Disciplina "Estatística Básica"
Aula:
Teorema Central de Limite e Assuntos Relacionados
Ministrante:
Prof. Dr. Vladimir Belitsky
Departamento de Estatística, IME-USP
Data de última modificação: 2023-09-15.
Anterior: 2019-05-03. Localização:
Aulas/MinhaAulaTCL/AulaTCL-2023-09-15

A aproximação de Binomial por Normal

A aproximação de distribuição Binomial por distribuição Normal é uma de muitas consequencias do TEOREMA CENTRAL DE LIMITE (TCL).

Concentramos nossa atenção nesta consequencia por que

(a) entre todas as consequenias do TCL, esta é uma das que mais se usa em cursos universitários sobre Probabilidade e Estatística Básicas de duração de um ou de dois semestres;

(b) a aproximação de Binomial por Normal permite uma fácil ilustração gráfica;

(c) quando falarmos sobre as condições gerais da aplicabilidade do TCL, a distribuição Binomial fornecerá um ambiente cómodo para a exemplificação de tais condições.

A aproximação de Binomial por Normal

Teorema 1 (aproximação de Binomial por Normal em sua formulação prática e sem detalhes).
Seja $B_{n,p}$ a variável aleatória Binomial com parâmetros n e p (isto é, $B_{n,p} \sim Bin(n;p)$, na notação introduzida quando definimos variáveis aleatórias binomiais; o índice "n, p" acrescentei por fins didáticos).
Seja Y a var. aleat. Normal de média np e variância $np(1-p)$ (isto é, $Y \sim \mathcal{N}(np, np(1-p))$, na notação introduzida quando definimos variáveis aleatórias normais).
Então, vale a seguinte aproximação: para quaisquer a e b (com $a < b$)

$$\boxed{\boldsymbol{P}[a \leq B_{n,p} \leq b] \approx \boldsymbol{P}[a \leq Y \leq b]}$$

assim como seus dois casos particulares:

$$\boxed{\boldsymbol{P}[B_{n,p} \leq b] \approx \boldsymbol{P}[Y \leq b]} \quad \text{e} \quad \boxed{\boldsymbol{P}[B_{n,p} \geq a] \approx \boldsymbol{P}[Y \geq a]}$$

A aproximação de Binomial por Normal

Esclarecimento: Com as palavras "*...em sua formulação prática e sem detalhes*" eu quiz dizer que na formulação do Teorema **??**, não pretendi entrar em detalhes sobre todas as condições que garantem a validade do teorema, nem na discussão acerca de qual boa é a aproximação alegada pelo teorema.

A aproximação de Binomial por Normal

Eis abaixo a formulação do Teorema **??** que seria dada pela variável aleatória Binomial, se ela soubesse falar e quizesse lhe explicar como ela possa ser aproximada com o uso do teorema: "Pegue aquela variável aleatória normal cuja média coincide com a minha média e cuja variância coincide com a minha variância, e ponha ela no meu lugar nas fórmulas que expressam minhas probabilidades; assim você terá uma aproximação dessas."

Para veremos que isso é correto, basta só lembrar que

$$\begin{aligned} B_{n,p} \sim Bin(n;p) &\implies \boldsymbol{E}[B_{n,p}] = np \\ B_{n,p} \sim Bin(n;p) &\implies \mathrm{Var}[B_{n,p}] = np(1-p) \end{aligned}$$

e notar que "média" se usa neste texto como o nome alternativo para a "esperança matemática".

A aproximação de Binomial por Normal; exemplos de aplicação

Exemplo 1. Suponha $B_{500;0,3} \sim Bin(500; 0,3)$. Aplique Teorema **??** para calcular aproximadamente o valor de

$$\boldsymbol{P}\left[130 \leq B_{500;0,3} \leq 155\right]$$

Solução: Para a aplicação do teorema, precisamos calcular a esperança matemática e a variância da variável aleatória B. Eis estas:

$$\begin{array}{l} \boldsymbol{E}[B_{n,p}] = np \implies \boldsymbol{E}[B_{500;0,3}] = 500 \times 0,3 = 150 \\ \mathrm{Var}[B_{n,p}] = np(1-p) \implies \mathrm{Var}[B_{500;0,3}] = 150 \times 0,3(1-0,3) = 105 \end{array}$$

Então, a probabilidade em interesse é, aproximadamente,

$$\begin{aligned} \boldsymbol{P}\left[130 \leq Y \leq 155\right] &= \boldsymbol{P}\left[\frac{130-150}{\sqrt{105}} \leq Z \leq \frac{155-150}{\sqrt{105}}\right] \\ &= \boldsymbol{P}\left[-1,95 \leq Z \leq 0,48\right] = 0,6588 \end{aligned}$$

A aproximação de Binomial por Normal; exemplos de aplicação

Exemplo 2. Se p for $0,4$, qual deve ser o valor de n para que ao tomar $B_{n;0,4} \sim Bin(n;0,4)$ teremos que $\boldsymbol{P}[B_{n;0,4} \leq 300] = 0,985$?
Solução:

$$0,985 = \boldsymbol{P}[B_{n;0,4} \leq 300] \approx \boldsymbol{P}[Y \leq 300] = \boldsymbol{P}\left[Z \leq \frac{300 - 0,4n}{\sqrt{0,4(1-0,4)n}}\right]$$

Como o limiar para 0,985 é 2,17, então n dá-se pela solução da equação

$$\frac{300 - 0,4n}{\sqrt{0,4(1-0,4)n}} = 2,17$$

Elevando os dois lados dessa ao quadrado e resolvendo a eq. quadrática, acha-se duas raízes: 680,68 e 826,375. A primeira não serve pois $300 - 0,4 \times 680,68$ é negativo. Arredondando a segunda, obtem-se a solução: 827 (não toco no assunto do arredondamento e no por quê esse deve ser para mais).

A aproximação de Binomial por Normal; exercícios

Exc. 1. Use a aproximação da binomial pela normal para estimar as seguintes probabilidades:

(a) $\boldsymbol{P}\big[20 \leq B \leq 30\big]$, se $B \sim \mathrm{Bin}(80; 0,3)$;

(b) $\boldsymbol{P}\big[B \leq 40\big]$, se $B \sim \mathrm{Bin}(100; 0,4)$;

(c) $\boldsymbol{P}\big[50 \leq B \leq 60\big]$, se $B \sim \mathrm{Bin}(80; 0,7)$;

(d) $\boldsymbol{P}\big[B \geq 50\big]$, se $B \sim \mathrm{Bin}(300; 0,25)$.

A aproximação de Binomial por Normal; exercícios

Nos Exercícios **??**, **??**, **??** a seguir, você precisa interpretar a pergunta da parte textual de tal forma que a resposta seja expressa em termos da probabilidade envolvendo uma variável aleatória binomial; depois, você calculará o valor dessa probabilidade usando o teorema sobre a aproximação de binomial por normal. A parte de aproximação não deve causar-lhe problemas; a parte difícil é a de interpretação.

A aproximação de Binomial por Normal; exercícios

Exc. 2. Sabe-se que 25% das crianças expostas a um particular agente infeccioso adquirem uma certa doença. Considere um grupo de 60 crianças.
Use a aproximação da binomial pela normal para calcular, aproximadamente, a probabilidade de que no mínimo 15 e no máximo 20 crianças do grupo adoençam.

A aproximação de Binomial por Normal; Exercícios

Ajuda ao Exc. ??. A frase *"sabe-se que 25% das crianças expostas a um particular agente..."* significa que trata-se de uma população de crianças. Entretatnto, o enunciado não especifica se essa população é só um grupo específico que ficou exposto a um agente infeccioso, ou que essa é a população de todas as crianças do mundo que potencialmente podem ser expostas ao agente. Também, o enunciado não esclarece se 25% é a proporção das crianças cujo genótipo ou o estado de saúde é tal que ao serem expostas ao agente, ficarão doente, ou se dentro de cada criança há uma "moeda" que dá "cara" com a probabilidade 25%, e que a exposição da criança ao agente infeccioso acarreta o lançamento de sua moeda que manda a criança adoentar caso sair "cara". Para qualquer das alternativas possíveis, o modelo do experimento aleatório enunciado é o mesmo: você representa cada criança da população por uma bola, você tinge de preto 25% das bolas, e coloca todas numa urna; você vai retirar dela, ao acaso e com reposição, 60 bolas, e você está interessado na probabilidade que as retiradas mostrem no mínimo 15 e no máximo 20 bolas pretas. É óbvio que o número de bolas pretas em 60 retiradas com reposição tem a distribuição Binomial$(60; 0,25)$.

