By @kakashi_copiador

Aula 02 - Equipe Exatas

CNU (Bloco 1 - Infraestrutura, Exatas e Engenharia) Conhecimentos Específicos - Eixo Temático 5 - Geoprocessamento e Análise de Dados - 2024 (Pós-Edital)

Autor:

Alexandre Vastella, Equipe Exatas Estratégia Concursos, Monik Begname de Castro

17 de Janeiro de 2024

Índice

DISTRIBUIÇÃO DE BERNOULLI

Uma variável aleatória discreta X com Distribuição de Bernoulli assume apenas **2 valores possíveis**, **0** ou **1**, em um experimento realizado uma **única vez**. Esse experimento é chamado de **Ensaio de Bernoulli** ou **Experimento de Bernoulli**.

Um exemplo clássico dessa distribuição é o **lançamento** de uma **moeda**.

Chamamos os resultados possíveis de **sucesso** (em que a variável assume o valor **X = 1**) ou **fracasso** (em que a variável assume o valor **X = 0**). Se estivermos interessados na face CARA, esta representaria o **sucesso** e COROA representaria o **fracasso** (ou o contrário, se estivéssemos interessados na outra face).

Nesse exemplo, não faz muita diferença qual face corresponde ao sucesso ou ao fracasso, porque a probabilidade de ambas é a mesma: 50%.

Agora, vamos considerar que estamos torcendo para que o resultado do lançamento de um dado seja um **múltiplo de 3**. Nesse caso, os resultados 3 e 6 correspondem ao **sucesso** e os **demais** resultados correspondem ao **fracasso**.

Assim, teríamos **2 resultados de sucesso** (em que X = 1) e **4 resultados de fracasso** (em que X = 0). Logo, as probabilidades seriam as seguintes:

$$P(X = 1) = \frac{2}{6} = \frac{1}{3}$$

$$P(X = 0) = \frac{4}{6} = \frac{2}{3}$$

Normalmente, indicamos a probabilidade de **sucesso** como $\boldsymbol{p}$ e a probabilidade de **fracasso** como $\boldsymbol{q}$.

$$\boldsymbol{p} = P(X = 1)$$

$$\boldsymbol{q} = P(X = 0)$$

Para esse exemplo, temos $\boldsymbol{p} = \frac{1}{3}$ e $\boldsymbol{q} = \frac{2}{3}$.

Outro exemplo de uma distribuição de Bernoulli é associar o sucesso a apenas **uma das faces** do dado, por exemplo, a face 3. Nesse caso, temos **1 resultado de sucesso** (X = 1) e **5 resultados de fracasso** (X = 0):

$$\boldsymbol{p} = P(X = 1) = \frac{1}{6}$$

$$\boldsymbol{q} = P(X = 0) = \frac{5}{6}$$

Como há apenas **2 resultados possíveis**, as probabilidades de sucesso e de fracasso são **complementares**, isto é, a **soma** dessas 2 probabilidades é igual a **1**:

$$\boldsymbol{p} + \boldsymbol{q} = \mathbf{1}$$

$$\boldsymbol{q} = \mathbf{1} - \boldsymbol{p}$$

A **probabilidade de sucesso** $\boldsymbol{p}$ é a **única** informação necessária para **caracterizar** uma distribuição de Bernoulli, uma vez que a probabilidade de fracasso é calculada a partir dela. Por isso, podemos indicar que uma variável X segue distribuição de Bernoulli como $X \sim Ber(\boldsymbol{p})$.

Esse dado que **caracteriza** uma distribuição de probabilidade é chamado de **parâmetro**.

Um **mesmo experimento** pode estar associado a variáveis aleatórias com distribuições **distintas** de probabilidade, dependendo de como você o analisa.

O lançamento de um dado, por exemplo, pode estar associado a uma variável **uniforme**, com 6 valores equiprováveis; ou a distribuições de **Bernoulli** com parâmetros distintos; dentre outras distribuições possíveis.

Agora, vamos calcular a **esperança** da distribuição de Bernoulli. A fórmula geral da esperança é:

$$E(X) = \sum x.P(X = x)$$

$$E(X) = 0 \times q + 1 \times p$$

$$\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{p}$$

Para calcular a **variância**, primeiro calculamos $E(X^2)$:

$$E(X^2) = \sum x^2.P(X = x)$$

$$E(X^2) = 0^2 \times q + 1^2 \times p$$

$$E(X^2) = \boldsymbol{p}$$

Logo, a variância é:

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(X) = p - p^2 = p(1-p)$$

$$\boxed{\boldsymbol{V(X) = p.q}}$$

Para o exemplo do dado, em que o sucesso corresponde a uma face múltipla de 3, vimos que $\boldsymbol{p} = \frac{1}{3}$ e $\boldsymbol{q} = \frac{2}{3}$. Nesse caso, a esperança e a variância são:

$$E(X) = \boldsymbol{p} = \frac{1}{3}$$

$$V(X) = \boldsymbol{p.q} = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

Para o exemplo em que o sucesso corresponde a uma face específica do dado, vimos que $\boldsymbol{p} = \frac{1}{6}$ e $\boldsymbol{q} = \frac{5}{6}$. Nesse caso, a esperança e a variância são:

$$E(X) = \boldsymbol{p} = \frac{1}{6}$$

$$V(X) = \boldsymbol{p.q} = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$$

ESQUEMATIZANDO

Distribuição de Bernoulli (p)

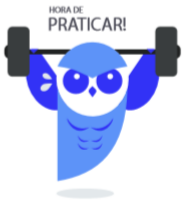
1 experimento: **Ensaio de Bernoulli**

2 resultados possíveis: sucesso ($X = 1$) ou fracasso ($X = 0$)

Probabilidade de sucesso: $P[X = 1] = \boldsymbol{p}$

Probabilidade de fracasso: $P[X = 0] = \boldsymbol{q} = \boldsymbol{1 - p}$

Esperança: $E(X) = \boldsymbol{p}$; Variância: $V(X) = \boldsymbol{p.q}$



(2017 – SES/DF) Considere o lançamento de um dado cúbico honesto cujas faces são numeradas de 1 a 6, após o qual é observado se o número da face voltada para cima é múltiplo de 3. Tendo em vista que um experimento como esse pode apresentar apenas dois resultados possíveis (sucesso ou falha), é correto afirmar que tal experiência denomina-se distribuição

a) de Bernoulli.

b) hipergeométrica.

c) de Poisson.

d) normal.

e) qui-quadrado.

Comentários:

A distribuição que trabalha com apenas 2 resultados possíveis (sucesso ou fracasso) para 1 ensaio (no caso, 1 lançamento do dado) é a distribuição de Bernoulli.

Gabarito: A.

(CESPE/2016 – TCE/PA) Se as variáveis aleatórias X e Y seguem distribuições de Bernoulli, tais que:

$$P[X = 1] = P[Y = 0] = 0{,}9$$

Então a média de Y é superior a 0,5.

Comentários:

A questão indaga sobre a **média (esperança) de Y**.

O enunciado informa que Y segue distribuição de Bernoulli, com probabilidade de **fracasso** de:

$$P(Y = 0) = q = 0{,}9$$

Logo, a probabilidade de **sucesso** é **complementar**:

$$p + q = 1$$

$$p = 1 - 0{,}9 = 0{,}1$$

Assim, a esperança de Y é:

$$E(Y) = p = 0{,}1$$

Que é **inferior** a 0,5.

Gabarito: Errado.

DISTRIBUIÇÃO BINOMIAL

Quando **repetimos** um **mesmo** Ensaio de **Bernoulli** (isto é, o experimento com **2 resultados possíveis**), damos origem à **Distribuição Binomial**.

Para termos uma distribuição binomial, é necessário que a **probabilidade de sucesso** de cada repetição seja a **mesma** (afinal, estamos repetindo um **mesmo** Ensaio de Bernoulli).

Além disso, as repetições precisam ser **independentes**, isto é, o resultado de um experimento **não** pode afetar o resultado de outro.

Os **parâmetros** dessa distribuição (os dados que a caracterizam) são o número $\boldsymbol{n}$ de repetições e a probabilidade de sucesso $\boldsymbol{p}$**.** Por isso, podemos indicar que uma variável X segue distribuição binomial como $X \sim B(\boldsymbol{n}, \boldsymbol{p})$.

A **Distribuição Binomial** pode ser considerada a **soma** de n variáveis com **Distribuição de Bernoulli** independentes, com **mesmo parâmetro** p.

Também é possível formar uma distribuição binomial pela **soma** de outras **distribuições binomiais**, com **mesmo parâmetro** p.

Por exemplo, sendo $n_X = 3$ o número de repetições da variável X e $n_Y = 4$ o número de repetições da variável Y, então a soma das variáveis $S = X + Y$ terá **distribuição binomial** com o seguinte número de repetições:

$$n_S = n_X + n_Y = 4 + 3 = 7$$

Um exemplo de distribuição binomial é o lançamento de um dado **n** = **3** vezes, em que o sucesso, corresponde à face 6 e o fracasso corresponde às demais faces.

Cada um desses lançamentos corresponde a um **Ensaio de Bernoulli**, em que podemos obter, em cada um deles, sucesso (X_{Ber} = 1) ou fracasso (X_{Ber} = 0), ou seja, **2** resultados possíveis.

Assim, o número de possíveis resultados para os 3 lançamentos é (princípio multiplicativo de combinatória):

2 x 2 x 2 = 8

As 8 possibilidades são:

{(0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), (1,1,1)}

A variável X com distribuição binomial representa o **número de sucessos** obtidos em todos os n lançamentos.

Sabendo que cada sucesso terá valor **1** e que cada fracasso terá valor **0**, então o número de sucessos pode ser calculado pela **soma** dos resultados dos Ensaios.

Para o nosso exemplo dos 3 lançamentos, a variável binomial pode assumir os seguintes valores:

- X = 0 (nenhum sucesso): {(0,0,0)}
- X = 1 (1 sucesso): {0,0,1), (0,1,0), (1,0,0)}
- X = 2 (2 sucessos): {(0,1,1), (1,0,1), (1,1,0)}
- X = 3 (3 sucessos): {(1,1,1)}

De maneira geral, a variável binomial pode assumir qualquer valor entre **0 e** $\boldsymbol{n}$.

Agora, vamos calcular a **probabilidade** de cada valor dessa variável binomial.

A probabilidade de obter a face 6 (**sucesso**) em um **único lançamento** é:

$$\boldsymbol{p} = \frac{1}{6}$$

E a probabilidade de não obter a face 6 (**fracasso**) em um **único lançamento** é complementar:

$$\boldsymbol{q} = 1 - p = \frac{5}{6}$$

Logo, a probabilidade de ter **fracasso** nos $\boldsymbol{n}$ = 3 lançamentos (0 sucesso: X = 0) corresponde à **interseção** de $n = 3$ fracassos. Por serem eventos **independentes**, a interseção é o **produto** das probabilidades:

$$P(X = 0) = \boldsymbol{q} \times \boldsymbol{q} \times \boldsymbol{q} = \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{125}{216}$$

Generalizando, para $\boldsymbol{n}$ **repetições**, a probabilidade de ter **0 sucesso** (**$\boldsymbol{n}$ fracassos**) é:

$$P(X = 0) = \underbrace{\boldsymbol{q} \times \boldsymbol{q} \times \ldots \times \boldsymbol{q}}_{\boldsymbol{n} \text{ vezes}} = \boldsymbol{q}^{\boldsymbol{n}}$$

Similarmente, a probabilidade de obter somente **sucessos** (nenhum fracasso) nos $\boldsymbol{n}$ = 3 lançamentos (X = 3) corresponde à **interseção** desses eventos **independentes**, dada pelo **produto**:

$$P(X = 3) = \boldsymbol{p} \times \boldsymbol{p} \times \boldsymbol{p} = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$$

Generalizando, para $\boldsymbol{n}$ **repetições**, a probabilidade de ter $\boldsymbol{n}$ **sucessos** (0 fracassos) é:

$$P(X=n)=\underbrace{p\times p\times\ldots\times p}_{\boldsymbol{n}\text{ vezes}}=p^{\boldsymbol{n}}$$

Já, a probabilidade de obter exatamente 1 sucesso (2 fracassos) é igual à probabilidade de obter 1 sucesso no **primeiro experimento** **OU** 1 sucesso no **segundo experimento** **OU** 1 sucesso no **terceiro experimento**:

$$P(X=1)=P(\{1,0,0\}\cup\{0,1,0\}\cup\{0,0,1\})$$

Como são eventos **mutuamente excludentes**, a probabilidade da união desses eventos é a **soma** dessas probabilidades:

$$P(X=1)=P(\{1,0,0\}\cup\{0,1,0\}\cup\{0,0,1\})=P(\{1,0,0\})+P(\{0,1,0\})+P(\{0,0,1\})$$

Agora, vamos calcular essas probabilidades para o nosso exemplo:

$$P(\{1,0,0\})=\ p\times q\times q=\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}=\frac{25}{216}$$

$$P(\{0,1,0\})=\ q\times p\times q=\frac{5}{6}\times\frac{1}{6}\times\frac{5}{6}=\frac{25}{216}$$

$$P(\{0,0,1\})=\ q\times q\times p=\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}=\frac{25}{216}$$

Como as probabilidades são todas **iguais**, em vez de somá-las, basta **MULTIPLICAR** o resultado por 3:

$$P(X=1)=3\times p\times q\times q=3\times\frac{25}{216}=\frac{75}{216}$$

E a probabilidade de obter exatamente 2 sucessos (1 fracasso) é igual à probabilidade de obter 1 fracasso no primeiro **OU** no segundo **OU** no terceiro experimento:

$$P(\{0,1,1\})=\ q\times p\times p=\frac{5}{6}\times\frac{1}{6}\times\frac{1}{6}=\frac{5}{216}$$

$$P(\{1,0,1\})=\ p\times q\times p=\frac{1}{6}\times\frac{5}{6}\times\frac{1}{6}=\frac{5}{216}$$

$$P(\{1,1,0\})=\ p\times p\times q=\frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}=\frac{5}{216}$$

Esses eventos também são mutuamente excludentes, então a probabilidade da união também corresponde à **soma** das probabilidades. Como elas são iguais, podemos **multiplicar** o resultado por 3:

$$P(X = 2) = 3 \times p \times p \times q = 3 \times \frac{5}{216} = \frac{15}{216}$$

Generalizando, para $\boldsymbol{n}$ **repetições**, vamos calcular a probabilidade de obter $\boldsymbol{k}$ **sucessos**, digamos, **nas primeiras $\boldsymbol{k}$ tentativas** e, portanto, $\boldsymbol{n - k}$ **fracassos** nas demais tentativas:

$$\underbrace{p \times p \times \ldots \times p}_{\boldsymbol{k}\text{ vezes}} \times \underbrace{q \times q \times \ldots \times q}_{\boldsymbol{n-k}\text{ vezes}} = p^{\boldsymbol{k}} \times q^{\boldsymbol{n-k}}$$

Porém, a ordem não precisa ser exatamente essa. Podemos obter k sucessos em **quaisquer** k tentativas. Por isso, devemos **multiplicar** esse resultado pelo número de maneiras de **reorganizar** os resultados de sucesso e fracasso.

Para isso, podemos simplesmente "escolher" quais serão as tentativas que resultarão em $\boldsymbol{k}$ **sucessos**, pois as outras $\boldsymbol{n - k}$ tentativas serão, necessariamente, **fracassos**.

O número de maneiras de "escolher" $\boldsymbol{k}$ tentativas, dentre $\boldsymbol{n}$ tentativas no total, corresponde à combinação de $\boldsymbol{k}$ elementos, dentre $\boldsymbol{n}$:

$$\boxed{C_{\boldsymbol{n},\boldsymbol{k}} = \frac{n!}{(n-k)! \times k!}}$$

Logo, a probabilidade de ter exatamente k sucessos (e, portanto, $n - k$ fracassos) é o produto da probabilidade $p^k \times q^{n-k}$ com a combinação $C_{n,k}$.

TOME NOTA!

$$\boxed{\boldsymbol{P(X = k) = C_{n,k} \times p^k \times q^{n-k}}}$$

A combinação $\boldsymbol{C_{n,k}}$ também pode ser indicado por $\boldsymbol{C_n^k}$ ou por $\boldsymbol{\binom{n}{k}}$.

Podemos calcular também a probabilidade de um **intervalo** ou de múltiplos valores da variável binomial.

Por exemplo, vamos primeiro calcular a probabilidade de obter 1 **OU** 2 sucessos, em 3 lançamentos de uma moeda (com p = q = 0,5). Como são eventos mutuamente exclusivos, devemos **somar** as probabilidades:

$$P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2)$$

A probabilidade de obter 1 sucesso é:

$$P(X = 1) = C_{3,1} \times 0{,}5^1 \times 0{,}5^2 = 3 \times 0{,}5 \times 0{,}25 = 0{,}375$$

A probabilidade de obter 2 sucessos é:

$$P(X = 2) = C_{3,2} \times 0{,}5^2 \times 0{,}5^1 = 3 \times 0{,}25 \times 0{,}5 = 0{,}375$$

Então, a probabilidade de obter 1 OU 2 sucessos é:

$$P(X = 1 \cup X = 2) = 0{,}375 + 0{,}375 = 0{,}75$$

Agora, vamos calcular a probabilidade de obter **pelo menos um** sucesso.

Em geral, para calcular a probabilidade de "pelo menos um", é mais fácil calcular a probabilidade **complementar**, isto é, a probabilidade de **nenhum**:

$$P(X \geq 1) = 1 - P(X = 0)$$

A probabilidade de obter 0 sucesso é:

$$P(X = 0) = C_{3,0} \times 0{,}5^3 \times 0{,}5^0 = 1 \times 0{,}125 \times 1 = 0{,}125$$

Logo, a probabilidade de obter **pelo menos um** sucesso é o complemento:

$$P(X \geq 1) = 1 - 0{,}125 = 0{,}875$$

Também podemos nos referir a uma distribuição binomial como uma **amostra de tamanho n** de uma **população** que segue uma **Distribuição de Bernoulli**.

Por exemplo, vamos supor uma população de peças, das quais 20% delas são **defeituosas**. Nesse caso, a seleção de uma **única peça** corresponde a um Ensaio de Bernoulli, pois há 2 resultados possíveis: **sucesso** (peça defeituosa) ou **fracasso** (peça não defeituosa).

