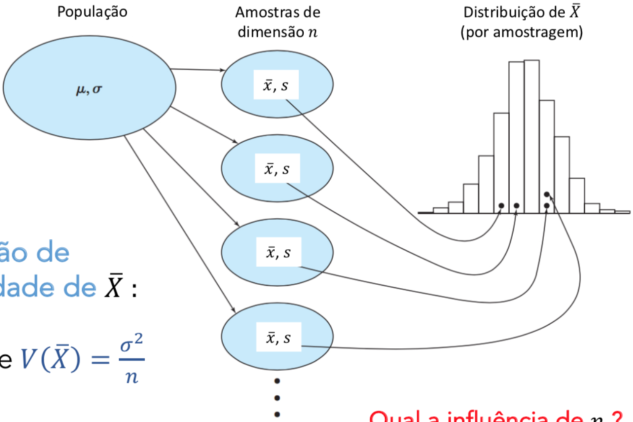
Distribuição por Amostragem
Média Amostral – Populações normais

Resultados Importantes

1. Sejam $X_1, X_2, \ldots X_n$, variáveis aleatórias independentes, tais que:
$$X_i \sim N\left(\mu_i, \sigma_i^2\right)$$
Então
$$X_1 + X_2 + \cdots + X_n \sim N\left(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)$$

2. Sejam agora $X_1, X_2, \ldots X_n$, v.a. independentes, e identicamente distribuídas, i.e.:
$$X_i \sim N\left(\mu, \sigma^2\right)$$
Então
$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuição por Amostragem
Média Amostral – Influência de *n*



Distribuição de Probabilidade de $\bar{X}$:

$E(\bar{X}) = \mu$ e $V(\bar{X}) = \frac{\sigma^2}{n}$

Qual a influência de n ?

Distribuição por Amostragem
Média Amostral – Influência de n

Exemplo

População	Tamanho da Amostra	Média Amostral
X	n	$\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

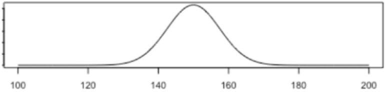
$X \sim N\left(150, 15^2\right)$

$$\begin{array}{ll} n = 4 & \overline{X} \sim N\left(150, 7.5^2\right) \\ n = 25 & \overline{X} \sim N\left(150, 3^2\right) \\ n = 225 & \overline{X} \sim N(150, 1) \end{array}$$

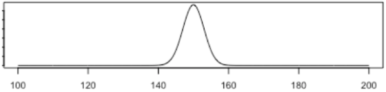
Distribuição por Amostragem
Média Amostral – Influência de n

A variabilidade de $\overline{X}$ diminui com o aumento do tamanho da amostra.

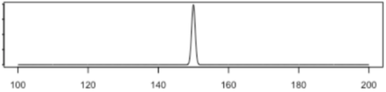
$\bar{X} \sim N(150, 7.5^2)$ $\quad n = 4$



$\bar{X} \sim N(150, 3^2)$ $\quad n = 25$



$\bar{X} \sim N(150,1)$ $\quad n = 225$



Distribuição por Amostragem
Média Amostral – Influência de n

Sejam $X_1, X_2, \ldots X_n$, variáveis aleatórias independentes, e identicamente distribuídas, com distribuição normal, i.e.:

$$X_i \sim N\left(\mu, \sigma^2\right)$$

Pretendemos analisar a distribuição de $\overline{X}$ para diferentes valores de n.

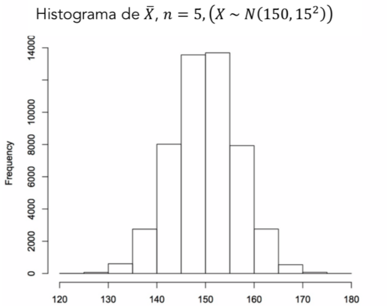
Distribuição por Amostragem

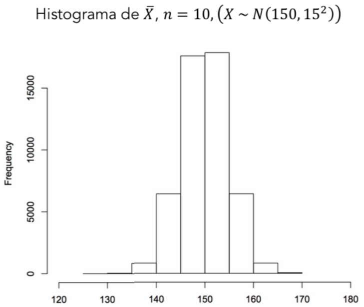
Média Amostral – Influência de n

Exemplo

Amostras aleatórias da distribuição normal, com média $\mu = 150$ e desvio padrão $\sigma = 15$, para