A aproximação de Binomial por Normal; Exercícios

Ajuda ao Exc. ??; continuação. Para fechar a solução, resta então só estimar $\boldsymbol{P}[15 \leq B \leq 20]$ com ajuda do Teorema **??**. Mas antes de fechar minha explicação, gostaria de ressaltar que a reposição das bolas retiradas é obrigatória para o modelo, pois sem a reposição a distribuição resultante não seria binomial, e, consequentemente, todo o exercício estaria alheio ao presente tema. Por outro lado, na realidade, a internação das crinaças adoecidas é – naturalmente – sem reposição. Para cociliar os opostos, pressuponha-se que a população real seja muito grande, de modo que a não devolução de 60 crianças não muda significativamente a proproção de bolas pretas na urna. Tal tipo de pressuposto geralmente não menciona-se no enunciado, pois a menção considera-se disnecessária.

A aproximação de Binomial por Normal; Exercícios

Ex. 3. Exatamente 30% da população da cidade apoiava o antigo prefeito que perdeu a última eleição. (Observe que as eleções já passaram e é por isto que sabe-se a proporção exata do eleitorado do antigo prefeito.) Qual é a probabilidade de que, dentre 100 moradores da cidade, escolhidos ao acaso, no mínimo 40 sejam do eleitorado deste candidato? Use a aproximação da binomial pela normal.

Nota: A princípio, um morador, uma vez escolhido, não será escolhido novamente. Isto faz com que o modelo probabilístico da escolha é aquele das retiradas sem reposição. Entretanto, sugiro assumir que a população da cidade é grande do modo que a escolha sem reposição pode ser aproximada pela a com reposição. Sem esse pressuposto e a consequente aproximação, a distribuição binomial não surge na solução do exercício.

Ajuda ao Exc. ??. Cada pessoa é uma bola na urna. As que apoiam o antigo prefeito – 30% em proporção de toda a população – são bolas pretas; as outras – brancas. Se escolhermos 100 bolas da urna, a distribuição do número de bretas dessa amostra terá a distribuição Binomial$(100; 0,3)$.

A aproximação de Binomial por Normal; Exercícios

Ex. 4. O dono da lanchonete do nosso instituto sabe que cada freguês escolhe carne ou frango com respectivas probabilidades 60% e 40%. Ele ofereceu almoço para 100 participantes de um encontro científico. Ele tem carne suficiente para todos, mas a quantidade de frango dá só para 45 pessoas. Qual é a probabildiade de todos os pedidos sejam satisfeitos? Dê uma estimativa desta probabilidade usando a aproximação de binomial pela normal.

A aproximação de Binomial por Normal; Exercícios

Ajuda ao Exc. ??. Ou: todas as pessoas são bolas numa emensa urna, e 40% das bolas preferem frango, e por isso são pintadas de preto, enquanto que 60% preferem carne, e foram pintadas de branco. Dessa urna, a sorte escolha ao acaso 100 bolas e manda-as à lanchonete. A quantidade das "bolas" que pedem carne tem a distribuição Binomial$(100; 0,4)$. O resto resolve-se pela abordagem que você viu na solução do exercício sobre o Papai Noel que levou brinquedos aos filhos de José.

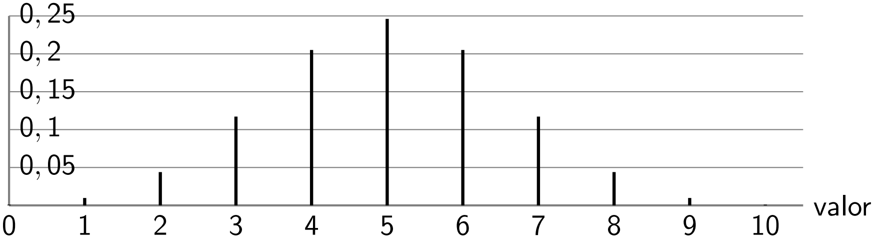
Ou: Cada um dos 100 participanete tem uma moeda que dá "frango" com probabilidade 0,4 e "carne" com 0,6. Amoeda está no estómago ou no bolso – não importa onde exatamente. Ao chegar ao balcão da lanchonete, a pessoa lança sua moeda (ou a sorte lança a moeda no estómago), e dependendo do resultado do lançamento pede frango ou carne. Naturalmente, a quantidade de pedidos de franco terá a distribuição Binomial$(100; 0,4)$.

A aproximação de Binomial por Normal; ilustração

Gostaria de ilustrar a aproximação supraformulada, mas tenho a dificuldade que reside na incompatibilidade entre as propriedades apresentadas abaixo e marcadas por • e ∘:

• as probabilidades envolvendo variável aleatória Binomial ilustram-se pela sua FUNÇÃO DE PROBABILIDADE, que é uma "coleção de palitos", como a no exemplo abaixo

FUNÇÃO DE PROBABILIDADE de $B \sim Bin(10;\, 0,5)$



0.0009765625	0.0097656250	0.0439453125	0.1171875000
0.2050781250	0.2460937500	0.2050781250	0.1171875000
0.0439453125	0.0097656250	0.0009765625	(valores)

A aproximação de Binomial por Normal; Ilustração

○ por outro lado, as probabilidades envolvendo variável aleatória Normal ilustram-se por áreas debaixo de sua FUNÇÃO DE DENSIDADE, como a do exemplo abaixo

FUNÇÃO DE DENSIDADE de $Y \sim \mathcal{N}(5; 2,5)$

A aproximação de Binomial por Normal; Ilustração

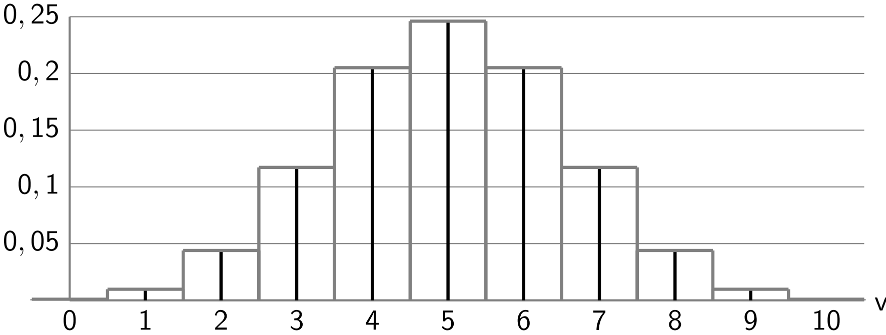
Então, incompatibilidade entre $\bullet$ e $\circ$ supramencionado é que $\boldsymbol{P}[a \leq B_{n,p} \leq b]$ interpreta-se como o comprimento dos palitos da função de probabilidade de $B_{n,p}$ cuja abcissas estão entre a e b, enquanto que $\boldsymbol{P}[a \leq Y \leq b]$ interpreta-se como a área da figura apoiada no intervalo $[a, b]$ e com teto feito da função de densidade de Y. São óbvias as dificuldades de comparação visual entre um comprimento e uma área.

Para resolver o conflito $\bullet \leftrightarrow \circ$, eu vou reapresentar a distribuição da variável aleatória Binomial da maneira que permite expressar suas probabilidade em termos de áreas e não em termos de comprimentos.

Veja a construção na transparência seguinte.

A aproximação de Binomial por Normal; Ilustração

Função de probabilidade em preto;
sua apresentação alternativa em cinza



Devido à construção (as alturas dos prédios são iguais às alturas dos palitos da função de probabilidade) e devido à propriedade particular da distribuição binomial (seus valores são inteiros entre 0 e n) temos que a soma dos palitos com abcissas entre a e b é quase (veja o detalhamento de "quase" abaixo) a área da figura apoíada em $[a, b]$ e com telhado dado pela curva cinza.