A **probabilidade de sucesso** desse experimento equivale à **proporção de peças defeituosas** na população:

$$p = 20\% = 0{,}2$$

E a probabilidade de fracasso é complementar: $q = 1 - p = 0{,}8$

Suponha que vamos selecionar uma **amostra de $n = 5$** peças ao acaso. O número de peças defeituosas encontradas na amostra segue uma **distribuição binomial** com parâmetros $n = 5$ e $p = 0{,}2$.

Assim, a probabilidade de obter $k = 3$ peças defeituosas (portanto, $n - k = 2$ peças não defeituosas) é:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 3) = C_{5,3} \times 0{,}2^3 \times 0{,}8^2$$

A combinação de 3 elementos, dentre 5, é:

$$C_{5,3} = \frac{5!}{(5-3)! \times 3!} = \frac{5 \times 4 \times 3!}{2! \times 3!} = \frac{5 \times 4}{2} = 10$$

Logo, a probabilidade de obter 3 é igual a:

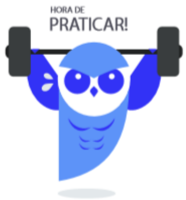
$$P(X = 3) = 10 \times 0{,}008 \times 0{,}64 = 0{,}0512$$

FIQUE ATENTO!

Para que as seleções sejam **independentes** (condição necessária para termos uma **distribuição binomial**), isto é, para que o resultado de uma **não influencie** no resultado da outra, a **proporção** de peças defeituosas **não** pode ser alterada a cada extração.

Para que essa condição seja satisfeita, temos duas alternativas:

- A seleção das peças é feita **com reposição**, isto é, a peça selecionada é **devolvida**; ou
- A população é **infinita** (ou **grande o suficiente**, em comparação com o tamanho da amostra, para permitir tal aproximação).



(2020 – Universidade do Estado do Pará – *Adaptada*) Julgue as seguintes afirmações:

I. As distribuições de Bernoulli e Binomial apresentam as mesmas características e, portanto, os mesmos parâmetros.

II. Repetições independentes de um ensaio de Bernoulli, com a mesma probabilidade de ocorrência de "sucesso", dão origem ao modelo Binomial;

Comentários:

Em relação à afirmativa I, o parâmetro de uma distribuição de Bernoulli é a probabilidade de sucesso p; e os parâmetros de uma distribuição binomial são a probabilidade de sucesso p e o número de repetições n.

Portanto, a afirmativa I está errada.

Em relação à afirmativa II, o modelo Binomial realmente consiste em repetições independentes de um ensaio de Bernoulli, com a mesma probabilidade de sucesso p. Logo, a afirmativa II está certa.

Gabarito: I – Errado; II – Certo.

(CESPE/2016 – Auditor de Controle Externo TCE/PA) Considere um processo de amostragem de uma população finita cuja variável de interesse seja binária e assuma valor 0 ou 1, sendo a proporção de indivíduos com valor 1 igual a p = 0,3.

Considere, ainda, que a probabilidade de cada indivíduo ser sorteado seja a mesma para todos os indivíduos da amostragem e que, após cada sorteio, haja reposição do indivíduo selecionado na amostragem.

A partir dessas informações, julgue o item subsequente.

Se for coletada uma amostra de tamanho n = 20, o número total de observações sorteadas com valor 1 terá distribuição binomial com parâmetros n e p.

Comentários:

A variável binária corresponde a uma distribuição de Bernoulli.

Coletando uma amostra de tamanho n, então o número de observações com o atributo sucesso corresponde a uma variável binomial, com parâmetros n e p.

Gabarito: Certo.

(CESPE/2016 – TCE/PA) Se as variáveis aleatórias X e Y seguem distribuições de Bernoulli, tais que

$$P[X = 1] = P[Y = 0] = 0{,}9$$

então X + Y segue uma distribuição binomial com parâmetros n = 2 e p = 0,3, se X e Y forem variáveis aleatórias independentes.

Comentários:

A distribuição binomial é caracterizada por repetições **independentes** de Ensaios de Bernoulli, com o **mesmo parâmetro p**.

O enunciado informa que:

- $P[X = 1] = 0{,}9$, ou seja, a probabilidade de sucesso de X é $p_X = 0{,}9$.
- $P[Y = 0] = 0{,}9$, ou seja, a probabilidade de fracasso de Y é $q_Y = 0{,}9$. Portanto, a probabilidade de sucesso de Y é: $p_Y = 1 - 0{,}9 = 0{,}1$

Como $p_X \neq p_Y$, então X + Y **não** segue uma distribuição binomial.

Gabarito: Errado.

(2018 – Câmara de Goiânia) Considere uma variável aleatória X com distribuição binomial e parâmetros p = 1/3 e n = 4. Qual é a probabilidade de X = 2?

a) 4/81

b) 1/9

c) 2/9

d) 8/27

Comentários:

A probabilidade P(X = k) de uma distribuição binomial é dada por:

$$P(X = k) = C_{n,k}.p^k.(1-p)^{n-k}$$

Para p = 1/3, n = 4 e k = 2, temos:

$$P(X = 2) = \frac{4!}{2!.2!}.\left(\frac{1}{3}\right)^2.\left(\frac{2}{3}\right)^2 = \frac{4\times 3}{2}.\frac{1}{9}.\frac{4}{9} = \frac{8}{27}$$

Gabarito: D

(2016 – ANAC) Em um determinado município, 70% da população é favorável a um certo projeto. Se uma amostra aleatória de cinco pessoas dessa população for selecionada, então a probabilidade de exatamente três pessoas serem favoráveis ao projeto é igual a

a) 40,58%

b) 35,79%

c) 42,37%

d) 30,87%

e) 37,46%

Comentários:

Considerando que a pessoa pode ser favorável ou não (não há outra possibilidade) e que o resultado da seleção de uma pessoa **não** afeta o de outra, então temos uma distribuição **binomial**.

Sabemos que:

- a proporção de pessoas favoráveis é p = 70% = 0,7 (logo, q = 1 – p = 0,3); e
- serão selecionadas n = 5 pessoas.

Então, a probabilidade de selecionar k = 3 pessoas favoráveis, P(X = 3), é dada por:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 3) = C_{5,3} \times (0{,}7)^3 \times (0{,}3)^2$$

A combinação de 3 elementos, dentre 5, é:

$$C_{5,3} = \frac{5!}{(5-3)! \times 3!} = \frac{5 \times 4 \times 3!}{2! \times 3!} = \frac{5 \times 4}{2} = 10$$

Assim, P(X = 3) é:

$$P(X = 3) = 10 \times 0{,}343 \times 0{,}09 = 0{,}3087 = 30{,}87\%$$

Gabarito: D

(FGV/2018 – ALE/RO) Uma moeda é lançada quatro vezes. A probabilidade de saírem mais caras do que coroas é de

a) $\frac{4}{16}$

b) $\frac{5}{16}$

c) $\frac{6}{16}$

d) $\frac{7}{16}$

e) $\frac{8}{16}$

Comentários:

Para saírem **mais caras** do que coroas em 4 lançamentos de uma moeda, é necessário que saiam 3 **OU** 4 caras. Assim, temos uma distribuição binomial com n = 4 e p = ½ (e também q = ½).

A probabilidade de saírem k = 4 caras (4 sucessos e 0 fracasso) é:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 4) = C_{4,4} \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^0 = 1 \times \frac{1}{16} \times 1 = \frac{1}{16}$$

A probabilidade de saírem k = 3 caras (3 sucessos e 1 fracasso) é:

$$P(X = 3) = C_{4,3} \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^1 = 4 \times \frac{1}{8} \times \frac{1}{2} = \frac{4}{16}$$

Portanto, a probabilidade de obter 3 OU 4 caras, sabendo que são eventos mutuamente exclusivos, é:

$$P(X=4)+P(X=3)=\frac{1}{16}+\frac{4}{16}=\frac{5}{16}$$

Gabarito: B.

(VUNESP/2019 – TJ/SP) Em uma eleição, sabe-se que 40% dos eleitores são favoráveis ao candidato X e o restante ao candidato Y. Extraindo uma amostra aleatória, com reposição, de tamanho 3 da população de eleitores, obtém-se que a probabilidade de que no máximo 1 eleitor da amostra seja favorável ao candidato X é igual a

a) 35,2%

b) 64,8%

c) 36,0%

d) 43,2%.

e) 78,4%

Comentários:

O enunciado informa que 40% dos eleitores são favoráveis a X (sucesso) e que os demais são favoráveis a Y (fracasso), ou seja, a seleção de uma pessoa ao acaso segue distribuição de Bernoulli.

Logo, a seleção de 3 pessoas **com reposição** (seleções independentes) configura uma **distribuição binomial** com n = 3 e p = 0,4 (q = 1 – p = 0,6).

Nessa distribuição, a probabilidade de encontrar k sucessos é dada por:

$$\boldsymbol{P(X=k)=C_{n,k}\times p^k\times q^{n-k}}$$

A probabilidade de obter **no máximo 1** eleitor favorável corresponde a obter 0 ou 1 sucesso:

$$\boldsymbol{P=P(X=0)+P(X=1)}$$

Para k = 0, temos:

$$P(X=0)=C_{3,0}\times 0{,}4^0\times 0{,}6^3=1\times 1\times 0{,}36=0{,}216$$

Para k = 1, temos:

$$P(X=1)=C_{3,1}\times 0{,}4^1\times 0{,}6^2=3\times 0{,}4\times 0{,}36=0{,}432$$

Assim, a probabilidade desejada é:

$$\boldsymbol{P(X=0\cup X=1)=0,216+0,432=0,648=64,8\%}$$

Gabarito: B

Esperança e Variância

E quanto à **esperança** dessa distribuição?

Vamos considerar o experimento de lançar uma moeda 100 vezes. Quantos resultados "CARA" você espera? 50, certo? Qual é a intuição por trás desse valor?

Se a probabilidade de sucesso é p e se estamos realizando esse experimento n vezes, espera-se que o número de sucessos obtidos mantenha essa **proporção**. Ou seja, o valor esperado é:

$$\boxed{E(X) = n \times p}$$

Para o exemplo dos 3 lançamentos de uma moeda, temos $p = \frac{1}{2}$ e $n = 3$. Então, o número de vezes que **esperamos** obter a face CARA é:

$$E(X) = n \times p = 3 \times \frac{1}{2} = \frac{3}{2} = 1{,}5$$

Mas esse valor não é inteiro!

Tudo bem! Algumas vezes obteremos mais CARAS, outras vezes menos, de modo que, **em média**, obteremos 1,5 CARA.

E a **variância** da distribuição binomial? A sua fórmula é:

$$\boxed{V(X) = n \times p \times q}$$

Para esse mesmo exemplo, com $q = \frac{1}{2}$, a variância é:

$$V(X) = n \times p \times q = 3 \times \frac{1}{2} \times \frac{1}{2} = \frac{3}{4} = 0{,}75$$

E o **desvio padrão** é a **raiz quadrada da variância**:

$$DP(X) = \sqrt{n \times p \times q}$$

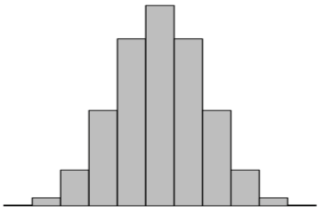
INDO MAIS FUNDO!

A combinação de k elementos, dentre n, é igual à combinação de $n - k$ elementos, dentre n: $C_{n,k} = C_{n,n-k}$.

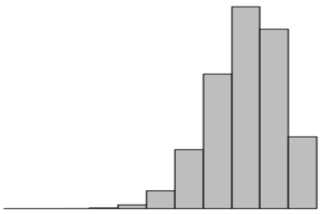
Ou seja, o **número de maneiras** de obter 0 sucesso é igual ao de obter n sucessos; o **número de maneiras** de obter 1 sucesso é igual ao de obter $n - 1$ sucessos; etc.

Assim, se tivermos $\boldsymbol{p = q}$, a **probabilidade** de obter 0 sucesso será igual à **probabilidade** de obter n sucessos; a probabilidade de obter 1 sucesso será igual à de $n - 1$ sucessos etc.

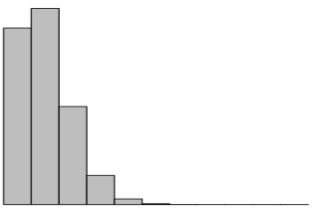
Portanto, temos uma distribuição **simétrica** se $\boldsymbol{p = q}$, isto é, se $\boldsymbol{p = q = 0,5}$, conforme ilustrado a seguir.



Quando a probabilidade de sucesso for maior $\boldsymbol{p > 0,5}$, teremos uma distribuição **assimétrica negativa**, conforme ilustrado a seguir.



Analogamente, quando a probabilidade de sucesso for menor $\boldsymbol{p < 0,5}$, teremos uma distribuição **assimétrica positiva**, conforme ilustrado a seguir.



Apesar de a distribuição binomial ser assimétrica sempre que $p \neq 0{,}5$, sempre que o produto $n \times p$ for um número **inteiro**, esse será o valor da **média,** da **mediana** e da **moda**!

Distribuição Binomial (n, p)

n Ensaios de Bernoulli **independentes**, com probabilidade de sucesso p

$$P(X = k) = C_{n,k}.p^k.q^{n-k}$$

Esperança: $E(X) = n.p$; Variância: $V(X) = n.p.q$

Observe que:

- A **esperança da binomial** é igual a n **vezes a esperança da Bernoulli**; e
- A **variância da binomial** é igual a n **vezes a variância da Bernoulli**.

$$\boldsymbol{E(X_{Binomial}) = n.E(X_{Bernoulli})}$$

$$\boldsymbol{V(X_{Binomial}) = n.V(X_{Bernoulli})}$$

(2019 – Universidade Federal do Acre) Seja X uma variável aleatória com distribuição binomial de parâmetros n e p. Então, pode-se dizer que a variância de X é dado por.

a) n

b) n.p

c) n.p(1 – p).

d) n.p^2

e) n.p^2(1 – p)

Comentários:

A variância de uma variável com distribuição binomial é dada por:

$$V(X) = n.p.q = n.p.(1-p)$$

Gabarito: C

(CESPE/2013 – TRT 17ª Região) Em toda distribuição binomial, a média será menor que a variância.

Comentários:

Em uma distribuição binomial, a média é E(X) = n.p e a variância V(X) = n.p.q.

Como q < 1 (por ser uma probabilidade), então:

$$V(X) = n.p.q < n.p.1 = E(X)$$

Portanto, a média é sempre **maior** que a variância.

Gabarito: Errado.

(2016 – IBGE) Quando um pesquisador vai a campo e aborda pessoas na rua para serem entrevistadas, o número de pessoas que aceita responder à pesquisa segue uma distribuição binomial.

Se o valor esperado dessa distribuição é 8, e sua variância é 1,6, então a probabilidade de uma pessoa aceitar responder à pesquisa é de

a) 1,6%

b) 16%

c) 20%

d) 50%

e) 80%

Comentários:

O enunciado informa que o valor esperado de uma variável com distribuição binomial é E(X) = 8.

Sabendo que, para uma distribuição binomial, temos E(X) = n.p, então:

$$E(X) = n.p = 8$$

$$n = \frac{8}{p}$$

O enunciado informa, ainda, que a variância é V(X) = 1,6.

Sabemos que:

$$V(X) = n.p.q = n.p.(1-p) = n.p - n.p^2$$

Então:

$$V(X) = n.p - n.p^2 = 1{,}6$$

Considerando que $n = \frac{8}{p}$, temos:

$$V(X) = \frac{8}{p}.p - \frac{8}{p}.p^2 = 1{,}6$$

$$8 - 8.p = 1{,}6$$

$$8.p = 6{,}4$$

$$p = 0{,}8 = 80\%$$

Gabarito: E.

(FCC/2014 – Auditor Fiscal da SEFAZ/RJ) Sabe-se que:

I. X é uma variável aleatória com distribuição binomial com média 2p e variância (2p-2p2).

II. Y é uma variável aleatória com distribuição binomial com média 5p e variância (5p-5p2).

III. A probabilidade de X ser inferior a 2 é igual a 15/16.

Nessas condições, a probabilidade de Y ser superior a 3 é igual a

a) 3/1.024

b) 1/64

c) 5/512

d) 15/1.024

e) 7/512

Comentários:

Sendo X uma variável com distribuição binomial, com média 2p, então:

$$E(X) = n_X.p = 2p$$

$$n_X = 2$$

O enunciado informa que a probabilidade de X ser inferior a 2 é P(X < 2) = 15/16, isto é:

$$P(X < 2) = P(X = 0) + P(X = 1) = 15/16$$

Sabendo que a probabilidade da distribuição binomial é $P(X = k) = C_{n,k}.p^k.q^{n-k}$, então para n = 2, temos:

$$P(X = 0) = C_{2,0}.p^0.q^2 = 1 \times 1 \times q^2 = q^2 = (1 - p)^2$$

$$P(X = 1) = C_{2,1}.p^1.q^1 = 2.p.q = 2.p.(1 - p)$$

Sabendo que a soma dessas probabilidades é igual a 15/16, então:

$$(1 - p)^2 + 2.p.(1 - p) = \frac{15}{16}$$

$$1 - 2p + p^2 + 2.p - 2p^2 = \frac{15}{16}$$

$$1 - p^2 = \frac{15}{16}$$

$$p^2 = 1 - \frac{15}{16} = \frac{1}{16}$$

Extraindo a raiz de ambos os lados da equação, sabendo que $p \geq 0$, temos:

$$p = \frac{1}{4}$$

Sabendo que, para Y, temos E(Y) = 5.p, então o número n de repetições de Y é:

$$E(Y) = n_Y.p = 5p$$

$$n_Y = 5$$

A probabilidade de Y ser superior a 3 corresponde à probabilidade de Y ser igual 4 OU igual a 5:

$$P(Y > 3) = P(Y = 4) + P(Y = 5)$$

Sabendo que p = ¼ (logo q = 1 – p = ¾) e n_Y = 5, temos:

$$P(Y = 4) = C_{5,4} \times \left(\frac{1}{4}\right)^4 \times \left(\frac{3}{4}\right)^1 = 5 \times \frac{1}{4^4} \times \frac{3}{4} = \frac{15}{1024}$$

$$P(Y = 5) = C_{5,5} \times \left(\frac{1}{4}\right)^5 \times \left(\frac{3}{4}\right)^0 = 1 \times \frac{1}{4^5} \times 1 = \frac{1}{1024}$$

Assim, a probabilidade de Y ser superior a 3 é:

$$P(Y > 3) = \frac{15}{1024} + \frac{1}{1024} = \frac{16}{1.024} = \frac{1}{64}$$

Gabarito: B

(FGV/2022 – TCU) A média e a variância de uma distribuição binomial são, respectivamente, 20 e 4. O número de ensaios (n) dessa distribuição é:

a) 20

b) 22

c) 25

d) 50

e) 100

Comentários:

Segundo o enunciado, a média de uma distribuição binomial é igual a 20 e a variância é igual a 4:

$$E(X) = n \times p = 20$$

$$V(X) = n \times p \times q = 4$$

A fórmula da variância é igual à da esperança, multiplicada por q:

$$V(X) = \underbrace{n \times p}_{E(X)} \times q = E(X) \times q$$

Sabendo que E(X) = 20 e que V(X) = 4, podemos calcular a probabilidade de fracasso:

$$V(X) = 20 \times q = 4$$

$$q = \frac{4}{20} = \frac{1}{5}$$

A probabilidade de sucesso é complementar:

$$p = 1 - q = 1 - \frac{1}{5} = \frac{4}{5}$$

Sabendo que $E(X) = n \times p = 20$, podemos calcular o número n de ensaios:

$$E(X) = n \times \frac{4}{5} = 20$$

$$n = \frac{20}{\frac{4}{5}} = 20 \times \frac{5}{4} = \frac{100}{4} = 25$$

Gabarito: C

DISTRIBUIÇÃO NORMAL

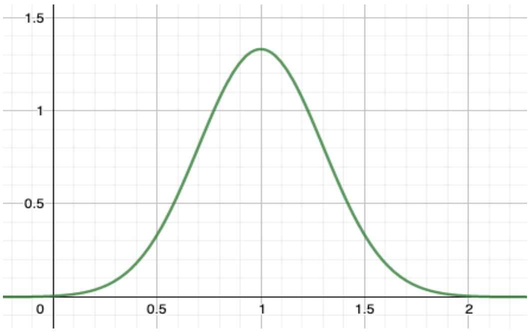
A distribuição **normal**, também chamada de **gaussiana**, é uma das distribuições contínuas mais importantes! A função densidade de probabilidade (f.d.p.) dessa distribuição é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in (-\infty, \infty)$$

Essa f.d.p. é bastante complicada, não é? Mas não se preocupe! Você não vai precisar integrar ou derivar!