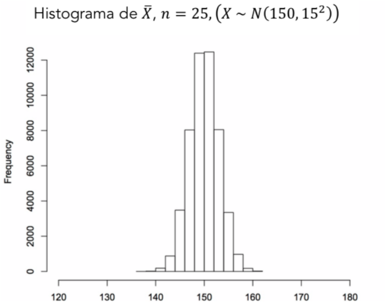
$$n = 5, 10, 25, 64$$

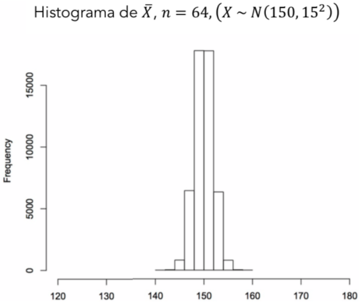




Distribuição por Amostragem
Média Amostral – Influência de n

Exemplo





Distribuição por Amostragem
Média Amostral – Influência de *n*

Qual a influência do valor de n?

- Note-se que um valor de $\overline{X}_{64}$ não é necessariamente mais próximo da média $\mu = 150$ do que um valor de $\overline{X}_5$
- Contudo, a probabilidade da média amostral estar mais próxima da média populacional (i.e. pertencer a um intervalo $\mu \pm \varepsilon$, com ε pequeno) é superior para $\overline{X}_{64}$

Teorema do Limite Central

Distribuição por Amostragem

Teorema do Limite Central (TLC)

Na distribuição por amostragem, vimos o caso em que a população tinha distribuição normal. E se não tiver?

Nesse caso usa-se um resultado teórico chamado Teorema do Limite Central, que é válido mesmo que a distribuição populacional não seja conhecida, e que diz o seguinte:

Se o tamanho da amostra n, for grande, então a média amostral $\overline{X}$ tem distribuição aproximadamente normal.

Na prática, a situação mais usual é desconhecermos a distribuição da população, e esse é um dos motivos pelos quais a distribuição normal é tão importante em estatística.

Distribuição por Amostragem

Teorema do Limite Central (TLC)

O TLC é um dos resultados mais importantes em Teoria das Probabilidades.

Esta é uma versão simplificada que corresponde à situação de v.a. **independentes e identicamente distribuídas** (i.i.d.)

TLC:

Seja $X_1, X_2, \ldots$, uma sucessão de variáveis aleatórias i.i.d., com

$$E\left(X_i\right) = \mu \quad e \quad V\left(X_i\right) = \sigma^2, \quad i = 1, 2, \ldots$$

Então, sendo $Z \sim N(0,1)$, temos:

$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = P(Z \leq z) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq z\right)$$

Distribuição por Amostragem

Teorema do Limite Central (TLC)

Quão grande tem que ser n para que a distribuição de $\overline{X}$ seja suficientemente próxima da distribuição normal?

A resposta a esta questão depende de quão próximo da distribuição normal é a distribuição da população.

Nesta unidade curricular usaremos **a regra** $n \geq 30$, que habitualmente garante uma **boa aproximação**.

Proporção Amostral

Distribuição por Amostragem

Proporção Amostral

O Teorema do Limite Central garante que a distribuição por amostragem da média $\overline{X}$ se aproxima da distribuição normal quando o tamanho da amostra aumenta.

Suponha-se agora que a **população** é **dicotómica** (há apenas duas observações possíveis):

- sucesso (1), com probabilidade p
- insucesso (0), com probabilidade $1 - p$.

Proporção Amostral

Seja $\widehat{P}$ a **proporção de sucessos** numa amostra de tamanho n retirada aleatoriamente dessa população.

Na verdade, $\widehat{P}$ **é uma média amostral** (soma dos 1 e 0 dividida por n).

Pelo Teorema do Limite Central, para n suficientemente grande, $\widehat{P}$ **tem distribuição normal**.

Distribuição por Amostragem

Proporção Amostral

Note-se que, numa amostra de tamanho n,

$$\widehat{P} = \frac{Y}{n}$$

onde Y é a v.a. representativa do nº de sucessos numa amostra de tamanho n.

Como

$$Y \sim B(n, p)$$

então

$$\mu_Y = np \quad \text{e} \quad \sigma_Y = \sqrt{np(1-p)}$$

Calculemos agora $\mu_{\hat{P}}$ e $\sigma_{\hat{P}}$.