A aproximação de Binomial por Normal; Ilustração

Para ver o "quase", escolhe $a = 3$ e $b = 8$. Então, $\boldsymbol{P}[3 \leq B_{10;0,5} \leq 8]$, que é a soma dos comprimentos dos palitos entre 3 e 8 é igual à área da figura delimitada por $3 - \frac{1}{2}$ à esquerda e $8 + \frac{1}{2}$ à direita. Entretanto, no que segue-se, vamos apresentar $\boldsymbol{P}[3 \leq B_{10;0,5} \leq 8]$ pela outra área, que é a área delimitada por 3 à esquerda e 8 à direita. Por isso que a apresentação será "quase" precisa. O erro, porém, é pequeno, se n for grande (que sempre é o caso na prática). Por exemplo, no caso do exemplo (no qual n ainda não é tal grande como seria na prática), é

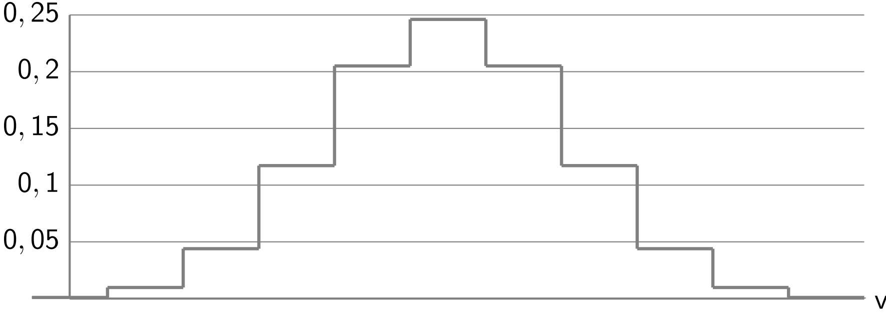
$$\frac{1}{2} \times \left(\boldsymbol{P}[B_{10;0,5} = 3] + \boldsymbol{P}[B_{10;0,5} = 8]\right)$$

Dá para deduzir dessa fórmula (após ela ser generalizada) que o erro do "quase" é, de fato, pequeno e ainda que seu valor está diminuindo se n crescer enquanto a e b ficam intactos.

A aproximação de Binomial por Normal; Ilustração

Então, construimos e justificamos a apresentação da distribuição Binomial por uma função de densidade, no sentido de que a probabilidade da variável aleatória Binomial assumir valores entre *a* e *b* é a área da figura debaixo da função de densidade delimitada dos lados por *a* e *b*. Tal função de densidade está formada por tetos do histograma feito na transparência anterior:

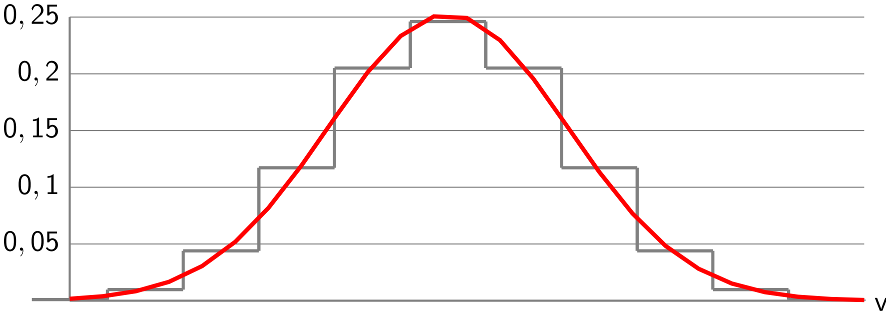
"função de densidade" da $Bin(10; 0.5)$



A aproximação de Binomial por Normal; Ilustração

Acontece que a função de densidade da distribuição $\mathcal{N}(5; 0.25)$ passa por perto da função de densidade da distribuição Binomial $Bin(10; 0.5)$, como a figura abaixo mostra. Tal proximidade faz com que a área debaixo de uma delas delimitada por a e b, seja quase igual à área debaixo da outra, delimitada por mesmos a e b.

"função de densidade" e função de densidade



A aproximação de Binomial por Normal; Ilustração

Como uma das áreas corresponde à (recorde, a variável aleatória $B_{10;0,5}$ abaixo é $Bin(10; 0.5)$)

$$\boldsymbol{P}\left[a \leq B_{10;0,5} \leq b\right]$$

e como a outra área corresponde à (recorde, a variável aleatória Y abaixo é $\mathcal{N}(5; 0.25)$)

$$\boldsymbol{P}\left[a \leq Y \leq b\right]$$

então, a afirmação do Teorema **??** que alega que

$$\boxed{\boldsymbol{P}\left[a \leq B \leq b\right] \approx \boldsymbol{P}\left[a \leq Y \leq b\right]}$$

foi ilustrada por desenhos.

Observe que essa foi uma ILUSTRAÇÃO mas não uma DEMONSTRAÇÃO. Em outras palavras, mostramos O QUE acontece como a razão da proximidade das probabilidades mas não mostramos POR QUE isso acontece.

TCL; formulação

Para mostrar o POR QUÊ, preciso formular o Teorema Central de Limite (TCL).

Tome uma sequencia arbitrária de variáveis aleatórias independentes e identicamente distribuidas. Chame estas por $X_1, X_2, X_3, \dots$ (a sequencia é infinita).

Chame por μ a esperança matemática de cada uma (todas elas têm a mesma) e chame por σ^2 a variância de cada uma (todas elas têm a mesma).

Para cada n natural, tome as n primeiras variáveis aleatórias sa série

$$\underbrace{X_1, X_2, \dots, X_n}_{n}, X_{n+1}, X_{n+2}, \dots$$

e monte a nova variável aleatória a ser chamada por $\mathcal{CP}_n$ (a notação estranha "$\mathcal{CP}$" terá sua justificativa)

$$\mathcal{CP}_n = \frac{X_1 + X_2 + \cdots + X_n \; - \; n\mu}{\sigma\sqrt{n}}$$

TCL, formulação

O Teorema Central de Limite

Para quaisquer a e b fixos antemão, a probabilidade

$$\boldsymbol{P}\left[a \leq \mathcal{CP}_n \leq b\right]$$

converge, conforme n cresce, à probabilidade

$$\boldsymbol{P}\left[a \leq Z \leq b\right], \text{ onde } Z \text{ é a variável aleatória normal padrão.}$$

A convergência significa que o valor da diferença entre as duas tende ao 0 conforme n crescer ao ∞. A velocidade do "desaparecemento" da diferença depende de
1) valores de a e de b;
2) a distribuição do "ingrediente" da variável aleatória $\mathcal{CP}_n$ (quer dizer da distribuição de X_i)
e esses não serão discutidos por mim aqui.

TCL, formulação

Observe que uma vez que μ é a esperança de cada ingrediente (quer dizer, de cada X_i), e que σ é o desvio padrão de cada ingrediente, então $n\mu$ é a esperança da soma $X_1 + \ldots + X_n$ e $\sqrt{n}\sigma$ é o desvio padrão dela. Portanto, a variável aleatória

$$\frac{X_1 + X_2 + \cdots + X_n \; - \; n\mu}{\sigma\sqrt{n}}$$

tem esperança nula e desvio padrão 1.

Consequentemente, cada $\mathcal{CP}_n$ tem esperança nula e desvio padrão 1. Tal variável aleatória chama-se $\mathcal{C}$entralizada (pois sua esperança é nula) e $\mathcal{P}$adronizada (pois sua variância é 1).

Portanto, o Teorema Central de Limite fala sobre a convergência de sequencia de variáveis aleatórias centralizadas e padronizadas.

TCL, formulação

Entretanto, para a convergência da sequencia

$$\mathcal{CP}_1, \mathcal{CP}_2, \ldots, \mathcal{CP}_n, \ldots \text{ a } Z \sim \mathcal{N}(0, 1^2)$$

não é suficiente que cada $\mathcal{CP}_n$ seja $\mathcal{C}$entralizada e $\mathcal{P}$adronizada.

O fenómeno (de convergência) afirmado em TCL ocorre devido à estrutura particular da variáveis na sequencia: todas elas são "feitas" por somas parciais de uma mesma sequencia de ingredientes:

$$\begin{array}{lll}
\mathcal{CP}_1 & \text{adveio de} & X_1 \\
\mathcal{CP}_2 & \text{adveio de} & X_1 + X_2 \\
\mathcal{CP}_3 & \text{adveio de} & X_1 + X_2 + X_3 \\
 & \ldots & \\
\mathcal{CP}_n & \text{adveio de} & X_1 + X_2 + \ldots + X_n \\
\vdots & \vdots & \vdots
\end{array}$$

onde $X_1, X_2, \ldots, X_n \ldots$ são variáveis aleatórias independetes e identicamente distribuidas.

Um comentário sobre o nome “Teorema Central de Limite”.

É errado pensar que o nome “Teorema Central de Limite” foi dado ao teorema acima formulado devido ao fato que o mesmo trata de convergência de variáveis aleatórias centradas. Se fosse isso, o nome seria “Teorema de Limite Central ou Centrado”. O adjetivo “central” usado no nome reflete a posição central do teorema em toda a área de Probabilidade e Estatística. Se fosse para modificar o nome com o intuíto de esclarecer ao máximo o caráter do teorema, esse seria “Teorema sobre Limite que ocupa lugar Central na Teoria de Probabilide e Estatística”.