Observe que essa função depende apenas dos **parâmetros μ** (média) e σ^2 (variância), que são parâmetros **independentes**.

No gráfico abaixo, temos uma f.d.p. com distribuição normal. Observe que a curva apresenta um formato de **sino**, que é uma característica de todas as variáveis normais.



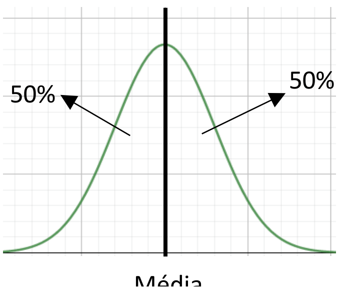
As distribuições normais são **simétricas**, ou seja, tem-se:

Média = Mediana = Moda

Logo, o valor de μ divide a distribuição em **duas partes iguais**. Sabendo que a área total, sob toda a curva, corresponde à probabilidade de todo o Espaço Amostral e, portanto, a 100%, então:

$$\boxed{P(X > \mu) = P(X < \mu) = 50\%}$$

A probabilidade de um intervalo corresponde à **área** sob a f.d.p. limitada por esse intervalo. Assim, a igualdade acima pode ser ilustrada como no gráfico seguir:



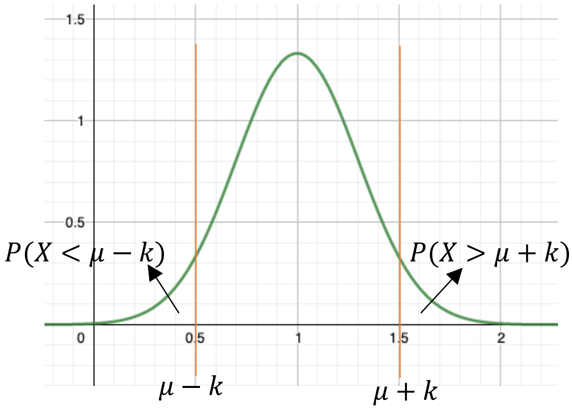
No exemplo do gráfico anterior, temos $\mu = 1$, logo:

$$P(X > 1) = P(X < 1) = 50\%$$

Mas a simetria não implica somente nisso. A partir da média, toda a distribuição de probabilidades para os valores superiores é igual à distribuição para os valores inferiores.

Assim, para qualquer k real, a probabilidade de a variável ser maior do que $\mu + k$ é igual à probabilidade de ser menor do que $\mu - k$:

$$\boxed{P(X > \mu + k) = P(X < \mu - k)}$$

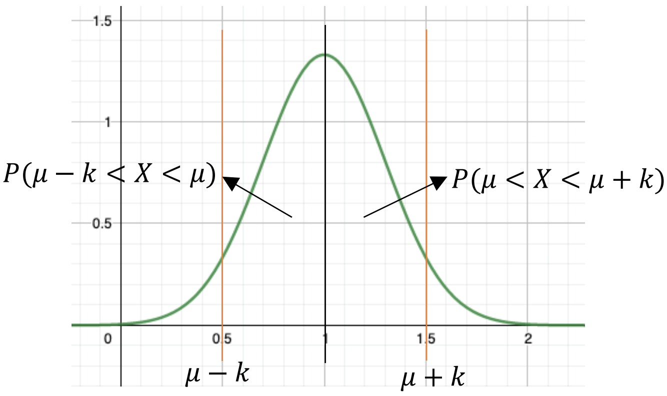


Em relação ao nosso exemplo, em que $\mu = 1$, temos:

- Para $k = 1$: $P(X > 2) = P(X < 0)$
- Para $k = 2$: $P(X > 3) = P(X < -1)$
- Para $k = 2{,}5$: $P(X > 3{,}5) = P(X < -1{,}5)$
- ...

Similarmente, as probabilidades associadas aos intervalos entre a média e esses limites $\mu + k$ e $\mu - k$ também são iguais, conforme equação abaixo e gráfico a seguir:

$$\boxed{P(\mu < X < \mu + k) = P(\mu - k < X < \mu)}$$



Em relação ao nosso exemplo em que $\mu = 1$, temos:

- Para $k = 1$: $P(1 < X < 2) = P(0 < X < 1)$
- Para $k = 2$: $P(1 < X < 3) = P(-1 < X < 1)$
- Para $k = 2{,}5$: $P(1 < X < 3{,}5) = P(-1{,}5 < X < 1)$
- ...

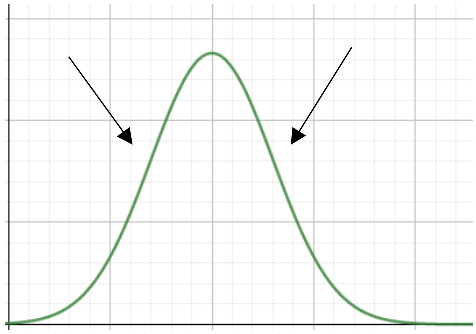
Podemos observar, ainda, que a curva normal apresenta **duas assíntotas**.

De modo geral, uma assíntota ocorre quando uma curva se **aproxima** cada vez mais a uma **reta**, porém **sem tocá-la**. A curva normal se **aproxima do eixo x** (eixo das abcissas) tanto para $x \to -\infty$, quanto para $x \to +\infty$. Por isso, dizemos que a curva normal é **duplamente assintótica**.

Além disso, existem **dois pontos de inflexão** na curva normal.

Pontos de inflexão são aqueles em que a **concavidade** da curva **muda**.

No início da curva normal, a concavidade está voltada para cima. No ponto (aproximado) indicado pela seta da esquerda, a **concavidade muda** para baixo, e no ponto (aproximado) indicado pela seta da direita, a **concavidade muda** novamente para cima.



Esses pontos de inflexão ocorrem precisamente a **1 desvio padrão** da média, ou seja, em $\boldsymbol{\mu - \sigma}$ e em $\boldsymbol{\mu + \sigma}$.

(FGV/2010 – SEAD-AP – *Adaptada*) Em relação à distribuição normal, julgue as afirmativas a seguir.

I – A função de densidade de probabilidade é simétrica em relação à média.

II – O valor da mediana é igual ao valor da média.

III – A média de uma variável aleatória com distribuição normal pode ser negativa.

Comentários:

Sabemos que a distribuição normal é simétrica em relação à média (logo, a afirmativa I está correta). Por ser simétrica, ela apresenta média = mediana (logo, a afirmativa II está correta).

A distribuição normal pode ter qualquer valor de média, inclusive negativa. Por exemplo, se estivermos tratando do lucro das empresas que vão à falência, provavelmente, a média será negativa. Logo, a afirmativa III está correta.

Resposta: Todas corretas.

Distribuição Normal Padrão

Para calcular os valores de probabilidade, temos uma **tabela** que relaciona os valores de intervalo da variável aos respectivos valores de probabilidade.

Essa tabela, inserida abaixo, se refere a uma distribuição normal $\boldsymbol{N(0,1)}$, isto é, com média $\boldsymbol{\mu = 0}$ e variância $\boldsymbol{\sigma^2 = 1}$, chamada de normal **padrão** ou **reduzida**, que denotamos por $\boldsymbol{Z}$.

Pelo gráfico anterior à tabela, deduzimos que os seus valores correspondem à probabilidade entre a média $\mu = 0$ e o valor de z indicado. Assim, os campos da tabela informam a probabilidade $P(0 < Z < z)$.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

Z (cont)	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

E quanto aos valores de z?

O valor de z começa a ser lido na primeira coluna (que apresenta as unidades e os décimos de z) e termina de ser lido na primeira linha (que apresenta os centésimos de z). Assim, a probabilidade $P(0 < Z < z)$ é o valor que está no campo, cuja linha corresponda à unidade e ao décimo de z e cuja coluna corresponda ao centésimo de z.

Por exemplo, para encontrar o valor de $P(0 < Z < 1{,}96)$, precisamos buscar o número que está na linha 1,9 e na coluna 0,06, conforme indicado abaixo. Podemos observar que $P(0 < Z < 1{,}96) = 0{,}475$.

Z	...	0,05	0,06	0,07	...
...	...	...	...	...	...
1,8	...	0,4678	0,4686	0,4693	...
1,9	...	0,4744	0,475	0,4756	...
2	...	0,4798	0,4803	0,4808	...
...	...	...	...	...	...

Também podemos fazer o caminho inverso, qual seja, encontrar o valor de z que corresponde à probabilidade desejada.

Vamos encontrar o valor de z tal que $P(0 < Z < z) = 0{,}40$, por exemplo. Para isso, devemos buscar o valor 0,40 nos campos da tabela. Como não consta exatamente esse valor, somente 0,3997 e 0,4015, optamos pelo valor mais próximo, isto é, 0,3997.

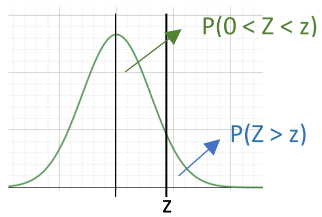
Este se encontra na linha 1,2 e na coluna 0,08, conforme indicado a seguir. Logo, concluímos que z = 1,28.

Z	...	0,07	0,08	0,09
...	...	0,3577	0,3599	0,3621
1,1	...	0,379	0,381	0,383
1,2	...	0,398	0,3997	0,4015
1,3	...	0,4147	0,4162	0,4177
...	...	...	...	...

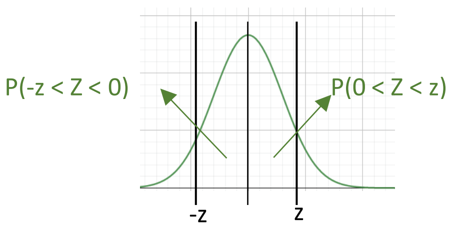
Para resolver questões envolvendo a tabela normal padrão é importante lembrar que essa distribuição é simétrica, com média $\mu = 0$.

TOME NOTA!

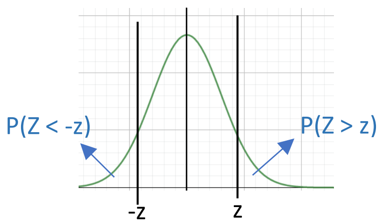
$$\boldsymbol{P(0 < Z < z) = 0,5 - P(Z > z)}$$



$$\boldsymbol{P(-z < Z < 0) = P(0 < Z < z)}$$

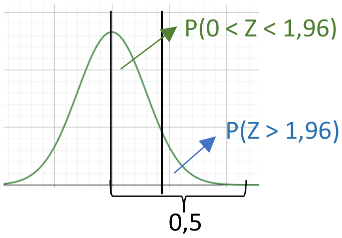


$$\boldsymbol{P(Z < -z) = P(Z > z)}$$

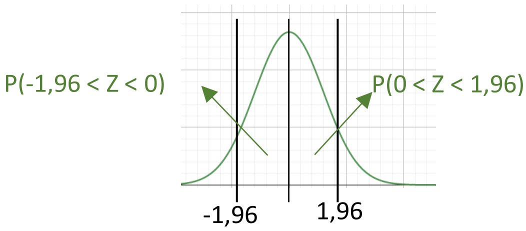


Supondo, por exemplo, z = 1,96, vimos que $P(0 < Z < 1{,}96) = 0{,}475$. Logo:

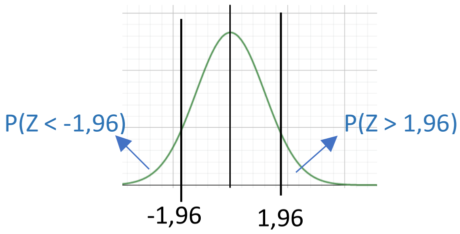
$$\boldsymbol{P(Z > 1,96) = 0,5 - P(0 < Z < 1,96)} = 0{,}5 - 0{,}475 = 0{,}025$$



$$P(-1,96 < Z < 0) = P(0 < Z < 1,96) = 0{,}475$$

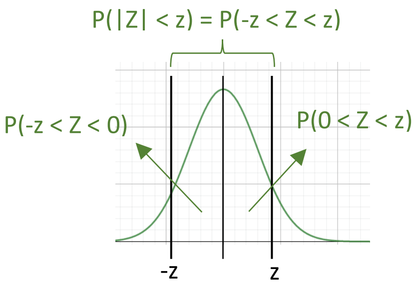


$$P(Z < -1,96) = P(Z > 1,96) = 0{,}5 - 0{,}475 = 0{,}025$$



A questão pode solicitar e/ou fornecer a probabilidade em módulo, da forma $P(|Z| < z)$ ou $P(|Z| > z)$. Para resolvê-las, é importante lembrar que a probabilidade $P(|Z| < z)$ corresponde a:

$$\boldsymbol{P(|Z| < z)} = P(-z < Z < z) = P(-z < Z < 0) + P(0 < Z < z)$$



Pela simetria da normal padrão, temos $P(-z < Z < 0) = P(0 < Z < z)$, logo:

$$\boldsymbol{P(|Z| < z) = 2 \times P(0 < Z < z)}$$

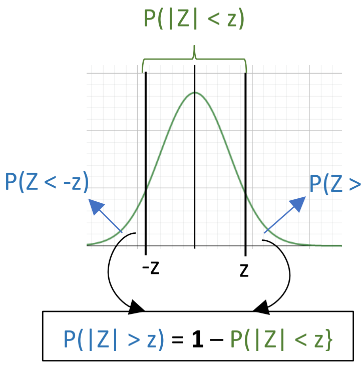
Supondo, por exemplo, z = 1,96, vimos que $P(0 < Z < 1{,}96) = 0{,}475$. Logo:

$$P(|Z| < 1.96) = 2 \times P(0 < Z < 1.96) = 2 \times 0{,}475 = 0{,}95$$

E a probabilidade $P(|Z| > z)$ pode ser calculada pela fórmula da probabilidade complementar:

$$\boldsymbol{P}(|\boldsymbol{Z}| > \boldsymbol{z}) = 1 - P(|Z| < z)$$

$$\boldsymbol{P(|Z| > z) = 1 - 2 \times P(0 < Z < z)}$$

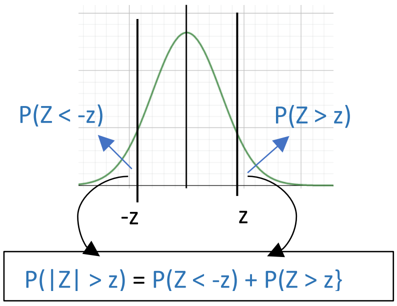


Supondo, por exemplo, z = 1,96, vimos que $P(0 < Z < 1{,}96) = 0{,}475$. Logo:

$$P(|Z| > z) = 1 - 2 \times P(0 < Z < z) = 1 - 2 \times 0{,}475 = 1 - 0{,}95 = 0{,}05$$

Ou, também podemos calcular $P(|Z| > z)$, aplicando-se o raciocínio análogo ao que fizemos anteriormente:

$$\boldsymbol{P}(|\boldsymbol{Z}| > \boldsymbol{z}) = P(Z < -z \cup Z > z) = P(Z < -z) + P(Z > z)$$



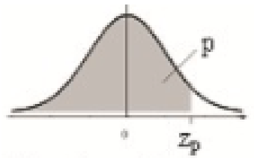
Pela simetria da normal padrão, temos $P(Z < -z) = P(Z > z)$, logo:

$$\boldsymbol{P}(|\boldsymbol{Z}| < \boldsymbol{z}) = 2 \times P(Z > z)$$

Para z = 1,96, em que $P(Z > 1{,}96) = 0{,}025$, temos:

$$P(|Z| > z) = 2 \times P(Z > z) = 2 \times 0{,}025 = 0{,}05$$

Existem, ainda, outros tipos de tabela para a distribuição normal padrão, que apresentam as probabilidades para outros intervalos, como por exemplo para a região indicada a seguir:



Esse tipo de tabela indica as probabilidades da forma $P(-\infty < Z < z)$, que correspondem à **função da distribuição normal acumulada**.

Transformação entre Distribuições Normais

Mas, e se a média da distribuição for diferente de zero e/ou a variância for diferente de 1?

Para isso, fazemos uma **transformação** de uma **distribuição normal qualquer** para a **distribuição normal padrão**, conforme fórmula indicada abaixo:

$$z = \frac{x-\mu}{\sigma}$$

Com essa transformação, encontramos os valores de **z** na distribuição normal padrão **associados** aos valores de **x** da distribuição normal de interesse, com média μ e desvio padrão σ.