Distribuição por Amostragem

Proporção Amostral

$$\mu_{\hat{P}} = \frac{1}{n}E(Y) = \frac{1}{n}np = p$$

$$\sigma_{\hat{P}} = \sqrt{V(\widehat{P})} = \sqrt{V\left(\frac{Y}{n}\right)} = \sqrt{\frac{1}{n^2}V(Y)}$$

$$= \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}}$$

Portanto, **se** n **for grand**e, a distribuição por amostragem da proporção $\widehat{P}$ pode ser **aproximada** por uma **distribuição normal** de média p e desvio padrão $\sqrt{\frac{p(1-p)}{n}}$.

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Se $\widehat{P}$ tem distribuição **aproximadamente normal**, então

$$Y = n\widehat{P}$$

tem também distribuição **aproximadamente normal**.

Como a **distribuição de Y é binomial**, conclui-se que, se n for **grande**, a **distribuição binomial** $B(n, p)$, pode ser **aproximada por uma distribuição normal** de média,

$$\mu = np$$

e desvio padrão,

$$\sigma = \sqrt{np(1-p)}$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

O Teorema de **Moivre-Laplace** é uma das primeiras tentativas de uso da distribuição normal como aproximação de outras distribuições.

Este teorema, estabelece que, **se X for uma v.a. com distribuição binomial**, de parâmetros n e p, e

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

então, se n é suficientemente grande ($n \to \infty$), Y **tem uma distribuição aproximadamente normal**, com média 0 e variância 1, e escreve-se

$$Y \dot{\sim} N(0,1)$$
$$n \to \infty$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Vimos já que uma v.a. binomial $Bi(n, p)$, pode ser considerada como a soma de n v.a. independentes de Bernoulli $B(1, p)$.

Sendo $X \sim Bi(n, p)$ e $Z \sim N(0, 1)$, para n elevado, tem-se:

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq z\right) \approx P(Z \leq z)$$

Nota: Nesta aproximação substitui-se uma **distribuição discreta** (binomial), por uma distribuição contínua (normal), o que torna necessária a chamada **correção de continuidade** que veremos mais à frente.

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Condições de aproximação

Há diversos critérios que podem ser utilizados, a título indicativo, para definir as condições de aproximação.

Na prática, as condições que vamos utilizar, são as seguintes:

- se X tem distribuição binomial de parâmetros n e p, i.e.:

$$X \sim Bi(n, p)$$

- e se

$$n > 25 \quad \text{e} \quad \min\{np, n(1-p)\} > 5$$

então

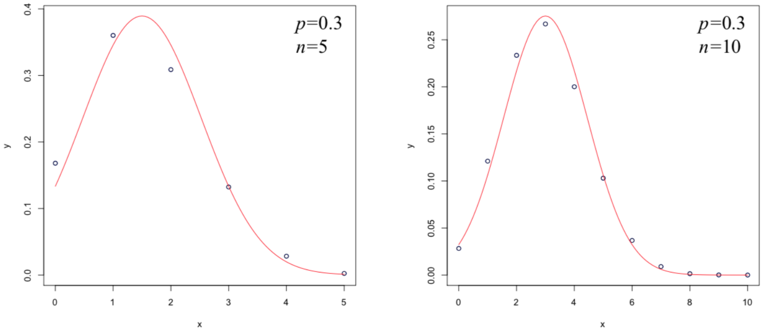
$$\frac{X - np}{\sqrt{np(1-p)}} \dot{\sim} N(0, 1)$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Ilustração gráfica do resultado anterior

$Bi(n,p)$ - pontos; $N\left(np, np(1-p)\right)$ - linha contínua

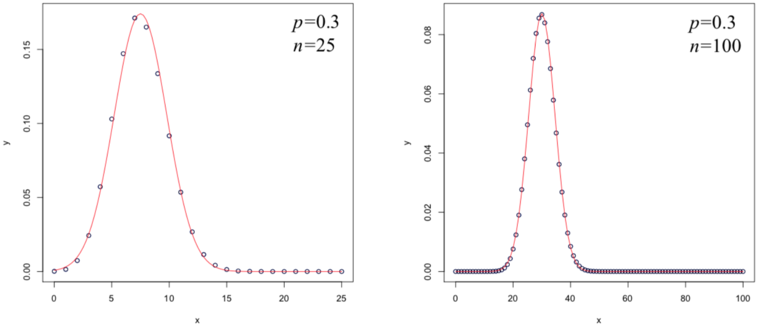


Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Ilustração gráfica do resultado anterior