Do TCL para o Teorema sobre a aproximação da binomial pela normal.

Para a dedução da aproximação da binomial pela normal a partir da TCL, vamos tomar no lugar da sequencia de ingredientes

$$X_1, X_2, X_3, \ldots$$

uma sequencia de variáveis aleatórias Bernoulli com o mesmo valor do parâmetro p; quer dizer, na nossa dedução agora

$$\text{cada } X_i \sim \text{Bernoulli}(p) \text{ e todas são independentes}$$

Alias, é fácil conceber essa sequencia: imagine que alguém lança sem parar uma mesma moeda, e denote por p a probabilidade da moeda dar "cara" em qualquer lançamento (a moeda não precisa ser honesta); imagine que você observa a série de lançamentos e, para cada $i = 1, 2, 3, \ldots$, sua i-ésima observação está codificadas por X_i da seguinte maneira: X_i assume 1 se i-ésimo lançamento der "cara", e assume 0 se der "coroa".

Do TCL para o Teorema sobre a aproximação da binomial pela normal.

Para a sequencia epsecifica assim escolhida tem-se:

$$\boldsymbol{E}[X_i] = p, \quad \mathrm{Var}[X_i] = p(1-p)$$

e ao definir $B_{n,p}$ como

$$B_{n,p} = X_1 + X_2 + \ldots + X_n$$

tem-se que

$$B_{n,p} \sim Bin(n;p)$$

Então, a variável aleatória $\mathcal{CP}_n$ tratada no TCL adquire, no caso, a seguinte expressão:

$$\mathcal{CP}_n = \frac{B_{n,p} - np}{\sqrt{np(1-p)}}$$

Do TCL para o Teorema sobre a aproximação da binomial pela normal.

Então, o TCL, sendo formulada para nosso caso, afirma que para quaisquer a e b fixos antemão,

$$\boldsymbol{P}\left[a \leq \frac{B_{n,p} \ - \ np}{\sqrt{np(1-p)}} \leq b\right]$$

converge, conforme n cresce, à probabilidade

$\boldsymbol{P}\left[a \leq Z \leq b\right]$, onde Z é a variável aleatória normal padrão.

A convergência afirmada acarreta no que, para n fixo mas grande,

$$\boldsymbol{P}\left[a \leq \frac{B_{n,p} \ - \ np}{\sqrt{np(1-p)}} \leq b\right] \approx \boldsymbol{P}\left[a \leq Z \leq b\right]$$

Do TCL para o Teorema sobre a aproximação da binomial pela normal.

Então, ao fixar um n grande, podemos dizer que as variáveis aleatórias

$$\frac{B_{n,p} \ - \ np}{\sqrt{p(1-p)}\sqrt{n}} \text{ e } Z$$

são "idênticas".

Multiplicaremos ambas por $\sqrt{np(1-p)}$, e depois acrescentamos às duas np. A "identidade" está preservada. Logo, as variáveis aleatórias

$$B_{n,p} \text{ e } \sqrt{np(1-p)}\, Z \ + \ np$$

são "idênticas".

A primeira delas tem a distribuição $Bin(n,p)$, e a segunda $\mathcal{N}\left(np,\ np(1-p)\right)$ (devido às propriedades de Normais listadas e explicadas adiante). A "identidade" entre as duas é exatamente a afirmação do Teorema sobre a aproximação de Binomial por Normal.

Propriedades das variáveis normais; teoria

Na dedução do Teorema sobre a aproximação de Binomial por Normal a partir do TCL usamos a propriedade que diz que se

$$Z \text{ tenha a distribuição Normal Padrão}$$

então

$$\sqrt{np(1-p)}\, Z + np$$

tem a distribuição Normal com média np e a variância $np(1-p)$.

tal propriedade é uma caso particular de uma série de propriedades que estão apresentadas abaixo.

Propriedades das variáveis normais; teoria

Sabemos que:
(a) se $Y = X + b$ então $\boldsymbol{E}[Y] = \boldsymbol{E}[X] + b$ e $\mathrm{Var}[Y] = \mathrm{Var}[X]$;
(b) se $Y = cX$ então $\boldsymbol{E}[Y] = c\boldsymbol{E}[X]$ e $\mathrm{Var}[Y] = c^2\mathrm{Var}[X]$.

As variáveis aleatórias <u>normais</u> obedecem estas propriedades, mas há muito mais:
(a) se $X \sim \mathcal{N}(\mu, \sigma^2)$ e se $Y = X + b$ então $Y \sim \mathcal{N}(\mu + b, \sigma^2)$ (observe que isto implica, em particular, que $\boldsymbol{E}[Y] = \boldsymbol{E}[X] + b$ e $\mathrm{Var}[Y] = \mathrm{Var}[X]$);
(b) se $X \sim \mathcal{N}(\mu, \sigma^2)$ e se $Y = cX$ então $Y \sim \mathcal{N}(c\mu, c^2\sigma^2)$ (observe que isto implica, em particular, que $\boldsymbol{E}[Y] = c\boldsymbol{E}[X]$ e $\mathrm{Var}[Y] = c^2\mathrm{Var}[X]$);
(c) se $X \sim \mathcal{N}(\mu, \sigma^2)$, se $V \sim \mathcal{N}(m, s^2)$, se X e V são variáveis aleatórias independentes e se $Y = X + V$ então $Y \sim \mathcal{N}(\mu + m, \sigma^2 + s^2)$ (em palavras: a soam de duas variáveis aleatórias normais independentes é uma variável aleatória normal cuja média é a soma das médias das variáveis aleatórias do soma e cuja variância é a soma das variâncias das variáveis aleatórias da soma).

Propriedades das variáveis normais; teoria

Observe que a propriedade (c) não vale para a maioria das outras distribuições estudadas até o momento (por exemplo, soma de duas uniformes independentes não é uma uniforme, soma de duas binomiais independentes com difernetes valores do parâmetro p não é binomial).

Observe uma propriedade particular decorrente das (a)-(c): se $X \sim \mathcal{N}(\mu, \sigma^2)$, então $(-X) \sim \mathcal{N}(-\mu, \sigma^2)$. Esta permite não só somar variáveis aleatórias normais, mas também subtrair, por exemplo:
se $X \sim \mathcal{N}(\mu, \sigma^2)$, e se $Y \sim \mathcal{N}(\mu, \sigma^2)$, e X e Y são independentes então

$$X - Y \text{ tem a distribuição } \mathcal{N}(0, 2\sigma^2).$$

Vale ainda notar que

$$X - X \text{ é identicamente zero}$$

Por favor, verifique que você entende a diferneça entre as duas.

Propriedades das variáveis normais; exercícios

Ex. 5(a) A variável aleatória X tem distribuição $\mathcal{N}(2; (3)^2)$. Qual é a distribuição de $\frac{1}{3}X$?
(b) As variáveis aleatórias X_0, X_1, X_2, X_3, X_4 são independentes e cada uma tem a distribuição $\mathcal{N}(1; (3)^2)$. Qual é a distribuição da variável aleatória $4X_0 - (X_1 + X_2 + X_3 + X_4)$?

Ex. 6(a) A variável aleatória X tem distribuição $\mathcal{N}(8; (2)^2)$. Qual é a distribuição de $\frac{1}{2}X$?
(b) As variáveis aleatórias X_0, X_1, X_2, X_3 são independentes e cada uma tem a distribuição $\mathcal{N}(1; (2)^2)$. Qual é a distribuição da variável aleatória $(X_1 + X_2 + X_3) - 3X_0$?

Ex. 7(a) A variável aleatória X tem distribuição $\mathcal{N}(1; (2)^2)$. Qual é a distribuição de $3X$?
(b) As variáveis aleatórias X_0, X_1, X_2, X_3, X_4 são variáveis aleatórias independentes e cada uma tem a distribuição $\mathcal{N}(3; (4)^2)$. Qual é a distribuição de $4X_0 - (X_1 + X_2 + X_3 + X_4)$?

Propriedades das variáveis normais; exercícios

Ex. 8(a) A variável aleatória X tem distribuição $\mathcal{N}(3; (4)^2)$. Qual é a distribuição de $\frac{1}{3}X$?

(b) As variáveis aleatórias X_0, X_1, X_2 são independentes e cada uma tem a distribuição $\mathcal{N}(3; (5)^2)$. Qual é a distribuição da variável aleatória $(X_1 + X_2) - 2X_0$?