Vamos supor uma distribuição normal com média $\mu = 1$ e desvio padrão $\sigma = 3$, em que estamos interessados no valor de $x = 7$. A transformação desse valor para a distribuição normal padrão é:

$$z = \frac{x-\mu}{\sigma} = \frac{7-1}{3} = 2$$

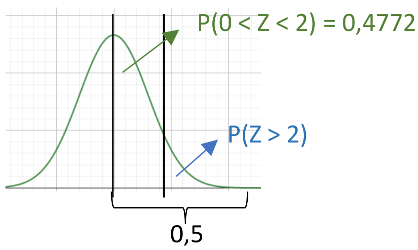
Isso significa que os **intervalos** associados a **z = 2** na distribuição **normal padrão** apresentam a **mesma probabilidade** daqueles associados a $\boldsymbol{x = 7}$ **na distribuição** $\boldsymbol{X}$, com média $\mu = 1$ e desvio padrão $\sigma = 3$.

Por exemplo:

$$\boldsymbol{P(X > 7) = P(Z > 2)}$$

Pela tabela da normal padrão, temos $P(0 < Z < 2) = 0{,}4772$, logo:

$$P(Z > 2) = 0{,}5 - 0{,}4772 = 0{,}0228$$



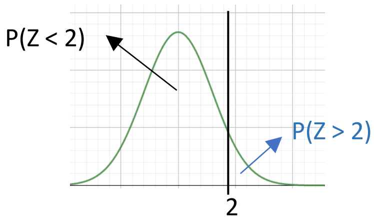
Portanto, $P(X > 7) = 0{,}0228 = 2{,}28\%$.

Analogamente, temos:

$$\boldsymbol{P(X < 7) = P(Z < 2)}$$

Sabemos que $P(Z > 2) = 0{,}5 - 0{,}4772 = 0{,}0228$, logo:

$$P(Z < 2) = 1 - P(Z > 2) = 1 - 0{,}0228 = 0{,}9772$$



Portanto, $P(X < 7) = 0{,}9772$.

Para encontrar intervalos envolvendo outros valores, por exemplo P(4 < X < 7), precisamos aplicar a transformação para ambos os valores. Para x = 4, temos:

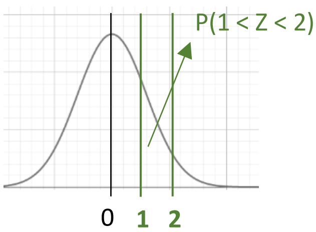
- $$z = \frac{x - \mu}{\sigma} = \frac{4 - 1}{3} = 1$$

Sabendo que a transformação para x = 7 é z = 2, então podemos concluir que:

$$P(4 < X < 7) = P(1 < Z < 2)$$

A probabilidade P(1 < Z < 2) pode ser calculada como:

$$P(1 < Z < 2) = P(0 < Z < 2) – P(0 < Z < 1)$$



Pela tabela da distribuição normal, observamos que P(0 < Z < 1) = 0,3413 e que P(0 < Z < 2)=0,4772, logo:

$$P(1 < Z < 2) = 0{,}4772 – 0{,}3413 = 0{,}1359$$

Assim, concluímos que P(4 < X < 7) = 0,1359 = 13,59%

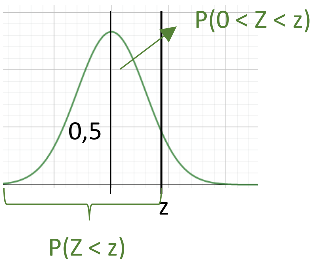
Também podemos fazer o **caminho inverso**, encontrando o **valor de x** em uma distribuição com média μ e desvio padrão σ, a partir de uma **probabilidade desejada**.

Para isso, primeiro encontramos o valor de **z** correspondente a essa probabilidade desejada, utilizando a tabela da normal padrão. Em seguida, aplicamos a fórmula da transformação.

Por exemplo, podemos calcular o valor de x para o qual a probabilidade de $P(X < x) = 0{,}8$, para a distribuição normal com os mesmos parâmetros do exemplo anterior ($\mu = 1$ e $\sigma = 3$).

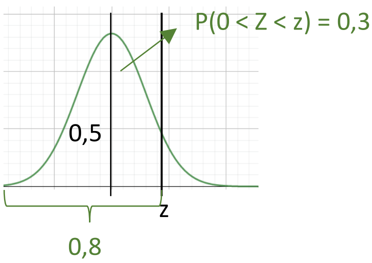
Considerando a simetria em torno de zero da normal padrão, temos que:

$$P(0 < Z < z) = P(Z < z) - 0{,}5$$



Assim, precisamos encontrar, na tabela normal padrão, o valor de z que corresponde a:

$$P(0 < Z < z) = 0{,}8 - 0{,}5 = 0{,}3$$



Pela tabela da distribuição normal padrão, observamos que esse valor é $z = 0{,}84$, pois P(0<Z<0,84) = 0,2995, que é o valor da tabela mais próximo de 0,3.

Substituindo os valores conhecidos na fórmula transformação ($\mu = 1$, $\sigma = 3$ e $z = 0{,}84$), podemos encontrar o valor de x, que delimita uma probabilidade $P(X < x) = 0{,}8$:

$$z = \frac{x - \mu}{\sigma}$$

$$0{,}84 = \frac{x - 1}{3}$$

$$x = 3 \times 0{,}84 + 1 = 3{,}52$$

Agora, vamos calcular as probabilidades associadas a intervalos genéricos, em função do **desvio padrão** σ.

Para calcular a probabilidade $\boldsymbol{P(\mu - \sigma < X < \mu + \sigma)}$, utilizamos a seguinte transformação, para $\boldsymbol{x = \mu + \sigma}$:

$$z = \frac{\boldsymbol{x} - \mu}{\sigma} = z = \frac{\boldsymbol{\mu + \sigma} - \mu}{\sigma} = 1$$

Para $\boldsymbol{x = \mu - \sigma}$, temos:

$$z = \frac{\boldsymbol{x} - \mu}{\sigma} = z = \frac{\boldsymbol{\mu - \sigma} - \mu}{\sigma} = -1$$

Portanto:

$$\boldsymbol{P(\mu - \sigma < X < \mu + \sigma)} = P(-1 < Z < 1) = \boldsymbol{P(-1 < Z < 0) + P(0 < Z < 1)}$$

De maneira geral, quando os intervalos são da forma $P(\mu - k < X < \mu + k)$, em que os extremos são **equidistantes** da média, basta fazermos a transformação para um dos extremos, pois o outro estará associado ao **mesmo valor de z**, porém multiplicado por –1.

Pela tabela da curva normal, temos $P(0 < Z < 1) = 0{,}3413$. Pela simetria da normal padrão, em torno da média 0, temos:

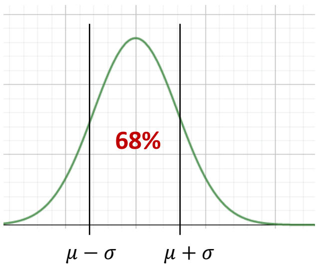
$$P(-1 < Z < 0) = P(0 < Z < 1) = 0{,}3413$$

Logo:

$$P(-1 < Z < 1) = P(-1 < Z < 0) + P(0 < Z < 1) = 0{,}3413 + 0{,}3413 = 0{,}6826$$

$$\boldsymbol{P(\mu - \sigma < X < \mu + \sigma) \cong 68\%}$$

Ou seja, a probabilidade de uma variável normal qualquer se afastar da média (para cima ou para baixo) em até **1 desvio padrão** é aproximadamente **68%**, conforme ilustrado a seguir.



Para $\boldsymbol{P(\mu - 2\sigma < X < \mu + 2\sigma)}$, temos:

$$z = \frac{x - \mu}{\sigma} = \frac{\mu + 2\sigma - \mu}{\sigma} = 2$$

Portanto:

$$\boldsymbol{P(\mu - 2\sigma < X < \mu + 2\sigma)} = P(-2 < Z < 2) = \boldsymbol{P(-2 < Z < 0)} + \boldsymbol{P(0 < Z < 2)}$$

Pela tabela, temos $P(0 < Z < 2) = 0{,}4772$. Considerando a simetria da normal padrão, temos:

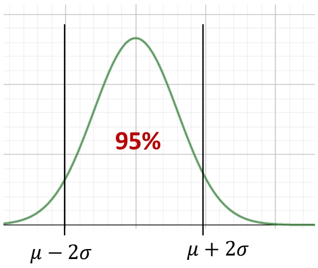
$$P(-2 < Z < 0) = P(0 < Z < 2) = 0{,}4772$$

E a probabilidade desejada é:

$$P(-2 < Z < 2) = P(-2 < Z < 0) + P(0 < Z < 2) = 0{,}4772 + 0{,}4772 = 0{,}9544$$

$$\boldsymbol{P(\mu - 2\sigma < X < \mu + 2\sigma) \cong 95\%}$$

Ou seja, a probabilidade de uma variável normal qualquer se afastar da média (para cima ou para baixo) em até **2 desvios padrão** é aproximadamente **95%**, como ilustrado abaixo.



Para $\boldsymbol{P(\mu - 3\sigma < X < \mu + 3\sigma)}$, obtemos, pela transformação, $z = 3$, logo:

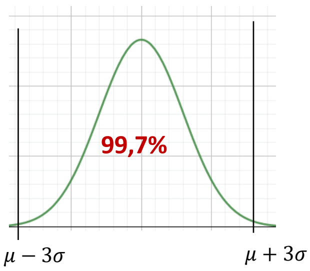
$$\boldsymbol{P(\mu - 3\sigma < X < \mu + 3\sigma)} = P(-3 < Z < 3) = \boldsymbol{P(-3 < Z < 0)} + \boldsymbol{P(0 < Z < 3)}$$

Pela tabela, temos $P(0 < Z < 3) = 0{,}4987$ e, pela simetria, $P(-3 < Z < 0) = 0{,}4987$, logo:

$$P(-3 < Z < 3) = P(-3 < Z < 0) + P(0 < Z < 3) = 0{,}4987 + 0{,}4987 = 0{,}9974$$

$$\boldsymbol{P(\mu - 3\sigma < X < \mu + 3\sigma) \cong 99{,}7\%}$$

Ou seja, a probabilidade de uma variável normal qualquer se afastar da média (para cima ou para baixo) em até **3 desvios padrão** é aproximadamente **99,7%**, ilustrado abaixo.



Essas probabilidades (**68%** para $\boldsymbol{\mu \pm \sigma}$; **95%** para $\boldsymbol{\mu \pm 2\sigma}$ e **99,7%** para $\boldsymbol{\mu \pm 3\sigma}$) compõem a chamada **Regra Empírica**. Algumas questões (não muitas) exigem que você memorize essas probabilidades.

ESQUEMATIZANDO

Distribuição Normal: $N(\mu, \sigma^2)$
Simétrica, com formato de sino, definida em **toda a reta real**
Regra Empírica: 68% para $\boldsymbol{\mu \pm \sigma}$; **95%** para $\boldsymbol{\mu \pm 2\sigma}$ e **99,7%** para $\boldsymbol{\mu \pm 3\sigma}$
Transformação para a **Normal Padrão** N(0,1): $\boldsymbol{z = \frac{x-\mu}{\sigma}}$

HORA DE PRATICAR!

(FGV/2010 – SEAD-AP – *Adaptada*) Em relação à distribuição normal, julgue as afirmativas a seguir:

I – Se X tem distribuição normal com média μ e variância σ^2 então a variável $Z = \frac{(X-\mu)}{\sigma^2}$ tem distribuição normal padrão.

II – A probabilidade de que uma variável Z que tenha distribuição normal padrão seja maior que 5 é aproximadamente igual a 0.

Comentários:

Em relação à afirmativa I, a transformação para a Normal Padrão é $Z = \frac{(X-\mu)}{\sigma}$, ou seja, a divisão é pelo desvio padrão, não pela variância. Por isso, a afirmativa I está incorreta.

Em relação à afirmativa II, a distribuição se concentra em 3 desvios-padrão para ambos os lados (mais de 99% se encontram nesse intervalo). De fato, as tabelas da normal padrão, em geral, fornecem valores até z=3,99 porque valores a probabilidade de Z ser maior que isso é praticamente nula. Logo, a afirmativa II está correta.

Resposta: I – incorreta, II – correta.

(VUNESP/2009 – CETESB) Para um determinado horário, considerando-se todos os dias de um período, ao se calcular a média de congestionamento de trânsito em km obtém-se o valor μ e desvio padrão σ. Considerando-se que os valores obtidos pela variável e suas respectivas probabilidades constituem uma distribuição normal, no intervalo de $(\mu - \sigma)$ até $(\mu + \sigma)$, a percentagem dos dados contidos é cerca de

a) 25%

b) 50%

c) 68%

d) 94%

e) 99%

Comentários:

Essa questão exige o conhecimento da **Regra Empírica**. Vimos que o intervalo de $(\mu - \sigma)$ até $(\mu + \sigma)$ concentra 68% da distribuição normal.

Gabarito: C

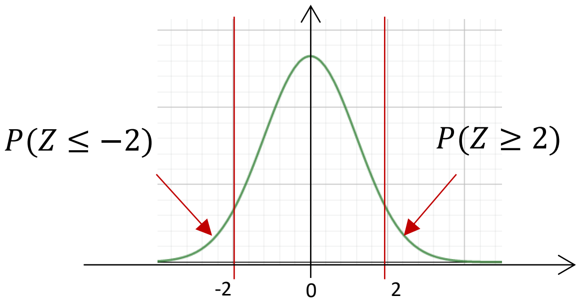
(CESPE/2016 – Analista da FUNPRESP-JUD) A simetria de Z implica que P(Z ≥ 2) = 1 – P(Z ≤ -2).

Comentários:

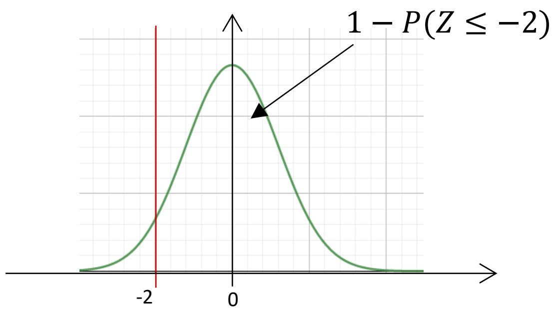
A simetria da curva normal com média igual a zero implica na seguinte relação entre P(Z ≥ 2) e P(Z ≤ -2):

$$P(Z \geq 2) = P(Z \leq -2)$$

Essa relação está ilustrada no gráfico abaixo.



A relação descrita no enunciado iguala a probabilidade $P(Z \geq 2)$ à probabilidade $1 - P(Z \leq -2)$. Esta corresponde a toda a região indicada abaixo:



Podemos observar que $P(Z \geq 2)$ é bem menor que 50%, enquanto $1 - P(Z \leq -2)$ é bem maior que 50%, ou seja:

$$P(Z \geq 2) \neq 1 - P(Z \leq -2)$$

Logo, o item está errado.

Gabarito: Errado.

(FGV/2022 – EPE) O salário médio dos funcionários de uma empresa é normalmente distribuído com média de R$ 2.500,00 e desvio padrão de R$ 1.500,00. A empresa divide os funcionários em 5 classes, a saber: M, N, O, P e Q, onde "M" é a classe com melhor salário e "Q" a classe com menor salário.

Se apenas 5% dos funcionários dessa empresa estão na classe "M", o menor valor do salário do funcionário para ele pertencer à classe "M" é

[Considere que P(Z ≤ 1,64) = 0,95.]

a) 3900,00

b) 4170,00

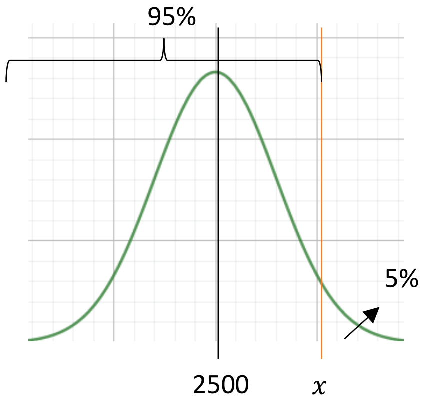
c) 4960,00

d) 5160,00

e) 5350,00

Comentários:

O enunciado informa que os salários seguem distribuição normal com média $\mu = R\$\ 2.500{,}00$ e desvio padrão $\sigma = R\$1.500{,}00$; e pede o valor do menor salário da classe M, associada aos 5% melhores salários, conforme ilustrado a seguir:



Sabendo que 5% (ou 0,05) da distribuição é maior do que o valor buscado, então 95% (ou 0,95) da distribuição é **menor** e o enunciado informa justamente que P(Z ≤ 1,64) = 0,95.

Assim, devemos aplicar a transformação para z = 1,64, sabendo que a média é 2500 e o desvio padrão é 1500:

$$z = \frac{x - \mu}{\sigma}$$

$$1{,}64 = \frac{x - 2500}{1500}$$

$$x - 2500 = 1{,}64 \times 1500 = 2460$$

$$x = 4960$$

Gabarito: C

(VUNESP/2014 – EMPLASA) O tempo de vida da população de um determinado país tem distribuição normal com a média igual a 68 anos e o desvio padrão igual a 11.

Considere os valores da tabela e a fórmula $Z = \frac{X-\mu}{\sigma}$:

Z	Distribuição normal reduzida
0,5	0,1915
1,0	0,3413
1,5	0,4332
2,0	0,4772
2,5	0,4938

A probabilidade de uma pessoa viver mais do que 90 anos é de

a) 15,87%

b) 6,68%

c) 4,82%

d) 3,36%

e) 2,28%

Comentários:

Sendo $\mu = 68$ e $\sigma = 11$, temos a seguinte transformação para $x = 90$:

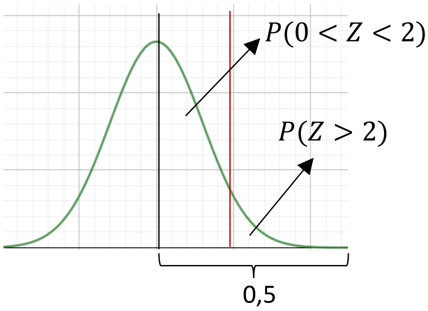
$$z = \frac{x - \mu}{\sigma} = \frac{90 - 68}{11} = \frac{22}{11} = 2$$

Pela tabela, que apresenta a probabilidade $P(0 < Z < z)$, temos:

$$P(0 < Z < 2) = 0{,}4772$$

Logo:

$$P(Z > 2) = 0{,}5 - P(0 < Z < 2) = 0{,}5 - 0{,}4772 = 0{,}0228$$



Gabarito: E.