$Bi(n, p)$ - pontos; $N(np, np(1-p))$ - linha contínua

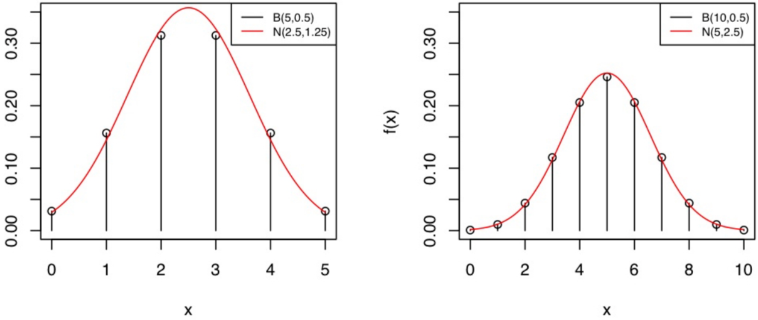


Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Ilustração gráfica do resultado anterior

$Bi(n,p)$ - pontos; $N(np, np(1-p))$ - linha contínua

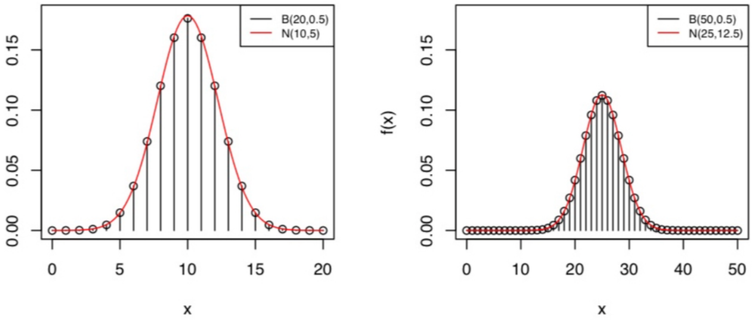


Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Ilustração gráfica do resultado anterior

$Bi(n,p)$ - pontos; $N\left(np, np(1-p)\right)$ - linha contínua



Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.)

Para melhorar a aproximação de uma distribuição discreta (p.e. binomial) por uma distribuição contínua (p.e. normal), é comum usar-se o que se designa por **correção de continuidade**, que consiste na associação de um intervalo a cada valor da variável aleatória discreta.

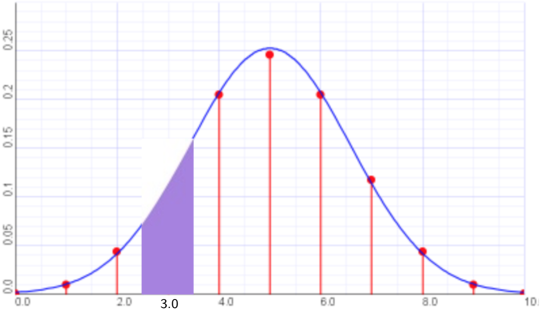
Por exemplo, ao valor 3 da variável aleatória discreta, associamos o intervalo $[2.5, 3.5]$, como se ilustra na figura a seguir.

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.)

Ao valor 3 da variável discreta, associamos o intervalo $[2.5, 3.5]$:



Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.)

A primeira coisa a fazer é verificar se são satisfeitas as condições de aproximação, que são:

$$X \sim Bi(n, p) \quad \text{e} \quad n > 25 \quad \text{e} \quad \min\{np, n(1-p)\} > 5$$

Satisfeitas estas condições, adotamos o procedimento adequado a cada caso.

Por exemplo, para calcular $P(X \leq k)$:

$$P(X \leq k) = P(X < k + 0.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} < \frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$\approx \phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.)

Se queremos $P(X \geq k)$, procedemos de modo análogo:

$$P(X \geq k) = P(X > k - 0.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} > \frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$\approx 1 - \phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

onde ϕ é a função de distribuição para a normal reduzida.

Esse procedimento (subtrair e somar 0.5) é conhecido como correção de continuidade de Fisher, especialmente recomendado **quando n não for muito grande**.

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.) – **Exemplo 1**

Analisemos o caso do cálculo de $P(X = k)$, mas vejamos primeiro, por exemplo, o caso em que $k = 3$.