TCL e a vida real

Vamos analisar o caso particular do TCL: a aproximação de Binomial por Normal.
Recorde: escolhemos n grande e somamos n variáveis aleatórias Bernoulli(p) independentes entre si:

$$B_{n,p} = X_1 + X_2 + \ldots + X_n$$

O TCL afirma que

$$\mathcal{CP}_n = \frac{B_{n,p} - np}{\sqrt{np(1-p)}}$$

tem distribuição (aproximadamente) Normal Padrão.

Re-escrevo:

$$\mathcal{CP}_n = \frac{X_1 - p}{\sqrt{np(1-p)}} + \frac{X_2 - p}{\sqrt{np(1-p)}} + \cdots + \frac{X_n - p}{\sqrt{np(1-p)}}$$

e parto para a análise de cada fração dessa somatória.

TCL e a vida real

Recordo: cada X_i tem a distribuição

x	0	1
$P[X_i = x]$	$1-p$	p

Portanto, cada $\frac{X_i - p}{\sqrt{np(1-p)}}$ tem

x	$-\frac{p}{\sqrt{np(1-p)}}$	$\frac{1-p}{\sqrt{np(1-p)}}$
$P\left[\frac{X_i - p}{\sqrt{np(1-p)}} = x\right]$	$1-p$	p

Observe: se n for muuuito grande então,

$$\text{tanto } -\frac{p}{\sqrt{np(1-p)}} \text{ quanto } \frac{1-p}{\sqrt{np(1-p)}}$$

são muuuito pequenos. Portanto, podemos concluir que cada $\frac{X_i - p}{\sqrt{np(1-p)}}$ contribui pouquissimo à somatória de todas. Isso nos leva à seguinte interpretação do TCL:

TCL e a vida real

Interpretação do TCL: Se somarmos grande número de variáveis aleatórias independentes e identicamente distribuidas, sendo que cada uma delas tráz contribuição infima ao valor da soma, então a soma terá distrbuição aproximadamente Normal Padrão.

Vale lembrar que no caso analisado, $\boldsymbol{E}[X_i] = p$ e $\mathrm{Var}[X_i] = p(1-p)$ (pois $X_i \sim \mathit{Bernoulli}(p)$). Portanto,

$$\boldsymbol{E}\left[\frac{X_i \; - \; p}{\sqrt{np(1-p)}}\right] = 0, \text{ e } \mathrm{Var}\left[\frac{X_i \; - \; p}{\sqrt{np(1-p)}}\right] = \frac{1}{n}$$

Isso dá uma correção à interpretação feita acima. Agora ela é assim:

TCL na vida real

Interpretação (melhorada) do TCL: Se somarmos n variáveis aleatórias independentes e identicamente distribuidas, cada uma com esperança 0 e com variância $1/n$, então a soma terá distrbuição aproximadamente Normal Padrão.

Observe que a frase "com variância $1/n$" implica, em particular, que a contribuição de cada variável á soma é ínfima (é o que foi dito na primeira Interpretação do TCL), mas além disso, exige que a variância de cada contribuição esteja numa relação com o número de variáveis na soma.

TCL na vida real

Valem ainda duas correções à Interpretação do TCL dada acima; ambas podems ser provadas mas não mo meu curso.

Correção 1 à Interpretação (melhorada) do TCL: Se a Natureza errar um pouco e somar n variáveis independentes e identicamente distribuidas, mas com a esperança diferente de 0 e a variância diferente de $1/n$, então a aproximação da soma dar-se-á não por Normal Padrão, mas por uma distribuição Normal com media possivelmente difernete de 0 e com variância possivelmente diferente de 1.

Correção 2 à Interpretação (melhorada) do TCL: A aproximação por Normal continua valendo se as variáveis estejam com distribuições diferentes (mas não muito) e ainda estejam dependentes (mas não muito).

TCL na vida real

Concluimos:

Interpretação informal do TCL: Se uma atributo numa população forma-se como uma soma de grande quantidade de variáveis aleatórias independentes, mas com pequeno impacto da cada uma no valor sa soma, então a distribuição da frequencia relativa por esse atributo será aproximadamente normal.

É por isto que em diversas situações da vida real, frequencias relativas populacionais têm distribuição aproximadamente normal:

altura de pessoas
peso de pessoas, etc.

Mas não qualquer quantidade teria a distribuição normal; por exemplo, o tempo de vida de uma lâmpada tem distribuição (aproximadamente) exponencial. A razão: o tempo de vida obedece a lei da Física que não enquadra-se no esquema de "'soma de...".

TCL na vida real

Pergunta:
SE A NATUREZA FEZ COM QUE UMA CERTA FREQUENCIA RELATIVA DE UMA CERTA POPULAÇÃO TENHA DISTRIBUIÇÃO APROXIMADAMENTE NORMAL, COMO É QUE DESCOBRIR ISSO?
Resposta:
E precisa fazer o histograma da distribuição da frequencia relativa; se haver a função de densidade de uma Normal que passe por telhados dos prédios do histograma, então pode-se afirmar que haja a aproximação pela Normal no sentido formulado acima.
Os prédios do histograma devem ser enconstados, quer dizer, não deve haver "ruas" entre prédios, e os dados (do conjunto de dados que gerou a frequencia estudada) não deve ter lacunas grandes, isto é, os dados devem ser bem densos (claro, com a densidade diminuindo para as caudas).
Se haver a aproximação por Normal, então, tipicamente, toma-se a Normal com a média igual à média populacional e a variância igual á variância populacional.

Exemplo "eucaliptos" de uso do TCL na vida real

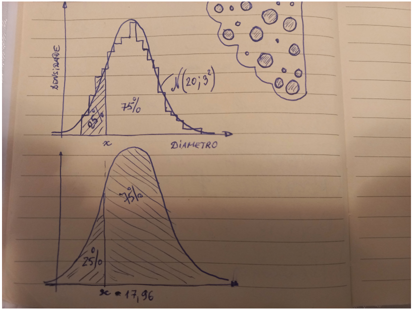
Ex. 9. O diâmetro de um eucalipto com 10 anos de idade, tem distribuição aproximadamente normal com média 20 cm e desvio padrão 3 cm. Uma empresa que plantou eucaliptos há dez anos, tem direito de desmatar 75% da área plantada. Para maximizar o lucro, a empresa pretende cortar as árvores com o maior volume de madeira. Para isso, é preciso achar o valor de x tal que 75% das árvores plantadas têm diâmetro de no mínimo x. Calcule este valor.

Haverá agora uma série de esclarecimentos. O primeiro deles não tem nada a ver com a Estatística mas ele é essencial para que você possa entender o problema central do exemplo e conseguir abordar o problema com ajuda da Estatítsica. O desenho abaixo é que dá o esclarecimento.

Exemplo “eucaliptos” de uso do TCL na vida real

No desenho, eu apresentei um pedaço do lote de terra no qual foram plantados os eucaliptos. Esses estão marcados por círculos (é como se fosse que a plantação foi fotografada do espaço). Os círculos são diferentes; isso acontece pois árvores não são idênticas. Acontece que a empresa que plantou as árvores tem o direito de desmatar 75% da plantação. De acordo com o plano inicial (traçado há dez anos) as árvores cortadas irão agora para produção de celulose. Para otimizar o lucro, a empresa tem o interesse em cortar as árvores mais volumosas, o que entende-se no caso como as árvores com o maior diâmetro (sim, no presente caso fazemos a simplificação: quanto maior o diâmetro, maior é o volume da árvore toda). Então, a empresa deseja achar o valor x tal que as árvores de diâmetro maior que x representem 75% de toda a plantação. Depois, os funcionários da empresa receberão a instrução: ”cortar toda e qualquer árvore cuja diâmetro excede x”, e com isso será alcançada a desejada otimização do volume cortado.

Exemplo “eucaliptos” de uso do TCL na vida real



Exemplo “eucaliptos” de uso do TCL na vida real

O segundo esclarecimento a ser fornecido agora já tem tudo a ver com a Teoria Estatítstica. A idéia central é que possuimos o histograma da distribuição de frequencia relativa pelo atributo "diâmetro" que observamos/medimos para os eucaliptos da plantação. Existem diversos caminhos que nos forneceriam tal histograma. Você pode imeginar que alguém foi e mediu todas as árvores da plantação. Você também pode imaginar que que houve um estudo abragente sobre os eucaliptos de idade de 10 anos e que nossa plantação é uma amostra representativa de toda a população. O importante para nós é a principal propriedade de histograma: os tetos de seus prédios representam a densidade para a frequencia relativa. Isso nos permite visualizar a tarefa de procura pelo limiar x que foi formulada acima: procuramos pelo x que separe o histograma em 25% (à sua esquerda) e 75% (à sua direita).