(FCC/2015 – Auditor Fiscal da SEFAZ/PI) Se Z tem distribuição normal padrão, então: P(Z < 0,4) = 0,655; P(Z < 1,2) = 0,885; P(Z < 1,6) = 0,945; P(Z < 1,8) = 0,964; P(Z < 2) = 0,977.

O efeito do medicamento A é o de baixar a pressão arterial de indivíduos hipertensos. O tempo, em minutos, decorrido entre a tomada do remédio e a diminuição da pressão é uma variável aleatória X com distribuição normal, tendo média μ e desvio padrão σ. Se o valor de μ é de 56 min e o valor de σ é de 10 min, a probabilidade de X estar compreendido entre 52 min e 74 min é igual a

a) 30,9%

b) 56,0%

c) 61,9%

d) 52,4%

e) 64,5%

Comentários:

A probabilidade de X estar entre 52 min e 74 min pode ser calculada a partir das transformações para x = 74 min e para x = 52 min, considerando a média de 56 min e desvio padrão de 10 min.

Para x = 74, temos:

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{74 - 56}{10} = \frac{18}{10} = 1{,}8$$

A transformação para x = 52 é:

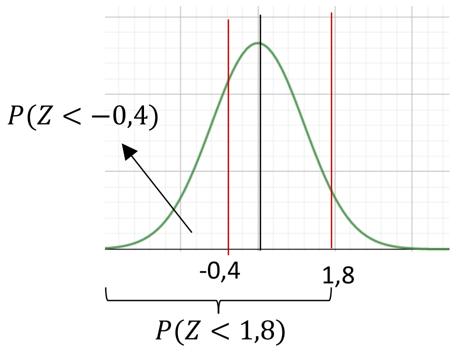
$$z = \frac{52 - 56}{10} = \frac{-4}{10} = -0{,}4$$

Então, a probabilidade de X estar compreendido entre 52 min e 74 min corresponde a:

$$P(52 < X < 74) = P(-0{,}4 < Z < 1{,}8)$$

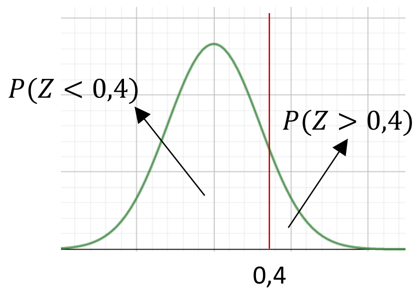
Essa probabilidade pode ser calculada como:

$$P(-0{,}4 < Z < 1{,}8) = P(Z < 1{,}8) - P(Z < -0{,}4)$$



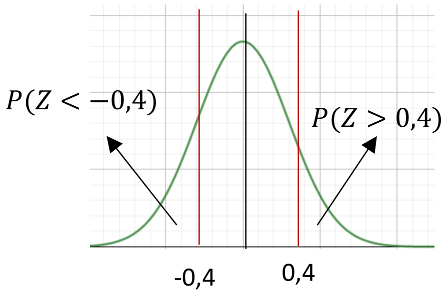
Pela tabela observamos que P(Z < 1,8) = 0,964. Além disso, temos P(Z < 0,4) = 0,655, logo, o seu complementar é:

$$P(Z > 0{,}4) = 1 - 0{,}655 = 0{,}345$$



Pela simetria da normal padrão, temos:

$$P(Z < -0{,}4) = P(Z > 0{,}4) = 0{,}345$$



Inserindo esses valores na equação acima, temos:

$$P(52 < X < 74) = P(Z < 1{,}8) - P(Z < -0{,}4) = 0{,}964 - 0{,}345 = 0{,}619$$

Gabarito: C

SOMA DE VARIÁVEIS E O TEOREMA CENTRAL DO LIMITE

Nesta seção, veremos a **soma** de variáveis com distribuição normal e, também, com outras distribuições. Ao final, veremos como aproximar uma distribuição binomial a uma normal.

Soma de Variáveis com Distribuição Normal

A **soma** de variáveis **independentes** com distribuição normal também segue uma distribuição **normal**, cuja média corresponde à **soma das médias** das variáveis e a variância corresponde à **soma das variâncias**.

Se $X_1, X_2, \dots, X_n$ são variáveis **independentes** que seguem distribuição **normal**, então a soma $Y = X_1 + X_2 + \cdots + X_n$ também segue distribuição **normal** com média e variância dadas respectivamente por:

$$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$V(Y) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

Ademais, a **diferença** entre duas variáveis **independentes** com distribuição normal também segue uma distribuição **normal**, cuja média e variância podem ser calculadas pelas propriedades dessas medidas.

Se X_1 e X_2 são variáveis **independentes** com distribuição **normal**, então a diferença $Y = X_1 - X_2$ segue distribuição **normal** com média e variância dadas respectivamente por:

$$E(Y) = E(X_1) - E(X_2)$$

$$V(Y) = V(X_1) + V(X_2)$$

Por exemplo, sendo $Y = X_1 - X_2$, em que X_1 e X_2 são variáveis normais independentes, com médias $E(X_1) = 3$ e $E(X_2) = 5$, e variâncias $V(X_1) = 4$ e $V(X_2) = 1$, então, Y terá **distribuição normal**, com parâmetros:

$$E(Y) = E(X_1) - E(X_2) = 3 - 5 = -2$$

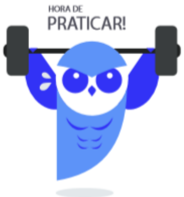
$$V(Y) = V(X_1) + V(X_2) = 4 + 1 = 5$$

Pontue-se que a **soma**, a **subtração**, a **multiplicação** ou a **divisão** de uma distribuição normal por uma **constante** real **também** segue **distribuição normal**, cuja média e variância podem ser calculadas pelas propriedades de esperança e variância.

Por exemplo, sendo X uma variável normal com média E(X) = 3 e variância V(X) = 4, então a variável Y = 2X – 6 terá **distribuição normal**, com média e variância dadas por:

$$E(Y) = E(2X - 6) = 2.E(X) - 6 = 2 \times 3 - 6 = 0$$

$$V(Y) = V(2X - 6) = 2^2.V(X) = 4 \times 4 = 16$$



(2020 – FADESP) A variável aleatória X tem distribuição normal com média μ = 2 e variância σ^2 = 9. Seja Y uma variável aleatória definida por Y = 2X + 1. Nestas condições, pode-se afirmar que Y tem distribuição

a) normal com média μ = 2 e variância σ^2 = 30.

b) qui-quadrado com média μ = 5 e variância σ^2 = 36.

c) normal com média μ = 5 e variância σ^2 = 9.

d) normal com média μ = 5 e variância σ^2 = 36.

Comentários:

Vimos que a soma de variáveis normais segue uma distribuição normal, mesmo quando somadas a uma constante. Logo, a variável Y possui distribuição normal, com média:

$$E(Y) = E(2X + 1) = 2.E(X) + 1$$

Sendo E(X) = 2, temos:

$$E(Y) = 2.2 + 1 = 5$$

A variância é:

$$V(Y) = V(2X + 1) = 4V(X)$$

Sendo V(X) = 36, então:

$$V(Y) = 4.9 = 36.$$

Gabarito: D.

(2018 – Petrobras) As variáveis aleatórias X e Y são independentes. A variável X segue uma distribuição Normal com média 4 e variância 16, e a Y segue uma distribuição Normal com média 9 e variância 1. A distribuição de X - Y é Normal com

a) média -5 e variância 15.

b) média -5 e variância 17.

c) média 5 e variância 15.

d) média 5 e variância 17.

e) média 13 e variância 15.

Comentários:

Vimos que a diferença de variáveis normais segue uma distribuição normal, com média:

$$\mu = \mu_X - \mu_Y = 4 - 9 = -5$$

E variância:

$$\sigma^2 = \sigma_X^2 + \sigma_Y^2 = 16 + 1 = 17$$

Gabarito: B.

(FCC/2015 – Analista do CNMP – *Adaptada*) Se Z tem distribuição normal padrão, então: P(Z < 0,5) = 0,591; P(Z < 1) = 0,841; P(Z < 1,15) = 0,8951; P(Z < 1,17) = 0,879; P(Z < 1,2) = 0,885; P(Z < 1,4) = 0,919; P(Z < 1,64) = 0,95; P(Z < 2) = 0,977; P(Z < 2,06) = 0,98; P(Z < 2,4) = 0,997.

Sejam X_1, X_2, ... , X_n variáveis aleatórias independentes e identicamente distribuídas com distribuição normal padrão. Seja a variável aleatória $Y = X_1 + X_2 + \cdots + X_n$.

Considerando essa informação, julgue a seguinte afirmativa.

Para n = 4, P(− 2 < Y < 1) = 0,432.

Comentários:

A soma de n = 4 variáveis aleatórias independentes com distribuição normal padrão N(0,1) segue distribuição normal com média e variância dadas por:

$$E(Y) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 0 + 0 + 0 + 0 = 0$$

$$V(Y) = V(X_1) + V(X_2) + V(X_3) + V(X_4) = 1 + 1 + 1 + 1 = 4$$

Logo, o desvio padrão é:

$$\sigma = \sqrt{V(Y)} = 2$$

Para calcular a probabilidade desejada, utilizaremos a fórmula da transformação para a normal padrão. O valor de z para y = 1 é:

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{1 - 0}{2} = 0{,}5$$

E o valor de z para y = -2 é:

$$z = \frac{-2 - 0}{2} = -1$$

Logo, a probabilidade desejada corresponde a:

$$P(-2 < Y < 1) = P(-1 < Z < 0,5) = P(Z < 0,5) - P(Z < -1)$$

Pela tabela fornecida, observamos que P(Z < 0,5) = 0,591. Temos também que P(Z < 1) = 0,841, logo o seu complementar é:

$$P(Z > 1) = 1 - 0,841 = 0,159$$

Pela simetria da normal padrão, temos:

$$P(Z < -1) = P(Z > 1) = 0,159$$

Assim, a probabilidade P(-2 < Y < 1) corresponde a:

$$P(Z < 0,5) - P(Z < -1) = 0,591 - 0,159 = 0,432$$

Resposta: Certo.

Teorema Central do Limite

Um dos motivos pelos quais a **Distribuição Normal** (ou de **Gauss**) é tão importante em Estatística decorre do **Teorema Central do Limite** (TLC), que trata da soma de variáveis que seguem uma distribuição **qualquer**:

*Para variáveis aleatórias $X_1, X_2, \dots, X_n$ **independentes e identicamente distribuídas** (i.i.d), a distribuição da soma $X_1 + X_2 + \dots + X_n$ **tende** a uma distribuição **normal**, à medida em que $\boldsymbol{n}$ **cresce**, cuja média e variância são dadas por:*

$$E(Y) = E(X_1) + E(X_2) + \dots + E(X_n) = n.E(X)$$

$$V(Y) = V(X_1) + V(X_2) + \dots + V(X_n) = n.V(X)$$

Note que as variáveis podem apresentar **qualquer** distribuição, mesmo assim, a sua soma seguirá **aproximadamente** uma distribuição **normal**.

A convergência de que trata o TLC é uma **convergência em distribuição**. A convergência em distribuição de uma sequência X_n a uma variável X significa que **as funções de distribuição acumuladas** de X_n e de X convergem.

As funções densidade de probabilidade não necessariamente convergem, ou seja, **não** se trata de uma convergência ponto a ponto. A convergência em distribuição é o tipo mais **fraco** de convergência, dentre os usuais.

Vamos supor que haja 100 variáveis **independentes** X_i, com i = 1, 2, ..., 100, todas com média **$E(X_i) = 3$** e variância **$V(X_i) = 4$**. Assim, sendo $Y = X_1 + X_2 + ... + X_{100}$, então a média e variância de Y serão:

$$E(Y) = 100 \text{ x } E(X_i) = 100 \text{ x } 3 = 300$$

$$V(Y) = 100 \text{ x } V(X_i) = 100 \text{ x } 4 = 400$$

Ainda que as variáveis **não** sejam **identicamente distribuídas**, isto é, ainda que apresentem distribuições **distintas**, mesmo assim, a sua soma terá, **aproximadamente**, uma distribuição normal, desde que as variáveis sejam **independentes** e tenham variâncias similares.

Nesse caso, a esperança e variância serão:

$$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$V(Y) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

Atenção: devemos somar as **variâncias**, **não** os desvios padrão!

(FGV/2017 – IBGE) Sejam $X_1, X_2, X_3, ..., X_{64}$ variáveis aleatórias discretas, com distribuição Binomial, todas com p = 0,25 e n = 12. Também são conhecidos valores da função distribuição acumulada da normal-padrão, mais especificamente: $\phi(2) = 0,977$, $\phi(1,5) = 0,933$, $\phi(1,25) = 0,894$.

No caso da extração de uma amostra (n = 64), a probabilidade de que a soma dos valores seja superior a 207 é igual a:

a) 0,023

b) 0,046

c) 0,067

d) 0,106

e) 0,134

Comentários:

Pelo Teorema Central do Limite, a soma de N variáveis independentes identicamente distribuídas X_i, cada uma com média E(X) e variância V(X), é uma variável Y, com distribuição aproximadamente normal, com média e variância dadas por:

$$E(Y) = N.E(X)$$

$$V(Y) = N.V(X)$$

Sendo X uma variável binomial com n = 12 e p = 0,25 (logo, q = 1 – p = 0,75), então:

$$E(X) = n.p = 12 \times 0{,}25 = 3$$

$$V(X) = n.p.q = 12 \times 0{,}25 \times 0{,}75 = 9/4$$

Assim, a distribuição da soma de N = 64 variáveis X terá média e variância dadas por:

$$E(Y) = N.E(X) = 64 \times 3 = 192$$

$$V(Y) = N.V(X) = 64 \times \frac{9}{4} = 144$$

O desvio padrão de Y é, portanto:

$$\sigma_Y = \sqrt{V(X)} = \sqrt{144} = 12$$

Assim, a probabilidade $P(Y > 207)$ pode ser calculada pela seguinte transformação:

$$z = \frac{y - \mu}{\sigma} = \frac{207 - 192}{12} = \frac{15}{12} = 1{,}25$$

O enunciado informa que $P(-\infty < Z < 1{,}25) = 0{,}894$. Logo:

$$P(Z > 1{,}25) = 1 - P(-\infty < Z < 1{,}25) = 1 - 0{,}894 = 0{,}106$$

Gabarito: D.

Aproximação da Binomial pela Normal

Como consequência do **Teorema Central do Limite**, podemos **aproximar** uma distribuição **binomial** a uma distribuição **normal** Y, quando o número de ensaios $\boldsymbol{n}$ for grande.

Nesse caso, a média e a variância da distribuição são, respectivamente:

$$E(Y) = n.p$$

$$V(Y) = n.p.(1 - p)$$

Assim, podemos aproximar a probabilidade de uma variável **binomial** X apresentar valores dentro de um intervalo $[a, b]$, ou seja, $P(a \leq X \leq b)$, à probabilidade de uma variável **normal** Y apresentar valores no referido intervalo:

$$P(a \leq X \leq b) \cong P(a < Y < b)$$

EXEMPLIFICANDO

Considerando uma variável binomial X, com $n = 50$ e $p = 0{,}5$, vamos calcular a probabilidade associada ao intervalo $[20, 30]$.

Pela distribuição binomial, teríamos que calcular:

$$P(20 \leq X \leq 30) = P(X = 20) + P(X = 21) + \cdots + P(X = 30)$$

$$P(20 \leq X \leq 30) = \binom{50}{20} 0{,}5^{20}. 0{,}5^{30} + \binom{50}{21} 0{,}5^{21}. 0{,}5^{29} + \cdots + \binom{50}{30} 0{,}5^{30}. 0{,}5^{20}$$

Cansativo, não é? O resultado dessa conta é aproximadamente 0,8811 (*confia em mim*).

Aproximando essa distribuição a uma variável Y que segue uma distribuição **normal**, com média $\mu_Y = n.p = 25$ e variância $\sigma_Y^2 = n.p.(1-p) = 12{,}5$ (desvio padrão $\sigma_Y \cong 3{,}5355$), temos a seguinte transformação para $x = 30$:

$$z = \frac{x-\mu}{\sigma} \cong \frac{30-25}{3{,}5255} \cong 1{,}41$$

Pela tabela normal, podemos observar que $P(0 < Z < 1{,}41) = 0{,}4207$.