Admitindo então que $X \sim B(n, p)$, e que são satisfeitas as condições de aproximação, tem-se:

$$P(X = 3) \approx P(3 - 0.5 < Y < 3 + 0.5)$$

onde

$$Y \sim N(np, np(1 - p))$$

De um modo geral:

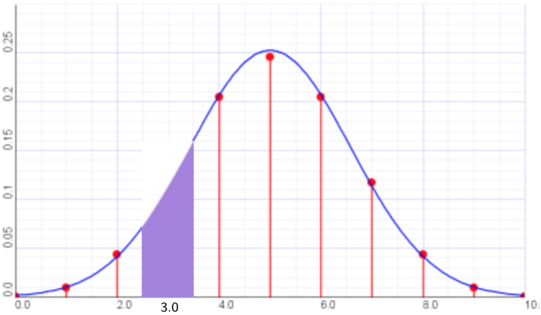
$$P(X = k) \approx P(k - 0.5 < Y < k + 0.5)$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade (C.C.) – **Exemplo 1**

$$P(X = 3) = P(2.5 \leq X \leq 3.5) \approx P(2.5 \leq Y \leq 3.5)$$



Se não fosse feito deste modo obteríamos o valor zero.

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Exemplo 2

Seja uma v.a. X, tal que:

$$X \sim Bi(n, p); \quad (n = 50,\ p = 0.3)$$

Calcular $P(X \geq 18)$.

Como conhecemos a distribuição de X, podemos calcular o **valor exato**, que é dado por:

$$P(X \geq 18) = \sum_{k=18}^{50} C_k^{50} 0.3^k (1 - 0.3)^{50-k} = 0.2178$$

Façamos a **aproximação à normal**, mas **sem C.C.**!

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Exemplo 2 – sem C.C.

Aproximando à normal, **sem C.C.**, virá:

$$X \sim Bi(n,p) \to Z = \frac{X - 50 \times 0.3}{\sqrt{50 \times 0.3(1-0.3)}} = \frac{X-15}{3.24} \dot{\sim} N(0,1)$$

$$P(X \geq 18) \approx P\left(Z \geq \frac{18-15}{3.24}\right) = P(Z \geq 0.93)$$

$$P(X \geq 18) \approx P(Z \geq 0.93) = 1 - P(Z < 0.93) = 0.1762$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Exemplo 2 – com C.C.

Aproximando à normal, com C.C., virá:

$$P(X \geq 18) \approx P\left(Z \geq \frac{17.5 - 15}{3.24}\right) = P(Z \geq 0.77)$$

$$P(X \geq 18) \approx 1 - P(Z \leq 0.77) = 0.2206$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade – **Exercício**

O tempo de vida (em anos), dos indivíduos de uma determinada espécie, segue uma distribuição normal.

A média da distribuição é 10 anos e o desvio padrão 3.7 anos.

Considerar um grupo de 200 indivíduos escolhidos aleatoriamente.

Qual a probabilidade de nesse grupo, pelo menos 50 indivíduos estarem vivos ao fim de 12 anos?

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade – **Exercício**

Sejam X e Y as variáveis aleatórias seguintes:

X – v.a. representativa do tempo de vida (em anos) dos indivíduos dessa espécie

Y – v.a. representativa do número de indivíduos entre os 200 escolhidos que estão vivos ao fim de 12 anos

Temos:

$$X \sim N\left(10, 3.7^2\right) \quad \text{e} \quad P(X \geq 12) \approx 0.2946$$

$$Y \sim Bi(200, p) \quad \text{onde} \quad \mathrm{p} = P(X \geq 12) \approx 0.2946$$

Distribuição por Amostragem

Aproximação da Distribuição Binomial pela Normal

Correção de Continuidade – **Exercício**

Verificam-se as condições de aproximação:

$$n > 25 \quad \text{e} \quad np \approx 58.92 > 5 \quad \Rightarrow \quad \frac{Y - np}{\sqrt{np(1-p)}} \dot{\sim} N(0,1)$$

Então

$$\frac{Y - 58.92}{\sqrt{41.562168}} \dot{\sim} N(0,1)$$

e

$$P(Y \geq 50) = P(Y > 49.5) = P\left(\frac{Y - 58.92}{\sqrt{41.562168}} > \frac{49.5 - 58.92}{\sqrt{41.562168}}\right)$$

i.e.:

$$P(Y \geq 50) \approx 1 - \phi(1.46) \approx 0.9279$$