Exemplo “eucaliptos” de uso do TCL na vida real

A solução da tarefa será feita com o auxílio do seguinte fato: a densidade do histograma pode ser aproximada pela função-densidade da distribuição Normal $\mathcal{N}(20, 3^2)$. Quem e como descobriu os parâmetros 20 e 3^2 não é o assunto que nos interessa no momento. Já aprópria aproximação é uma consequencia do Teorema Central de Limite. O presente exemplo é uma ilustração de emprego do tal teorema.
Então, a proximidade dos telhados do histograma à função de densidade específica permite reformular a tarefa da seguinte maneira: achar x que separe o sino que fica debaixo da função de densidade $\mathcal{N}(20, 3^2)$ em duas partes: a de 25% e a de 75%.

Exemplo “eucaliptos” de uso do TCL na vida real

Isso resolve-se pela conta abaixo, na qual V é uma variável aleatória auxiliar; ela tem a distribuição $\mathcal{N}(20, 3^2)$, mas ela não tem interpretação no âmbito do exemplo (explicitamente falando, na situação tratada no presente exemplo não há nenhuma aleatoriedade e é por isso que a variável aleatória V não possui interpretação alguma nos termos envolvidos na formulação ou na solução do problema do exemplo).

Achar x t.q. $0,75 = \boldsymbol{P}[V \geq x]$ onde $V \sim \mathcal{N}(20, 3^2)$. Eis a solução:

$$0,75 = \boldsymbol{P}\left[Z \geq \frac{x-20}{3}\right], \text{ onde } Z \sim \mathcal{N}(0, 1^2)$$

Da tabela da distribuição de Normal Padrão,

$$\frac{x-20}{3} = -0,68 \implies x = 17,96$$

A resposta é: se cortar todas as árvores cujo diametro seja $\geq 17,96$cm, então será cortado 75% da plantação (e, claro, alcança-se o objetivo de cortar o maior volume possível).

Exemplo "elevador" de uso do TCL na vida real

Ex. 10. A capacidade máxima de um elevador é de 500kg. Se a distribuição X dos pesos dos usuários é suposta $\mathcal{N}(70, 100)$: Qual a probabilidade de 7 passageiros ultrapassarem esse limite?

A pergunta é uma maneira – talvéz não muito boa – de dizer os seguinte:
se 7 pessoas foram retiradas ao acaso da população de usuários do elevador, qual é a probabilidade da soma de seus pesos ser > 500?
Para resolver o problema, precisa assumir que a retirada faz-se com a reposição. Isso garante que se

$$W_1, W_2, W_3, W_4, W_5, W_6, W_7$$

foram as variáveis aleatórias denotando os pesos das 7 pessoas, então elas sejam independentes e identicamente distribuidas.

Exemplo "elevador" de uso do TCL na vida real

Agora, a proximidade do histograma com a distribuição $\mathcal{N}(70, 100)$ permite substituir as W's por

$$V_1, V_2, V_3, V_4, V_5, V_6, V_7$$

onde as V's são independentes e cada uma tem distribuição $\mathcal{N}(70, 100)$. Observe que, diferentemente do que aconteceu no exercício anterior, agora cada V possui interprtação: é (aproximadamente) o peso de pessoa retirada ao acaso da população.

As propriedade de variáveis aleatórias normais citadas acima, garantem que

$$S = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 + V_7 \sim \mathcal{N}(7 \times 70, 7 \times 100)$$

Então o problema reduz-se à conta $\boldsymbol{P}[S \geq 500]$ onde $S \sim \mathcal{N}(7 \times 70, 7 \times 100)$.

Exemplo “elevador” de uso do TCL na vida real

Eis a conta (abaixo, Z denomina a variável aleatória Normal Padrão):

$$\begin{aligned} & \boldsymbol{P}[S \geq 500] \\ = \ & \boldsymbol{P}\left[Z \geq \frac{500-490}{\sqrt{700}}\right] \\ = \ & \boldsymbol{P}[Z \geq 0,3779] \\ \approx \ & \boldsymbol{P}[Z \geq 0,38] = 1 - \boldsymbol{P}[Z \leq 0,38] \\ = \ & 1 - 0,6480 = 0,3520 \end{aligned}$$

Então, a resposta é que $0,3520$ é a probabilidade de 7 pessoas aleatoreamente escolhidas da população $\mathcal{N}(70,100)$ terem seus pesos, em somatória, maior que 500 kg.

Exercícios: TCL na vida real

Ex. 11. Cada seção usada para construção de um oleoduto tem um comprimento médio de 5m e desvio padrão de 20cm. O comprimento total do oleoduto será de 8km.
(a) Se a firma construtora do oleoduto encomendar 1600 seções, qual a probabilidade de terem que comprar mais do que uma seção adicional (isto é, das 1600 seções somarem 7995m ou menos)?
(b) Qual a probabilidade do uso exato de 1599 seções, isto é, a soma das 1599 seções estar entre 8000 e 8005m?

Exercícios: TCL na vida real

Ex. 12. A máquina de empacotar um determinado produto o faz segundo uma distribuição normal, com a média que pode ser regulada, e com o desvio padrão cujo valor é 10g para qualquer que seja o valor da média.
(a) Em quanto deve ser regulado o peso médio para que apenas 10% dos pacotes tenham menos de 500g?
(b) Com a máquina assim regulada, qual é a probabilidade de que o peso de um pacote exceda os 600g?
(c) Suponha que o valor da média foi regulado para 500g. Ache a probabilidade que o peso total de 4 pacotes seja menor que 1980g.

Exercícios: TCL na vida real

Ex. 13. O peso de um ovo tem distribuição aproximadamente normal com média 30 gramas e desvio padrão 2 gramas. Um granjeiro adquiriu direito de vender 60% da produção de ovos para uma rede de supermercados. Para maximizar o lucro desta venda, ele quer vender os ovos mais pesados da sua produção. Para isso, ele precisa achar o valor de y tal que 60% dos ovos têm peso de no mínimo y. Calcule este valor.

Exercícios: TCL na vida real

Ex. 14. Analisando o histograma da altura de 10.000 alunos em uma universidade concluiu-se que a distribuição normal com média 170 cm e desvio-padrão igual a 5 cm é adequada para estudar a estatura probabilisticamente.
(a) Qual o número esperado de alunos com altura superior a 165cm?
(b) Qual o intervalo simétrico em torno da média que conterá 75% das alturas dos alunos?

Comentário:Esse exercício veio, sendo copiado, de um dos texto que ensinam a Estatística Básica. <u>Não fui eu</u> quem escreveu "Qual o número esperado de alunos com altura superior a 165cm?" Eu não sei o sentido de "o número esperado de alunos". Por mim, no lugar desse termo deveria ter aparecido "a proporção de alunos".

Exercícios: TCL na vida real

Ex. 15. O número de acidentados que chega por dia em certo hospital tem distribuição praticamente normal de média 75 e desvio padrão 8. Qual é a probabilidade de que, em um dia qualquer, cheguem
(a) pelo menos 75 acidentados?
(b) mais de 60 e menos de 80 acidentados?

Exercícios: TCL na vida real

Ex. 16. Uma clínica de emagrecimento recebe pacientes com peso seguindo uma distribuição normal de média 130 kg e desvio padrão 20 kg. Para efeito de determinar o tratamento mais adequado, os 25% dos pacientes de menor peso são classificados como "mais leves", enquanto que os 25% de maior peso, como "mais pesados". Determine os valores que delimitam cada uma dessas classificações.

Comentário: Esse é um exemplo do tipo de texto que surge quando uma faculdade – a de Fisioterapia, no caso, – pede de mim fazer exercícios cujo enredo seja mais próximo à realidade da profissão futura de seus alunos.

Exercícios: TCL na vida real

Ex. 17. O número de pedidos para compra de certo produto que uma companhia recebe por semana distribui-se normalmente, com média 125 pedidos e desvio padrão 30.
(a) Se em uma semana o estoque disponível é de 150 unidades, qual é a probabilidade de que todos os pedidos sejam atendidos?
(b) Qual deveria ser o estoque para que se tivesse 98% de probabilidade de que todos os pedidos fossem atendidos?
(c) Construa um intervalo centrado em torno da média que contenha 80% dos pedidos.

Exercícios: TCL na vida real

Ex. 18. A distribuição dos pesos de coelhos criados numa granja pode muito bem ser representada por uma distribuição Normal, com média 5 kg e desvio padrão 0,9 kg. Um abatedouro comprará 5000 coelhos e pretende classificá-los de acordo com o peso do seguinte modo: 15% dos mais leves como pequenos, os 50% seguintes como médios, os 20% seguintes como grandes e os 15% mais pesados como extras. Quais são os limites de peso para cada classificação?