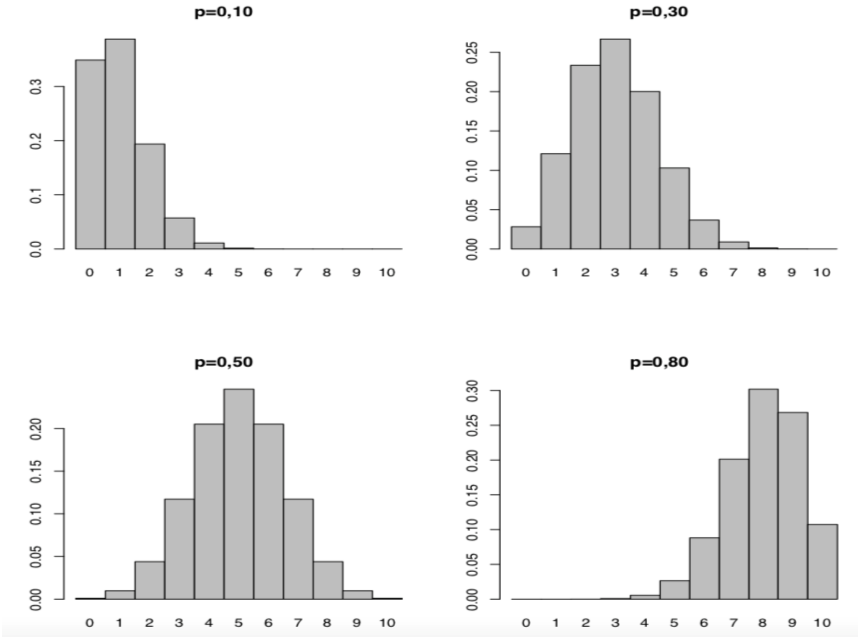
Considerando que o intervalo [20, 30] é simétrico em relação à média $\mu_Y = 25$, então:

$$P(20 \leq X \leq 30) \cong P(-1{,}41 < Z < 1{,}41) = 2 \times 0{,}4207 = 0{,}8414$$

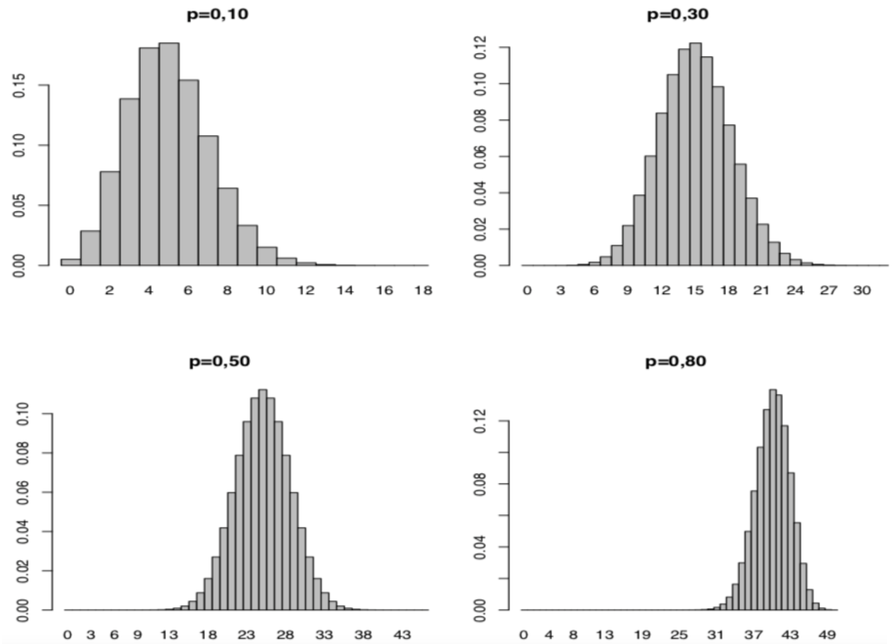
O número exato de ensaios necessários que permite essa aproximação é controverso (e depende do grau de precisão desejado). O que se sabe é que quanto **maior n** e quanto **mais próximo de $^1/_2$ é o valor de p**, mais a distribuição binomial se aproxima da normal.

A seguir, apresento os histogramas[1] de distribuições binomiais, para $n = 10$. Note que o histograma para $p = 0{,}5$ é bem próximo de uma normal, mesmo para um número relativamente pequeno de ensaios (10). Por outro lado, o histograma para $p = 0{,}1$ está bem distante de uma normal.

[1] Histogramas obtidos das aulas de Bacharelado em Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA/USP), disponível em https://www.ime.usp.br/~yambar/MAE0219/Aula%208%20Teorema%20do%20Limite%20Central/Aula%208-Teorema%20do%20Limite%20Central.pdf



Vejamos agora os histogramas para $n = 50$:



Nesse caso, todos os histogramas são próximos de uma curva normal, mesmo para $p = 0{,}1$.

Existe uma regra empírica que exige que o **produto** entre $\boldsymbol{n}$ e a **menor probabilidade entre $\boldsymbol{p}$ e $\boldsymbol{q}$** deve ser **maior que 5**, ou, para uma melhor aproximação, maior que 15.

Na prática, a questão deve fornecer elementos que permitirão concluir que a distribuição binomial pode ser aproximada a uma normal.

Correção de Continuidade

Por se tratar de uma distribuição **discreta** sendo aproximada a uma distribuição **contínua**, para melhorar a precisão dos resultados, é necessário introduzir uma **correção de continuidade**.

Para isso, devemos **aumentar** o intervalo da binomial da forma $[a, b]$, isto é, com os **extremos incluídos**, em **0,5** unidade para cada extremo. Em outras palavras, acrescentamos 0,5 unidade ao extremo superior e subtraímos 0,5 unidade do extremo inferior.

Assim, o intervalo da distribuição normal correspondente será $(a - 0{,}5, b + 0{,}5)$:

$$P(a \leq X \leq b) \cong P(a - 0{,}5 < Y < b + 0{,}5)$$

Por exemplo, o intervalo [2, 4] em uma distribuição binomial (discreta) corresponde ao seguinte intervalo na distribuição normal (contínua):

$$P(2 \leq X \leq 4) \cong P(2 - 0{,}5 < Y < 4 + 0{,}5)$$

$$P(2 \leq X \leq 4) \cong P(1{,}5 < Y < 4{,}5)$$

Se o intervalo desejado **não incluir** um dos extremos, ou ambos, então primeiro **ajustamos** o intervalo para incluí-lo(s).

Por exemplo, vamos calcular a probabilidade de obter **mais de 3** sucessos e **menos de 8** sucessos, o que corresponde ao intervalo (3, 8), isto é, **sem** os extremos.

Para utilizar a aproximação da binomial à normal, primeiro **ajustamos o intervalo** para que os extremos sejam **incluídos**.

Ora, o evento **mais de 3** sucessos e **menos de 8** sucessos, em uma distribuição binomial (discreta), equivale ao evento **4 ou mais** sucessos e **7 ou menos** sucessos, o que corresponde ao intervalo [4, 7], isto é, com os extremos incluídos.

$$P(3 < X < 8) = P(4 \leq X \leq 7)$$

Após o ajuste do intervalo, podemos aplicar a **correção de continuidade**, somando-se 0,5 unidade ao extremo superior e subtraindo 0,5 unidade do extremo inferior:

$$P(4 \leq X \leq 7) \cong P(4 - 0{,}5 < Y < 7 + 0{,}5)$$

$$P(4 \leq X \leq 7) \cong P(3{,}5 < Y < 7{,}5)$$

EXEMPLIFICANDO

Vamos, então, refazer o nosso exemplo com a correção de continuidade:

$$P(20 \leq X \leq 30) \cong P(19{,}5 < Y < 30{,}5)$$

Para $x = 30{,}5$, com $\mu = 25$ e $\sigma \cong 3{,}5355$, temos a seguinte transformação:

$$z = \frac{x-\mu}{\sigma} \cong \frac{30{,}5-25}{3{,}5255} \cong 1{,}56$$

Pela tabela normal, podemos observar que $P(0 < Z < 1{,}56) = 0{,}4406$. Logo:

$$P(20 \leq X \leq 30) \cong P(-1{,}56 < Z < 1{,}56) = 2 \times 0{,}4406 = 0{,}8812$$

Note que esse resultado é extremamente próximo do resultado exato, qual seja, 0,8811!

Logo, para resolver questões envolvendo a aproximação de uma distribuição binomial X a uma distribuição normal Y, precisamos:

i. Calcular a média e a variância da distribuição, $E(Y) = E(X) = n.p$ e $V(Y) = V(X) = n.p.q$
ii. Aplicar a correção de continuidade $P(a \leq X \leq b) \cong P(a - 0{,}5 < Y < b + 0{,}5)$
iii. Utilizar a transformação para a normal padrão $z = \frac{x-\mu}{\sigma}$
iv. Consultar a tabela da normal padrão para encontrar os valores de probabilidade

HORA DE
PRATICAR!

(CESPE/2013 – TRT 17ª Região) No que se refere a distribuições discretas, julgue o seguinte item.

A aproximação da distribuição binomial pela normal não se aplica com base no teorema limite central, visto que a binomial não se relaciona com uma soma de variáveis aleatórias.

Comentários:

O Teorema Central do Limite garante que, para n suficientemente grande, a distribuição binomial pode sim ser aproximada pela distribuição Normal.

Pontue-se, ainda, que a distribuição binomial é uma soma de variáveis de Bernoulli.

Gabarito: Errado.

(FGV/2017 – MPE/BA) A probabilidade de que uma decisão de 1ª instância da Justiça Federal do Paraná seja reformada pelo Tribunal Superior da 4ª Região é de 0,20. No momento 100 recursos aguardam por uma decisão dos Srs. Desembargadores daquele Tribunal.

São informados alguns valores da distribuição acumulada da normal-padrão:

Ø(1) = 0,87, Ø(1,28)=0,90 e Ø(2) = 98

Sem usar o ajuste de continuidade, a probabilidade de que mais de 24 decisões sejam reformadas é:

a) 13%

b) 10%

c) 8%

d) 5%

e) 2%

Comentários:

A questão trabalha com a aproximação da binomial à normal. Sabendo que a probabilidade de reforma de decisão é p = 0,20 e que há n = 100 recursos, então a média da distribuição é:

$$\mu = n \times p = 100 \times 0{,}2 = 20$$

Sendo a probabilidade de não reforma de decisão complementar q = 1 – p = 0,9, a variância é:

$$V(X) = n.p.q = 100 \times 0{,}2 \times 0{,}8 = 16$$

E o desvio padrão:

$$\sigma = \sqrt{V(X)} = \sqrt{16} = 4$$

Considerando que não deve ser utilizada a correção de continuidade, então a transformação de x = 24 para a normal padrão é:

$$z = \frac{x - \mu}{\sigma} = \frac{24 - 20}{4} = \frac{4}{4} = 1$$

Pelas informações fornecidas, observamos que a probabilidade acumulada é $P(Z \leq 1) = 0{,}87$. Logo, a probabilidade de haver mais decisões reformadas é complementar:

$$P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0{,}87 = 0{,}13 = 13\%$$

Gabarito: A.

DISTRIBUIÇÃO QUI-QUADRADO

A distribuição qui-quadrado resulta da **soma** de distribuições **normais reduzidas** (ou padrão) **independentes**, elevadas ao **quadrado**.

Em outras palavras, a distribuição qui-quadrado $\mathcal{X}_k^2$ corresponde à soma dos quadrados de k variáveis com distribuição normal padrão, $Z_i \sim N(0,1)$:

$$\mathcal{X}_k^2 = \sum_{i=1}^{k} (Z_i)^2$$

Dizemos que a distribuição apresenta ***k* graus de liberdade**, sendo ***k*** o número de variáveis **normais** que compõem a distribuição qui-quadrado. O número de graus de liberdade é o **único parâmetro** da distribuição.

Em particular, para $k = 1$, isto é, havendo uma única variável normal padrão elevada ao quadrado Z^2, temos uma distribuição qui-quadrado com **um** grau de liberdade, $\mathcal{X}_1^2$.

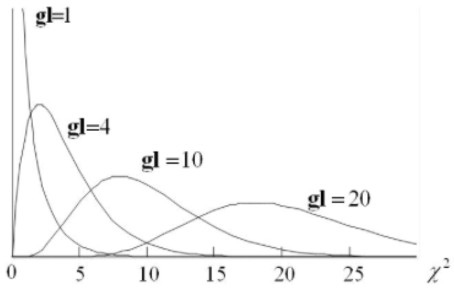
Pontue-se que, se as variáveis normais X_i **não forem reduzidas**, ou seja, se apresentarem média μ_i e desvio padrão σ_i, é necessário aplicar a **transformação** para a normal reduzida, antes de elevar ao quadrado, para formar a distribuição qui-quadrado padrão:

$$\mathcal{X}_k^2 = \sum_{i=1}^{k} (Z_i)^2 = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

Quando distribuições normais com média $\mu \neq 0$ **não padronizadas** são utilizadas para formar uma distribuição qui-quadrada, dizemos que a distribuição resultante é **não central**, com **parâmetro de não centralidade** igual à **média** μ das variáveis normais.

Para a distribuição qui-quadrado, os diferentes graus de liberdade reproduzem funções densidade de probabilidade distintas, conforme ilustrado a seguir[1].

[1] Gráfico do professor Ivan Balducci da Universidade Estadual Paulista, FOSJC/UNESP, disponível em https://slideplayer.com.br/slide/45581/.



Podemos observar que as variáveis com distribuição qui-quadrado são **positivas** (afinal correspondem à soma do **quadrado** de variáveis) e **assimétricas à direita**.

Observamos, ainda, que quanto **maior o grau de liberdade**, isto é, quanto mais variáveis normais ao quadrado são somadas, mais **simétrica** será a distribuição.

Inclusive, pelo **Teorema Central do Limite**, a distribuição qui-quadrado se **aproxima a uma distribuição normal**, à medida que o número de graus de liberdade $\boldsymbol{k}$ aumenta.

A **soma** de variáveis com distribuição qui-quadrado com $k_1, k_2, \dots, k_n$ graus de liberdade, segue uma distribuição **qui-quadrado** com $k_1 + k_2 + \cdots + k_n$ graus de liberdade.

A **média** da distribuição qui-quadrado é igual ao número de graus de liberdade $\boldsymbol{k}$; e a **variância** é igual ao dobro do número de graus de liberdade $\boldsymbol{2k}$:

$$\boldsymbol{E(\mathcal{X}_k^2) = k}$$

$$\boldsymbol{V(\mathcal{X}_k^2) = 2k}$$

A distribuição qui-quadrado também possui uma **tabela** que associa uma probabilidade $P(\mathcal{X}_k^2 < x)$ ou $P(\mathcal{X}_k^2 > x)$ a um valor positivo x, de acordo com o grau de liberdade k.

Pelo fato de a distribuição **não** ser **simétrica**, a distribuição acima da média é diferente daquela abaixo da média. Por isso, a tabela apresenta os valores de probabilidade para **toda** a distribuição.

A tabela a seguir apresenta os valores para $P(\mathcal{X}_k^2 < x)$, indicados na primeira linha. A primeira (assim como a última) coluna apresenta os graus de liberdade k, que a tabela indica como n.

Por exemplo, a mediana para uma distribuição com 5 graus de liberdade, isto é, o valor de x para o qual $P(\mathcal{X}_5^2 < x) = 50\%$, conforme se observa na linha $n = 5$ e na coluna $P = 0{,}5$, é $x = 4{,}351$.

	$P(\chi^2_n \leq x)$													
n	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	
1	3,93E-05	0,000157	0,000982	0,003932	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	3
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	4
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,832	15,086	16,750	5
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	6
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	7
8	1,344	1,647	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	8
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589	9
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188	10
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757	11
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217	28,300	12
13	3,565	4,107	5,009	5,892	7,041	9,299	12,340	15,984	19,812	22,362	24,736	27,688	29,819	13
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141	31,319	14
15	4,601	5,229	6,262	7,261	8,547	11,037	14,339	18,245	22,307	24,996	27,488	30,578	32,801	15
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000	34,267	16
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409	35,718	17
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805	37,156	18
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191	38,582	19
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566	39,997	20
21	8,034	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932	41,401	21
22	8,643	9,542	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289	42,796	22
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638	44,181	23
24	9,886	10,856	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,980	45,558	24
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314	46,928	25
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336	30,435	35,563	38,885	41,923	45,642	48,290	26
27	11,808	12,878	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,195	46,963	49,645	27
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278	50,994	28
29	13,121	14,256	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588	52,335	29
30	13,787	14,953	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892	53,672	30
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691	66,766	40
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154	79,490	50
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379	91,952	60
70	43,275	45,442	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	100,425	104,215	70
80	51,172	53,540	57,153	60,391	64,278	71,145	79,334	88,130	96,578	101,879	106,629	112,329	116,321	80
90	59,196	61,754	65,647	69,126	73,291	80,625	89,334	98,650	107,565	113,145	118,136	124,116	128,299	90
100	67,328	70,065	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	135,807	140,170	100

A distribuição qui-quadrado é um **caso particular** da distribuição **gama**, cuja f.d.p. é:

$$f(x) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, & se\ x \geq 0 \\ 0, & caso\ contrário \end{cases}$$

A f.d.p. da distribuição **qui-quadrado** é obtida com $\boldsymbol{\alpha = \frac{k}{2}}$ e $\boldsymbol{\beta = \frac{1}{2}}$.

Por exemplo, para $\boldsymbol{k = 2}$, temos uma distribuição **qui-quadrado** com 2 graus de liberdade, o que corresponde a uma distribuição **gama** com $\boldsymbol{\alpha = 1}$ e $\boldsymbol{\beta = \frac{1}{2}}$.

Nessa situação, em que $\alpha = 1$, a distribuição gama se reduz a uma distribuição **exponencial**, com $\beta = \lambda$. Então, uma distribuição **qui-quadrado** com **2 graus de liberdade** é também uma distribuição **exponencial**, com $\boldsymbol{\lambda = \beta = \frac{1}{2}}$.

A **média** dessa distribuição é $E(X) = \frac{1}{\lambda} = \mathbf{2}$ ou $E(\mathcal{X}_k^2) = k = \mathbf{2}$.

ESQUEMATIZANDO

Distribuição Qui-quadrado com $\boldsymbol{k}$ **graus de liberdade**	
Soma de $\boldsymbol{k}$ variáveis **normais padrão** elevadas ao quadrado: $\mathcal{X}_k^2 = \sum_{i=1}^{k}(Z_i)^2$	
Assimétrica à direita, com $x \geq 0$	
Esperança: $\boldsymbol{E(\mathcal{X}_k^2) = k}$;	**Variância**: $\boldsymbol{V(\mathcal{X}_k^2) = 2k}$

HORA DE PRATICAR!

(FCC/2015 – Analista do CNMP – *Adaptada*) Sejam X_1, X_2, ... , X_n variáveis aleatórias independentes e identicamente distribuídas com distribuição normal padrão e seja a variável aleatória $W = X_1^2 + X_2^2 + \cdots + X_n^2$. Considerando essa informação, julgue a seguinte afirmativa:

A variável W tem distribuição qui-quadrado com (n − 1) graus de liberdade.

Comentários:

Sendo $W = X_1^2 + X_2^2 + \cdots + X_n^2$, então W apresenta distribuição qui-quadrado com k = **n** graus de liberdade (não, n – 1).

Resposta: Errado.

(FGV/2021 – FunSaúde/CE) Se X_1, X_2,...X_n são n variáveis aleatórias independentes e identicamente distribuídas com distribuição N(μ,σ²), então a variável

$$Q = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

Tem a seguinte distribuição de probabilidades:

a) normal com média $n\mu$ e variância $n\sigma^2$.

b) qui-quadrado com n graus de liberdade

c) qui-quadrado com n-1 graus de liberdade.

d) t-Student com n graus de liberdade.

e) t-Student com n-1 graus de liberdade

Comentários:

A variável Q indicada no enunciado corresponde à soma de n variáveis normais, reduzidas à normal padrão, $Z_i = \frac{X_i - \mu_i}{\sigma_i}$, elevadas ao quadrado.

Assim, a variável Q apresenta distribuição qui-quadrado com k = n graus de liberdade.

Gabarito: B

(CESPE/2018 – Agente de Polícia Federal) Uma amostra aleatória simples Y_1, Y_2, ..., Y_{25} foi retirada de uma distribuição normal com média nula e variância σ^2, desconhecida. Considerando que $P(x^2 \leq 13) = P(x^2 > 41) = 0,025$, em que x^2 representa a distribuição qui-quadrado com 25 graus de liberdade, e que $S^2 = \sum_{i=1}^{25} Y_i^2$, julgue o item a seguir.

A variância da distribuição X^2 com 25 graus de liberdade é superior a 40.

Comentários:

A variância da distribuição qui-quadrado, com k = 25 graus de liberdade, é:

$$V(X) = 2k = 2 \times 25 = 50$$

Assim, a variância é superior a 40.

Gabarito: Certo.

(CESPE/2018 – Analista Administrativo – EBSERH) Considerando que X e Y sejam variáveis aleatórios mutuamente independentes que seguem distribuição normal padrão, julgue o próximo item.

A soma dos quadrados $Q = X^2 + Y^2$ segue uma distribuição exponencial com média igual a 2.

Comentários:

A soma de k variáveis independentes com distribuição normal padrão elevadas ao quadrado segue distribuição **qui-quadrado**, com k graus de liberdade. Nesse caso, temos uma variável com distribuição qui-quadrado com k = 2. A média dessa distribuição é:

$$E(X) = k = 2$$

Especialmente para k = 2, a distribuição qui-quadrado corresponde a uma distribuição exponencial, com $\lambda = \frac{1}{2}$.

Gabarito: Certo.

DISTRIBUIÇÃO T DE STUDENT

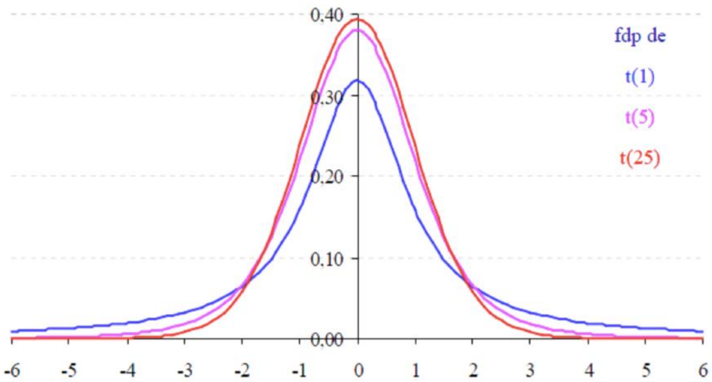
A variável T com distribuição ***t* de Student** (ou **t-Student**) é definida como:

$$T = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}}$$

em que Z é uma variável **normal padrão**, $\mathcal{X}_k^2$ é uma variável com distribuição **qui-quadrado**, com $\boldsymbol{k}$ graus de liberdade, sendo Z e $\mathcal{X}_k^2$ **independentes**.

A variável T apresenta **$\boldsymbol{k}$ graus de liberdade**, que também é o **único parâmetro** da distribuição.

O gráfico a seguir[1] apresenta as funções densidade de probabilidade para distribuições t-Student com 1, 5 e 25 graus de liberdade.



Observa-se que, assim como a normal padrão, a distribuição t-Student é **simétrica**, com média $\boldsymbol{\mu = 0}$ e formato de **sino**, assumindo valores em toda a reta real, $(-\infty, \infty)$.

Porém, em comparação com a normal, a curva **t-Student** é mais **baixa** na região **central** e mais **larga** nos **extremos**. Ou seja, a distribuição t-Student apresenta **maior variabilidade**.

A **variância** da distribuição de t-Student é dada por:

$$V(T) = \frac{k}{k-2},\ para\ k > 2$$

Pontue-se que a variância **não** é definida um grau de liberdade menor ou igual a 2 ($\boldsymbol{k \leq 2}$).

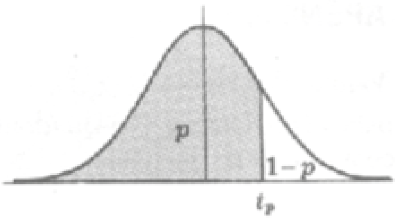
---

[1] Obtido na apresentação das aulas da Prof. Tarciana Liberal, da Universidade Federal da Paraíba, disponível em http://www.de.ufpb.br/~tarciana/Probabilidade2/Aula15.pdf

A **variância da distribuição** de t-Student é sempre **maior que 1**, para qualquer valor de k, pois a razão $\frac{k}{k-2}$ é sempre um valor **superior a 1**. Porém, à medida que k aumenta, essa razão se **reduz**, **aproximando-se de 1**.

Assim, à medida que o grau de liberdade ***k* aumenta**, a distribuição de t-Student se **aproxima** de uma curva **normal**, com **variância igual a 1** (outra consequência do **Teorema Central do Limite**).

A distribuição t-Student também apresenta uma **tabela** que associa aos valores de probabilidade os intervalos de valores de t, de acordo com o grau de liberdade, conforme indicado a seguir. Observe pelo gráfico anterior a essa tabela, que os seus valores de probabilidade são da forma $P(T < t)$.



ν	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,75}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	,158	,325	,727	1,000	1,376	3,08	6,31	12,71	31,82	63,66
2	,142	,289	,617	,816	1,061	1,89	2,92	4,30	6,96	9,92
3	,137	,277	,584	,765	,978	1,64	2,35	3,18	4,54	5,84
4	,134	,271	,569	,741	,941	1,53	2,13	2,78	3,75	4,60
5	,132	,267	,559	,727	,920	1,48	2,02	2,57	3,36	4,03
6	,131	,265	,553	,718	,906	1,44	1,94	2,45	3,14	3,71
7	,130	,263	,549	,711	,896	1,42	1,90	2,36	3,00	3,50
8	,130	,262	,546	,706	,889	1,40	1,86	2,31	2,90	3,36
9	,129	,261	,543	,703	,883	1,38	1,83	2,26	2,82	3,25
10	,129	,260	,542	,700	,879	1,37	1,81	2,23	2,76	3,17
11	,129	,260	,540	,697	,876	1,36	1,80	2,20	2,72	3,11
12	,128	,259	,539	,695	,873	1,36	1,78	2,18	2,68	3,06
13	,128	,259	,538	,694	,870	1,35	1,77	2,16	2,65	3,01
14	,128	,258	,537	,692	,868	1,34	1,76	2,14	2,62	2,98
15	,128	,258	,536	,691	,866	1,34	1,75	2,13	2,60	2,95
16	,128	,258	,535	,690	,865	1,34	1,75	2,12	2,58	2,92
17	,128	,257	,534	,689	,863	1,33	1,74	2,11	2,57	2,90
18	,127	,257	,534	,688	,862	1,33	1,73	2,10	2,55	2,88
19	,127	,257	,533	,688	,861	1,33	1,73	2,09	2,54	2,86
20	,127	,257	,533	,687	,860	1,32	1,72	2,09	2,53	2,84
21	,127	,257	,532	,686	,859	1,32	1,72	2,08	2,52	2,83
22	,127	,256	,532	,686	,858	1,32	1,72	2,07	2,51	2,82
23	,127	,256	,532	,685	,858	1,32	1,71	2,07	2,50	2,81
24	,127	,256	,531	,685	,857	1,32	1,71	2,06	2,49	2,80
25	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,79
26	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,78
27	,127	,256	,531	,684	,855	1,31	1,70	2,05	2,47	2,77
28	,127	,256	,530	,683	,855	1,31	1,70	2,05	2,47	2,76
29	,127	,256	,530	,683	,854	1,31	1,70	2,04	2,46	2,76
30	,127	,256	,530	,683	,854	1,31	1,70	2,04	2,46	2,75
40	,126	,255	,529	,681	,851	1,30	1,68	2,02	2,42	2,70
60	,126	,254	,527	,679	,848	1,30	1,67	2,00	2,39	2,66
120	,126	,254	,526	,677	,845	1,29	1,66	1,98	2,36	2,62
∞	,126	,253	,524	,674	,842	1,28	1,645	1,96	2,33	2,58

A primeira coluna apresenta os graus de liberdade, que a tabela chama de ν, e a primeira linha apresenta os valores de probabilidade acumulada, até os valores de t indicados nos campos. Por exemplo, para 5 graus de liberdade (5ª linha), o valor que delimita uma probabilidade $P(T < t) = 0{,}9$ (6ª coluna), é $t = 1{,}48$.

Lembre-se que a distribuição é **simétrica**, com média, mediana e moda iguais a $\mu = 0$. Assim:

$$P(T < -t) = P(T > t)$$

$$P(T < 0) = P(T > 0) = 0{,}5$$

Por isso, são apresentados apenas valores de probabilidades $P(T < t)$ maiores que 50%, uma vez que à esquerda da média, os valores são **simétricos**.

Assim como para a curva normal padrão, há outras formas de apresentar a tabela de t-Student.

INDO MAIS FUNDO!

Ressalte-se que a distribuição t-Student para $k = 1$ consiste na **razão entre duas variáveis normais padrão** independentes:

$$T = \frac{Z_1}{\sqrt{\frac{\chi_1^2}{1}}} = \frac{Z_1}{\sqrt{Z_2^2}} = \frac{Z_1}{Z_2}$$

Nesse caso específico, temos uma distribuição de **Cauchy**, que corresponde à **razão entre duas variáveis normais** independentes.

Essa distribuição também é **simétrica** e **mais achatada** em relação à normal. Porém, ela é uma distribuição "**patológica**", pois **não** possui **média** ou **variância** definidas.

ESQUEMATIZANDO

Distribuição t-Student, definida para todos os valores reais

$$T = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}}$$

Simétrica, com $\mu = 0$ e formato de **sino mais achatado** do que a normal

Variância: $V(X) = \frac{k}{k-2},\ para\ k > 2$

(2017 – TRF 2ª Região – *Adaptada*) Sobre a distribuição t-Student, julgue a afirmativa a seguir.

Considere uma variável aleatória Z com distribuição normal padrão e uma outra variável aleatória V com distribuição qui-quadrado e v graus de liberdade. Se Z e V forem independentes, então a variável aleatória $T = \frac{Z}{\sqrt{\frac{V}{v}}}$ tem distribuição t de *Student* com v graus de liberdade.

Comentários:

Vimos que a distribuição t-Student é definida como:

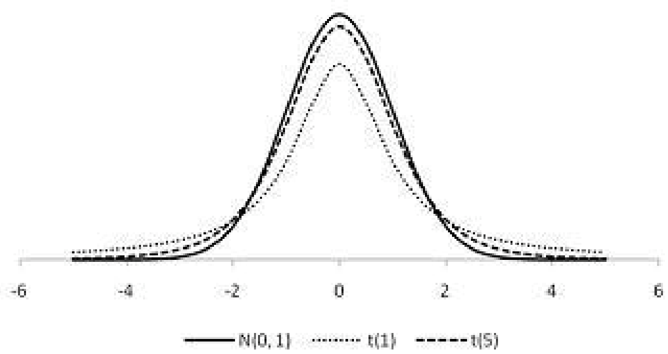
$$T = \frac{Z}{\sqrt{\frac{\mathcal{X}_k^2}{k}}}$$

Então, sendo V uma distribuição qui-quadrado com v graus de liberdade, temos:

$$T = \frac{Z}{\sqrt{\frac{V}{v}}}$$

Resposta: Certo.

(CESPE/2011 – Analista Judiciário do TJ/ES)



O gráfico acima mostra a função de densidade da distribuição normal padrão N(0, 1) e t(1) e t(5), que representam, respectivamente, as densidades da distribuição t de Student com 1 e 5 graus de liberdade.

Com base nesse gráfico, julgue o próximo item.

A distribuição N(0, 1) possui variância unitária, a t(5) possui variância igual a 5/3, e a variância da distribuição t-Student com 1 grau de liberdade é indefinida.

Comentários:

Vamos analisar as três afirmações.

Em relação à primeira afirmação, de fato, a distribuição normal padrão N(0,1) possui variância igual a 1 (unitária).

Em relação à segunda afirmação, a distribuição de t-Student com k = 5 graus de liberdade é:

$$V(X) = \frac{k}{k-2} = \frac{5}{3}$$

Em relação à terceira afirmação, a distribuição de t-Student com k = 1 grau de liberdade **não** tem variância definida.

Gabarito: Certo.

(FGV/2010 – FIOCRUZ) Duas variáveis aleatórias independentes X e Y são tais que X tem distribuição Normal com média 0 e variância 4 e Y pode ser escrita como $Y = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2$, em que os Z_i são independentes e identicamente distribuídos com distribuição normal padrão, i = 1, 2, 3, 4. Nesse caso, a seguinte variável tem distribuição t- Student

a) XY

b) $XY^{-0,5}$

c) $X^{-1}Y$

d) $4XY^{-0,5}$

e) 2XY

Comentários:

A variável t-Student é definida como:

$$T = \frac{Z}{\sqrt{\frac{\mathcal{X}_k^2}{k}}}$$

Sendo X uma variável normal com média $\mu = 0$ e variância V(X) = 4, para transformá-la na normal padrão precisamos dividir pelo desvio padrão $\sigma = \sqrt{V(X)} = 2$:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-0}{2} = \frac{X}{2}$$

Sendo Y a soma de 4 variáveis normais padrão independentes elevadas ao quadrado, então Y é uma variável qui-quadrado $\mathcal{X}_k^2$ com k = 4 graus de liberdade.

Sendo as variáveis independentes, então a seguinte razão apresenta distribuição de t-Student:

$$T = \frac{Z}{\sqrt{\frac{\mathcal{X}_k^2}{k}}} = \frac{\frac{X}{2}}{\sqrt{\frac{Y}{4}}} = \frac{\frac{X}{2}}{\frac{\sqrt{Y}}{2}} = \frac{X}{2} \times \frac{2}{\sqrt{Y}} = \frac{X}{\sqrt{Y}}$$

Essa razão pode ser representada como:

$$T = \frac{X}{Y^{\frac{1}{2}}} = X \times Y^{-\frac{1}{2}} = X \times Y^{-0,5}$$

Gabarito: B

(CESPE/2020 – Analista Judiciário do TJ/PA) Se X e Y são variáveis aleatórias normais independentes, tais que $X \sim N(0,1)$ e $Y \sim N(0,1)$, a razão $\frac{X}{Y}$ segue uma distribuição

a) de Cauchy

b) de Pareto

c) de Weibull

d) t de Student com 2 graus de liberdade

e) normal padrão

Comentários:

A razão entre variáveis independentes com distribuição normal segue uma distribuição de Cauchy, que corresponde à razão entre duas variáveis normais.

Note que X e Y são variáveis normais padrão. Nesse caso, temos também uma distribuição t de Student, porém com 1 grau de liberdade, e não 2, como descrito na alternativa D.

Gabarito: A.

DISTRIBUIÇÃO AMOSTRAL

Nesta seção, estudaremos a distribuição de probabilidade dos principais estimadores, chamada de **Distribuição Amostral**. Vale ressaltar que os **estimadores** são **variáveis aleatórias** e por isso apresentam distribuições de probabilidade.

Inicialmente, é importante pontuar que a distribuição dos elementos de uma **amostra** aleatória qualquer segue a **mesma** distribuição **populacional**.

Por exemplo, vamos considerar uma moeda com 2 faces, que vamos chamar de face 0 e face 1. Tratando-se de uma moeda equilibrada, a probabilidade de cada face é de $\frac{1}{2} = 0{,}5$ e a esperança e variância são, respectivamente:

$$E(X) = \sum x.P(X = x)$$

$$E(X) = 0 \times 0{,}5 + 1 \times 0{,}5 = 0{,}5$$

$$V(X) = \sum (x - \mu)^2.P(X = x)$$

$$V(X) = (0 - 0{,}5)^2 \times \frac{1}{2} + (1 - 0{,}5)^2 \times \frac{1}{2} = \frac{0{,}25 + 0{,}25}{2} = 0{,}25$$

Agora, suponha que vamos extrair uma **amostra** aleatórias de tamanho 3, ou seja, vamos lançar a moeda 3 vezes. Podemos representar essa amostra por X_1, X_2, X_3.

Considerando que os possíveis resultados das amostras são os mesmos da população (0 ou 1) e com as mesmas probabilidades de $\frac{1}{2} = 0{,}5$ para cada face, então temos:

$$E(X_1) = E(X_2) = E(X_3) = 0 \times 0{,}5 + 1 \times 0{,}5 = 0{,}5$$

$$V(X_1) = V(X_2) = V(X_3) = (0 - 0{,}5)^2 \times \frac{1}{2} + (1 - 0{,}5)^2 \times \frac{1}{2} = \frac{0{,}25 + 0{,}25}{2} = 0{,}25$$

Ou seja, a esperança e a variância de cada amostra são **iguais** às da população:

$$E(X_i) = E(X) = \mu$$

$$V(X_i) = V(X) = \sigma^2$$

Na verdade, toda a distribuição de probabilidade da **amostra** é **igual** à distribuição de probabilidade da **população** (as esperanças e as variâncias são iguais como consequência desse fato).

Quando as variáveis $X_1, X_2, X_3, \dots X_n$, que representam os elementos da **amostra**, são **independentes**, dizemos que elas **independentes e identicamente distribuídas** (**i.i.d.**), isto é, são independentes e apresentam a mesma distribuição (igual à da população).

Isso ocorre quando a **população** é **infinita** (ou muito grande em comparação com o tamanho da amostra) **ou** quando a **amostra** é extraída **com reposição**.

(FGV/2021 – FunSaúde/CE - *Adaptada*) Se X_1, X_2, ... X_n é uma amostra aleatória simples extraída de uma população infinita com determinada distribuição de probabilidades f(x), avalie se as afirmativas a seguir estão corretas.

I. X_1, X_2, ... X_n são independentes.

II. X_1, X_2, ... X_n são identicamente distribuídos.

III. Nem sempre cada X_i, i = 1,..., n, tem distribuição f(x).

Está correto o que se afirma em

a) I, apenas.

b) I e II, apenas.

c) I e III, apenas.

d) II e III, apenas.

e) I, II e III.

Comentários:

Essa questão trabalha com os conceitos fundamentais envolvendo distribuição amostral.

Sendo a população infinita, então os elementos da amostra (que o enunciado chamou de X_1, X_2, ... X_n) são variáveis aleatórias **independentes** que apresentam a **mesma distribuição** da população.

Assim, as afirmativas I e II estão corretas, enquanto a afirmativa III está incorreta, pois cada variável X_i apresenta a mesma distribuição f(x) da população (sempre).