Comentário:Os defensores de animais e os vegitarianos não precisam fazer esse exercício.

Exercícios: TCL na vida real

Ex. 19. Um teste de aptidão feito por pilotos de aeronaves em treinamento inicial requer que uma série de operações seja realizada em uma rápida sucessão. Suponha que o tempo necessário para completar o teste seja distribuído de acordo com uma Normal de média 90 minutos e desvio padrão 20 minutos. Para passar, o candidato deve completar o teste em menos de 80 minutos.

Exercícios: TCL na vida real

(a) Se 10 candidatos fazem o teste, qual a probabilidade de que pelo menos 3 sejam aprovados? *Dica: É a probabilidade de obter valor não maior que* 3 *na Distribuição Binomial com* $n = 10$ *e* p *igual a probabilidade de qualquer candidado ser aprovado.*

(b) Se 65 candidatos fazem o teste, quantos candidatos espera -se que passem? *Dica: No item (a) deste exercício, você já viu que se* K *candidados fazem o teste, então o número de aprovados segue a Distribuição Binomial cujo parámento* n *é igual a* K*,e cujo parâmetro* p *é igual à probabilidade de qualquer candidado passar no teste. Portanto, no presente item, surge a Distribuição Binomial com* $n = 65$ *e com valor de* p *que você deve ter calculado. Esta distribuição possui esperança matemática. À ela é que se refere no enunciado por "número esperado de candidados que passem". Posso reformular minha explicação assim: denote pela variável aleatória* X *o número de candidados que passem dentros dos 65 que façam o teste. Calcule então* $\boldsymbol{E}[X]$*, tomando em conta que, segundo ao enunciado,* $X \sim Bin(65, p)$*.*

(c) Qual é o valor do tempo máximo para se completar a prova de maneira que apenas 5% dos candidatos completam o teste num tempo inferior a esse valor?

Ex. 20. Uma enchedora automática de refrigerantes (era vodka, mas me obrigaram a substituir por refrigerante) está regulada para que o volume médio de líquido em cada garrafa seja de 1000 cm^3 e desvio padrão de 15 cm^3. Admita que o volume siga uma distribuição normal.
(a) Qual é a porcentagem de garrafas em que o volume de líquido é menor que 990 cm^3? *Observação: A potência* 3 *refere-se ao "cm" pois cm^3 é a unidade da medida do volume aquí usada; digo isso para evitar que você suspeite que 1000 e 15 devem ser elevados à potência* 3. *Desculpe por suspeitar que você possa ter tais intenções; é que já vi muitos alunos fazendo isso e resolvite avisar antecipadamente.*
(b) Qual é a porcentagem de garrafas em que o volume de líquido não se desvia da média em mais do que dois desvios padrões?

(c) Se 10 garrafas são selecionadas ao acaso, qual é a probabilidade de que, no máximo, 4 tenham volume de líquido superior a 1003 cm^3?

Ex. 21. Estudos meteorológicos indicam que a precipitação pluviométrica mensal em períodos de seca numa certa região pode ser considerada como seguindo a distribuição Normal de média 30 mm e variância 16 mm^2. *Observação: A potência 2 refere-se à medida de mensuração, isto é, a "mm", e não ao valor númerico. Mais epsecificamente: o valor numércio é 16 mesmo, e não* 16^2. *Note que neste exercício, diferentemente dos exercícios anteriores, é fornecido o valor da variância e não o do desvio padrão. Devido a isso, é agora o momento certo de te avisar sobre que a dimensão da variânica é o quadrado da dimensão de mensuração da quantidade cuja distribuição é dita aproximadamente normal. No presente caso, a quantidade é a precipitação pluviométrica. Esta está medida em milímetros. Logo a variância de sua distribuição tem dimensão* mm^2.

Exercícios: TCL na vida real

(a) Qual é a probabilidade de que a precipitação pluviométrica mensal no período da seca esteja entre 24 e 39 mm?
(b) Qual seria o valor da precipitação pluviométrica de modo que exista apenas 10% de chance de haver uma precipitação inferior a esse valor?
(c) Construa um intervalo central em torno da média que contenha 80% dos possíveis valores de precipitação pluviométrica.

Ex. 22. Suponha que X, o diâmetro interno (em milímetros) de um bocal, seja uma variável aleatória normalmente distribuída com média 13 e variância 1. Se X não atender a determinadas especificações, o fabricante sofrerá prejuízo. Especificamente, suponha que o lucro L (por bocal) seja a seguinte função de X:

$$L = \begin{cases} R\$20, & \text{se } 12 \leq X \leq 14 \\ R\$-3, & \text{se } X < 12 \\ R\$-2, & \text{se } X > 14 \end{cases}$$

Qual é o lucro esperado por bocal?

Exercícios: TCL na vida real

Observação: Para que possa resolver este exercício, você precisa entender que "o lucro esperado por bocal" corresponde à $\boldsymbol{E}[L]$*, onde L é como definido pelo enunciado. Segundo à definição,*

$$\boldsymbol{E}[L] = 20\boldsymbol{P}[12 \leq X \leq 14] - 3\boldsymbol{P}[X < 12] - 2\boldsymbol{P}[X > 14],$$

o que reduz a solução ao cálculo de probabilidades envolvendo uma variávela aleatória normal, a qual é X*, no caso. Obrigar você a fazer este ca ´lculo é o objetivo mirado pelo autor do texto. Creio que o cálculo seja fácil para você. Entretanto, o exercício é um tradicional causador de estranhesa. A barreira na procura de sua solução só pode estar, portanto, na interpretação do termo "lucro esperado por bocal". Qual foi o argumento que você inventou para se convencer da correção de minha sugestão que a interpretação desse é* $\boldsymbol{E}[L]$*?*

Exercícios: TCL na vida real

Ex. 23. As notas de Estatística dos alunos de uma determinada universidade distribuem-se normalmente, com média 6,4 e desvio padrão 0,8. O professor atribui graus A, B e C da seguinte forma:

Nota	Grau
$X \leq 5$	C
$5 < X \leq 7,5$	B
$7,5 < X \leq 10$	A

Em uma classe de 80 alunos, qual o número esperado de alunos com grau A? E com grau B? E com grau C?
Observação: Se você necessitar de ajuda para interpretar o termo "número esperado", veja a dica do item (b) do Ex. **??**.

Exercícios: TCL na vida real

Ex. 24. O diâmetro X de rolamentos esféricos produzidos por uma fábrica tem distribuição $\mathcal{N}(0,6140;(0,0025)^2)$. O lucro T de cada rolamento depende de seu diâmetro. Assim,
$T = 0,10$, se o rolamento for bom, isto é, se $0,610 \leq X \leq 0,618$;
$T = 0,05$, se o rolamento for recuperável, isto é, se $0,608 \leq X < 0,610$ ou se $0,618 < X \leq 0,620$;
$T = -0,10$, se o rolamento for defeituoso, isto é, se $X < 0,608$ ou se $X > 0,620$.
Calcule:
(a) As probabilidades de que os rolamentos sejam bons, recuperáveis e defeituosos.
(b) Calcule o lucro médio por rolamento fabricado.
Observação: A idéia da solução é a mesma que a do Ex. **??**.

Exercícios: TCL na vida real

Ex. 25. A durabilidade de um tipo de pneu da marca Rodabem é descrita por uma variável aleatória Normal de média 60.000 km e desvio padrão de 8.300 km.

(a) Se a Rodabem garante os pneus pelos primeiros 48.000 km, qual a proporção de pneus que deverão ser trocados pela garantia? *Observação: Espero que seja claro que a probabilidade de um pneu, escolhido ao acaso de uma lote, durar menos que 48 mil km é a mesma que a proporção dos pneus da lote que durariam menos que 48 mil km. Este vínculo entre probabilidade e proporção existe devido à própria definição do conceito de probabilidade.*

(b) O que aconteceria com a proporção do item (a), se a garantia fosse para os primeiros 45.000 km?

(c) Qual deveria ser a garantia (em km) de tal forma a assegurar que o fabricante trocaria sob garantia no máximo 2% dos pneus?

(d) Qual o intervalo central de durabilidade que contém 85% dos pneus fabricados pela Rodabem?

Exercícios: TCL na vida real

(e) Se você comprar 4 pneus Rodabem, qual será a probabilidade de que você utilizará a garantia (45.000 km) para trocar um ou mais destes pneus?

Ex. 26. O tempo de vida útil de uma lavadora de roupas automática tem distribuição aproximadamente Normal, com média de 3,1 anos e desvio padrão de 1,2 anos.
(a) Qual deve ser o valor do tempo de garantia dessa lavadora para que, no máximo, 15% das vendas originais exija substituição?
(b) Se esse tipo de lavadora tiver garantia de 1 ano, que porcentagem das vendas originais exigirá substituição?
Observação: Em outros exercícios que você encontra neste curso e/ou em outros cursos e livros didâticos, você vé que o tempo de vida útil está sendo modelada pela distribuição exponencial. Então, é a exponencial ou a Normal? A resposta é que tudo depende do processo físico que rege a duração de vida. Para uma lâmpada, por exemplo, a exponencial é mais apropreada. Mas quando a quebra pode ser causada por diversos fatores ou defeitos de fabricação, a modelagem do tempo de vida pela Normal pode ser mais precisa.

Exercícios: TCL na vida real

Ex. 27. Uma empresa produz televisores de 2 tipos, tipo A (comum) e tipo B (luxo), e garante a restituição da quantia paga se qualquer televisor apresentar defeito grave no prazo de seis meses. O tempo para ocorrência de algum defeito grave nos televisores tem distribuição normal, sendo que no tipo A com média 9 meses e desvio padrão 2 meses e, no tipo B, com média 12 meses e desvio padrão 3 meses. Os televisores de tipo A e B são produzidos com lucro de 1000 u.m. e 2000 u.m., respectivamente e, caso haja restituição, com prejuízo de 3000 u.m. e 8000 u.m., respectivamente.

(a) Calcule as probabilidades de haver restituição nos televisores do tipo A.

(b) Calcule o lucro médio para os televisores do tipo A.

Ajuda: *O lucro médio entende-se, naturalmente, como (a proporção dos aparelhos que não apresentaram defieto no prazo da garantia)$\times$ (lucro por aparelho vendido) $+$(a proporção dos aparelhos que defieto no prazo da garantia $\times$(prejuízo por aparelho restituido) onde prejuízo tem que ser tomado com sinal "$-$" na*

Exercícios: TCL na vida real

Ex. 28. Suponha que o tempo necessário para que estudantes completem uma prova tenha distribuição normal com média 90 minutos e desvio padrão 15 minutos.

(a) Qual é a proporção de estudantes que termina a prova

(i) em menos de 80 minutos?

(ii) em mais de 120 minutos?

(iii) entre 75 e 85 minutos?

(b) Qual é o tempo necessário para que 98% dos estudantes terminem a prova?

(c) Determinar o intervalo simétrico em torno do valor médio que contenha 70% dos valores do tempo para completarem a prova?

(d) Qual é a probabilidade de que, entre 5 estudantes escolhidos ao acaso, 3 deles completem a prova em menos de 80 minutos? (**Dica:** *É a probabilidade de obter* 3 *na Distribuição Binomial cujo parâmetro* n *é igual a* 5, *e cujo parâmetro* p *é igual à probabilidade de estudante retirado ao acaso da população terminar a prova em menos que* 80 *minutos.*)

Exercícios: TCL na vida real

Ex. 29. Numa certa população, o peso dos homens tem distribuição normal com média 75 kg e desvio padrão 10 kg, enquanto que o das mulheres é também normal com média 60 kg e desvio padrão 4 kg.

(a) Sorteando-se um homem qualquer, qual é a probabilidade dele ter peso acima de 65 kg?

(b) Sorteando-se uma mulher qualquer, qual é a probabilidade dela ter peso acima de 65 kg?

(c) Qual é a probabilidade de uma pessoa ter peso acima de 65 kg, sendo ela sorteada de um grupo em que o número de mulheres é o dobro do de homens? (**Obs.:** *Não é precisa fazer esse item. Nada que seja complicado na sua solução, mas não vale a pena voltar a atenção à distribuição condicional. Nada pareciso paraecerá na prova.*)

Exercícios: TCL na vida real

Ex. 30. Suponha que o tempo necessário para atendimento de clientes em uma central de atendimento telefônico siga uma distribuição normal de média de 8 minutos e desvio padrão de 2 minutos.

(a) Qual é a probabilidade de que um atendimento dure

(i) menos de 5 minutos;

(ii) mais do que 9,5 minutos;

(iii) entre 7 e 10 minutos?

(b) 75% das chamadas telefônicas requerem pelo menos quanto tempo de atendimento? Ou, em outras palavras, qual é o valor do limiar tal que exatamente 75% das chamadas mais curtas tenham suas durações abaixo desse limiar?

Observação: *Na realidade, na maioria dos casos do mundo real, o tempo necessário para atendimento de clientes em uma central de atendimento telefônico siga uma distribuição exponencial e não normal. Digo-lhe isso só para o título de conhecimento.*

Exercícios: TCL na vida real

Ex. 31. Sabe-se que a quantidade de ácido xanturênico excretado na urina de trabalhadores de uma indústria, que usa sulfeto de carbono como solvente, segue uma distribuição Normal com média 4,38 mg/15ml e desvio padrão 1,15mg/15ml. Determinar:

(a) A proporção de trabalhadores com quantidade de ácido xanturênico

(i) entre 2,20 e 4,00 mg/15ml;

(ii) acima de 5,50mg/15ml;

(b) a quantidade de ácido xanturênico que é superada por 80% dos trabalhadores.

Observação: *A solução deste exercício é mais curta que o texto do enunciado recheado por termos químicos e conceitos de saúde. Entretanto, o exercício apresenta perguntas reais colocadas para problema real cujos dados empíricas revelaram, de fato, que a quantidade do tal ácido na urina de tais trabalhadores se ajustava muito bem à distribuição Normal.*

Exercícios: TCL na vida real

Ex. 32. Doentes sofrendo de certa moléstia são submetidos a um tratamento intensivo, cujo tempo de cura é modelado por uma v.a. Normal de média 15 e desvio padrão 2 (em dias).

(a) Que proporção desses pacientes demora mais de 17 dias para se recuperar?

(b) Qual a probabilidade de que um paciente, escolhido ao acaso, apresente tempo de cura inferior a duas semanas?

(c) Que tempo de cura é necessário para recuperar 25% dos pacientes?

(d) Se 100 pacientes são escolhidos ao acaso, qual seria o número esperado de doentes curados em menos de 11 dias? (**Observação:** *Aqui há problema de duas etapas. Na primeira etapa, você retira 100 "bolas" da urna, ao acaso e com reposição, e observa suas "cores", sendo que na urna, inicialmente, há bolas pretas e brancas, e a proporção das pretas é igual à proporção dos doentes que curam-se em 11 dias ou menos. Então, o problema da primeira etapa é expressar o número de bolas pretas pela distribuição binomial com* $n = 100$ *e* p *igual à proporção d ebolas pretas*

Exercícios: TCL na vida real

Ex. 33. O número de vezes que um adulto respira, por minuto, depende da idade e varia grandemente de pessoa para pessoa. Suponha que a distribuição dessa variável aleatória seja normal com média de 16 e desvio padrão igual a 4.
(a) Um programa de exercícios respiratórios será oferecido a 10% das pessoas com respiração mais rápida. Como deve ser a respiração de uma pessoa para que ela seja incluída nesse programa?
(b) Em uma amostra de 100 pessoas, qual é o número esperado de pessoas cuja respiração excede a 22 vezes por minuto? (**Obs.:** *O esclarecimento acerca do "núero esperado" assim como da solução toda pode ser visto na observação ao item (d) do Exc.* **??**.)

Exercícios: TCL na vida real

Ex. 34. A nota média dos estudantes da USP no Provão de 2000 foi 465. Já no ano 2001, a nota média dos estudantes da USP foi 445. Os desvios-padrão, no entanto, foram iguais a 100, tanto em 2000 quanto em 2001. Em ambos os anos, a distribuição das notas dos estudantes da USP seguiu uma curva Normal.
(a) Qual a porcentagem de estudantes da USP que tirou mais que 600 no Provão de 2000?
(b) E no Provão de 2001?
(c) Qual porcentagem de estudantes da USP no Provão de 2000 tirou uma nota menor que a nota que somente 20% dos estudantes em 2001 não conseguiram ultrapassar? (**Observação:** Essa deve ser uma forma muito sofisticada de perguntar algo muito simples, mas eu não faço a menor idéia acerca daquilo que o autor dessa pergunta desejava. Se você, meu querido/a e talentoso/a aluno/a, conseguiu entender a pergunta, então dé sua resposta.)