Gabarito: B

Agora, estudaremos a distribuição dos estimadores mais utilizados, quais sejam, a **média**, a **proporção** e a **variância** amostrais.

Distribuição Amostral da Média

Para estimarmos a **média da população μ**, utilizamos como estimador a **média amostral $\bar{X}$**.

Sendo $X_1, X_2, \ldots, X_n$ os valores observados da amostra, a média amostral é a **razão** entre a soma dos valores observados, $X_1 + X_2 + \cdots + X_n$, e o número de elementos observados, n:

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

Assim como os demais estimadores, a média amostral é uma **variável aleatória**, uma vez que $\bar{X}$ **varia** de acordo com os valores observados da amostra $X_1, X_2, \ldots, X_n$.

Vamos ao exemplo da moeda lançada 3 vezes. Se o resultado for {0, 0, 1}, a média amostral será:

$$\bar{X} = \frac{0+0+1}{3} = \frac{1}{3} \cong 0{,}33$$

Se o resultado for {0, 1, 1}, por exemplo, a média amostral será:

$$\bar{X} = \frac{0+1+1}{3} = \frac{2}{3} \cong 0{,}67$$

*E qual seria a esperança desse estimador? Bem, sabendo que as faces possíveis são 0 e 1, cada uma com 50% de chance, esperamos que as **médias** desse experimento estejam em torno de 0,5.*

Ou seja, a **esperança da média amostral** é igual à **média populacional**?

$$E(\bar{X}) = \mu$$

E quanto à variância? A **variância da média amostral** é dada por:

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

Para o exemplo dos 3 lançamentos da moeda, em que V(X) = 0,25 e n = 3, a variância da média amostral é:

$$V(\bar{X}) = \frac{0{,}25}{3} \cong 0{,}08$$

A **variância** da **média amostral $\boldsymbol{V(\bar{X})}$** é **menor** do que a **variância populacional $\boldsymbol{V(X)}$**.

ESCLARECENDO!

Isso ocorre porque a **média** das observações $\bar{X}$ tende a ser um valor bem mais **próximo** da média populacional μ do que as observações **individuais** da amostra X_i. Afinal, no cálculo da média amostral, os valores acima da média **compensam** os valores abaixo da média.

Em outras palavras, as médias amostrais $\bar{X}$ **variam menos** do que as observações individuais X_i. Lembrando que as observações individuais da amostra seguem a mesma distribuição da população, então concluímos que a **variância da média amostral** $V(\bar{X})$ é **menor** do que a **variância da população $\boldsymbol{V(X)}$**.

Além disso, quanto **maior o tamanho da amostra $\boldsymbol{n}$**, **menor será a variância da média amostral**.

O **desvio padrão** (raiz quadrada da variância) de um estimador pode ser chamado de **erro padrão**.

O **erro padrão** (ou **desvio padrão**) da **média amostral** é dado por:

$$EP(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Ou seja, o erro padrão (ou desvio padrão) da média amostral pode ser calculado como a **raiz quadrada** da **variância da média amostral $\boldsymbol{\sqrt{V(\bar{X})}}$**, ou como a **razão** entre o **desvio padrão populacional** e a **raiz do número de elementos da amostra $\boldsymbol{\frac{\sigma}{\sqrt{n}}}$.**

Também podemos denotar o erro padrão (ou desvio padrão) da média amostral por $\boldsymbol{\sigma_{\bar{X}}}$.

Para o nosso exemplo, o erro padrão da média amostral pode ser calculado como:

$$EP(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{0{,}25}{3}} \cong 0{,}29$$

As fórmulas da esperança e variância podem ser obtidas pelas respectivas **propriedades**.

Sendo $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$, a **esperança** $E(\bar{X})$ é dada por:

$$E(\bar{X}) = E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = E\left(\frac{X_1}{n}\right) + E\left(\frac{X_2}{n}\right) + \cdots + E\left(\frac{X_n}{n}\right)$$

$$E(\bar{X}) = \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n)$$

Vimos que a esperança amostral é igual à esperança populacional, $E(X_i) = E(X) = \mu$:

$$E(\bar{X}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = n \times \frac{1}{n} \times \mu = \mu$$

Considerando que as variáveis $X_1, X_2, \dots, X_n$ são **independentes**, a **variância** $V(\bar{X})$ é:

$$V(\bar{X}) = V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = V\left(\frac{X_1}{n}\right) + V\left(\frac{X_2}{n}\right) + \cdots + V\left(\frac{X_n}{n}\right)$$

$$V(\bar{X}) = \frac{1}{n^2}V(X_1) + \frac{1}{n^2}V(X_2) + \cdots + \frac{1}{n^2}V(X_n)$$

Sabendo que a variância amostral é igual à variância populacional, $V(X_i) = V(X) = \sigma^2$:

$$\boldsymbol{V(\bar{X})} = \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 = n \times \frac{1}{n^2} \times \sigma^2 = \boldsymbol{\frac{\sigma^2}{n}}$$

A **variância** dos **elementos da amostra** X_i é **igual** à **variância populacional** $V(X_i) = \sigma^2$. O que é **diferente** é a **variância** da **média amostral**, $V(\bar{X}) = \frac{\sigma^2}{n}$.

Se a variância da população **não** for conhecida, ela precisa ser estimada a partir da amostra (variância amostral):

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Nessa situação, a estimativa para a **variância da média amostral** será:

$$V(\bar{X}) = \frac{s^2}{n}$$

(CESPE/2016 – TCE/PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal. Tendo essa informação como referência inicial, julgue o seguinte item.

Para essa amostra aleatória simples, o valor esperado da média amostral é igual à média populacional.

Comentários:

De fato, o **valor esperado da média amostral** (para uma amostra aleatória) é igual à **média populacional**.

Gabarito: Certo.

(2019 – UEPA) Considere uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população normal de média μ e variância σ² = 9. Então, a média e a variância de $\overline{X_n} = \frac{X_1, X_2, \ldots, X_n}{n}$, são, respectivamente,

a) μ e $\frac{3}{n}$.

b) $\frac{\mu}{n}$ e $\frac{9}{n}$.

c) μ e $\frac{9}{n}$.

d) μ e $\frac{n}{9}$.

Comentários:

A **média** (ou esperança) e a **variância** da média amostral são, respectivamente:

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{9}{n}$$

Gabarito: C.

(VUNESP/2015 – TJ-SP) Resultados de uma pesquisa declaram que o desvio padrão da média amostral é 32. Sabendo que o desvio padrão populacional é 192, então o tamanho da amostra que foi utilizada no estudo foi

a) 6.

b) 25.

c) 36.

d) 49.

e) 70.

Comentários:

O desvio padrão (ou erro padrão) da média amostral pode ser calculado como a razão entre o desvio padrão da população e a raiz do número de elementos da amostra:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

O enunciado informa que o desvio padrão da média amostral é $\sigma_{\bar{X}} = 32$ e o desvio padrão populacional é $\sigma = 192$. Logo:

$$32 = \frac{192}{\sqrt{n}}$$

$$\sqrt{n} = \frac{192}{32} = 6$$

$$n = (6)^2 = 36$$

Gabarito: C.

(FCC/2012 – TRE-SP) Uma variável aleatória U tem distribuição uniforme contínua no intervalo $[\boldsymbol{\alpha}, \mathbf{3}\boldsymbol{\alpha}]$. Sabe-se que U tem média 12. Uma amostra aleatória simples de tamanho n, com reposição, é selecionada da distribuição de U e sabe-se que a variância da média dessa amostra é 0,1. Nessas condições, o valor de n é

a) 80.

b) 100.

c) 120.

d) 140.

e) 150.

Comentários:

O que essa questão exige, em relação à matéria que acabamos de estudar, é a fórmula do desvio padrão da média amostral:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Assim, podemos escrever o tamanho amostral como:

$$\sqrt{n} = \frac{\sigma}{\sigma_{\bar{X}}}$$

$$n = \left(\frac{\sigma}{\sigma_{\bar{X}}}\right)^2 = \frac{\sigma^2}{\sigma_{\bar{X}}^2}$$

Ou seja, o tamanho amostral é a razão entre a variância populacional σ^2 (quadrado do desvio padrão populacional σ) e a variância da média amostral $\sigma_{\bar{X}}^2$ (quadrado do desvio padrão da média amostral $\sigma_{\bar{X}}$).

O enunciado informa que a variância da média amostral é $\sigma_{\bar{X}}^2 = 0,1$.

Pronto! A matéria desta aula acabou. Agora, para calcular a variância populacional, precisamos saber calcular a média e a variância da distribuição contínua uniforme.

A média (esperança) dessa distribuição é igual à média aritmética dos limites do intervalo, a e b. Sabendo que a = $\boldsymbol{\alpha}$, b = 3.$\boldsymbol{\alpha}$ e E(X) = 12, conforme dados do enunciado, temos:

$$E(X) = \frac{a+b}{2}$$

$$12 = \frac{\alpha + 3\alpha}{2} = \frac{4\alpha}{2} = 2\alpha$$

$$\alpha = 6$$

Ou seja, a = 6 e b = 3x6 = 18. Então, a variância é dada por:

$$V(X) = \frac{(b-a)^2}{12} = \frac{(18-6)^2}{12} = \frac{(12)^2}{12} = 12$$

Voltando à nossa fórmula para encontrar o tamanho amostral, temos:

$$n = \frac{\sigma^2}{\sigma_{\bar{X}}^2} = \frac{12}{0,1} = 120$$

Gabarito: C

Fator de Correção para População Finita

Os resultados da variância e do desvio padrão da média amostral são válidos para variáveis $X_1, X_2, \dots, X_n$ **independentes**, ou seja, quando a população é **infinita ou** quando as amostras são extraídas **com reposição**.

Quando isso não ocorre, ou seja, quando a **população é finita e** as amostras são extraídas **sem reposição**, precisamos fazer um **ajuste**.

Sendo n o tamanho da amostra e N o tamanho da população, precisamos multiplicar a **variância** da média amostral, pelo **fator de correção de população finita** $\boldsymbol{\frac{N-n}{N-1}}$.

$$\boxed{V(\bar{X}_*) = \frac{\sigma^2}{n} \times \boldsymbol{\frac{N-n}{N-1}}}$$

Observe que o fator de correção é **menor** do que **1**, pois $N - n < N - 1$. Assim, o ajuste **diminui** a variância da média amostral.

Sabendo que o erro (ou desvio) padrão é a raiz quadrada da variância, podemos ajustá-lo para populações finitas e amostras extraídas sem reposição, multiplicando-o pela raiz quadrada do fator de correção:

$$EP(\bar{X}_*) = \sqrt{V(\bar{X}_*)} = \sqrt{\frac{\sigma^2}{n} \times \boldsymbol{\frac{N-n}{N-1}}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\boldsymbol{\frac{N-n}{N-1}}}$$

Distribuição da Média Amostral e a Curva Normal

Quando a **população** segue distribuição **normal (ou gaussiana)** com variância conhecida, a **média amostral** também seguirá **distribuição normal**.

Como vimos anteriormente, a esperança, a variância e o desvio (ou erro) padrão da média amostral são:

$$E(\bar{X}) = \mu$$

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

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Assim, para calcular as **probabilidades** envolvendo a média amostral, nessa situação, utilizamos a seguinte **transformação** para a **normal padrão**:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Por exemplo, vamos supor uma população com média $\mu = 20$ e desvio padrão $\sigma = 2$. Para calcular a probabilidade de a **média de uma amostra** de tamanho $n = 16$ ser maior que $\bar{x} = 21$, fazemos:

$$z = \frac{21 - 20}{\frac{2}{\sqrt{16}}} = \frac{1}{\frac{2}{4}} = 2$$

Assim, a probabilidade $P(\bar{X} > 21)$ é igual à probabilidade $P(Z > 2)$, que pode ser obtida a partir da tabela normal padrão.

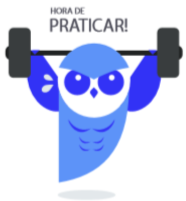
De modo equivalente, podemos dizer que a seguinte variável segue distribuição **normal padrão (ou reduzida)**, isto é, com média igual a 0 e desvio padrão igual a 1, que representamos como $N(0,1)$:

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

Ainda que a população **não** siga distribuição normal com variância conhecida, pelo **Teorema Central do Limite**, é possível **aproximar** a distribuição da média amostral a uma **normal**, também com média $E(\bar{X}) = \mu$, variância $V(\bar{X}) = \frac{\sigma^2}{n}$ e desvio padrão (ou erro padrão) $EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, quando o tamanho da amostra n for **grande** o bastante.

A possibilidade de tal aproximação **depende** do **tamanho da amostra** e da **distribuição da população**.

Quanto **maior** o tamanho da amostra e quanto mais **simétrica** for a distribuição da população, **melhor** será a aproximação. Usualmente, considera-se que para uma amostra grande, com $n \geq 30$, a aproximação será satisfatória, para **qualquer distribuição** populacional.



(CESPE/2014 – ANATEL) Com base no teorema limite central, julgue o item abaixo.

Sendo uma amostra aleatória simples retirada de uma distribuição X com média μ e variância 1, a distribuição da média amostral dessa amostra, $\bar{X}$, converge para uma distribuição normal de média nμ e variância 1, à medida que n aumenta.

Comentários:

A distribuição da média amostral $\bar{X}$ converge para uma distribuição normal à medida que n aumenta. Porém, a média dessa distribuição é $E(\bar{X}) = \mu$ e variância $V(\bar{X}) = \frac{V(X)}{n}$. Sabendo que $V(X) = 1$, a variância da média amostral é $V(\bar{X}) = \frac{1}{n}$ (e não 1).

Gabarito: Errado.

(CESPE 2018/PF) O tempo gasto (em dias) na preparação para determinada operação policial é uma variável aleatória X que segue distribuição normal com média M, desconhecida, e desvio padrão igual a 3 dias. A observação de uma amostra aleatória de 100 outras operações policiais semelhantes a essa produziu uma média amostral igual a 10 dias. Com referência a essas informações, julgue o item que se segue, sabendo que $P(Z > 2) = 0{,}025$, em que Z denota uma variável aleatória normal padrão.

O erro padrão da média amostral foi inferior a 0,5 dia.

Comentários:

O erro padrão da média amostral é:

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Em que n é o tamanho da amostra ($n = 100$); e σ é o desvio padrão amostral ($\sigma = 3$), logo:

$$EP(\bar{X}) = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0{,}3$$

Gabarito: Certo.

(CESPE/2019 – Analista Judiciário TJ) Um pesquisador deseja comparar a diferença entre as médias de duas amostras independentes oriundas de uma ou duas populações gaussianas. Considerando essa situação hipotética, julgue o próximo item.

Para que a referida comparação seja efetuada, é necessário que ambas as amostras tenham N ≥ 30.

Comentários:

Quando a população segue uma distribuição normal (ou gaussiana), a média amostral também seguirá uma distribuição normal, **independentemente do tamanho da amostra**. Logo, o item está errado.

Para fins de complementação, se a amostra for grande o suficiente (normalmente, consideramos isso para $n \geq 30$), a média amostral seguirá aproximadamente uma distribuição normal, mesmo que a população não siga distribuição normal.

Gabarito: Errado.

(FCC 2015/SEFAZ-PI) Instrução: Para responder à questão utilize, dentre as informações dadas a seguir, as que julgar apropriadas. Se Z tem distribuição normal padrão, então:

P(Z < 0,4) = 0,655; P(Z < 1,2) = 0,885; P(Z < 1,6) = 0,945; P(Z < 1,8) = 0,964; P(Z < 2) = 0,977.

Uma auditoria feita em uma grande empresa considerou uma amostra aleatória de 64 contas a receber. Se a população de onde essa amostra provém é infinita e tem distribuição normal com desvio padrão igual a R$ 200,00 e média igual a R$ 950,00, a probabilidade da variável aleatória média amostral, usualmente denotada por $\boldsymbol{X}$, estar situada entre R$ 980,00 e R$ 1.000,00 é dada por

a) 18,4%

b) 9,2%

c) 28,5%

d) 47,7%

e) 86,2%

Comentários:

O enunciado informa que a população segue distribuição normal, logo, a média amostral também terá distribuição normal. Assim, utilizamos a seguinte transformação para a normal padrão:

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Sabemos que o tamanho da amostra é n = 64, logo, $\sqrt{n} = 8$; que a média populacional é $\mu = 950$ e que o desvio padrão populacional é $\sigma = 200$. Substituindo esses valores, a transformação para $\overline{X_{inf}} = 980$ é:

$$Z_{\bar{X}} = \frac{980 - 950}{\frac{200}{8}} = 30 \times \frac{8}{200} = \frac{6}{5} = 1{,}2$$

Para $\overline{X_{sup}} = 1000$, temos:

$$Z_{\bar{X}} = \frac{1000 - 950}{\frac{200}{8}} = 50 \times \frac{8}{200} = 2$$

Ou seja, a probabilidade desejada corresponde a $P(1{,}2 < Z < 2)$, que pode ser calculada pelos dados fornecidos no enunciado:

$$P(1{,}2 < Z < 2) = P(Z < 2) - P(Z < 1{,}2) = 0{,}977 - 0{,}885 = 0{,}092 = 9{,}2\%$$

Gabarito: B

Distribuição Amostral da Proporção

Agora, vamos trabalhar com uma população em que **determinada característica** está presente em uma **proporção p** dessa população, por exemplo, 15% da população apresenta olhos azuis; 20% da população está doente, 1% da produção apresenta defeito, etc.

Um elemento qualquer da população X pode **apresentar** a característica estudada, o que chamamos de **sucesso** ($X = 1$), ou **não**, o que chamamos de **fracasso** ($X = 0$). A probabilidade de sucesso é $\boldsymbol{p}$ e a probabilidade de fracasso é $q = 1 - p$.

Essa população apresenta uma distribuição de **Bernoulli**, com parâmetro $\boldsymbol{p}$.

Sendo essa proporção populacional desconhecida, precisamos **estimá-la** a partir da proporção de sucessos encontrados na **amostra**, que indicamos por $\hat{\boldsymbol{p}}$.

Considerando que cada observação X_i da amostra será $X_i = 0$ ou $X_i = 1$ (assim como para a população), então a proporção de sucessos na amostra pode ser calculada como:

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Vamos supor que queremos estimar a proporção de **defeitos** em uma produção de medicamentos. Para isso, extraímos uma amostra de 10 medicamentos, que apresentou o seguinte resultado, em que 0 representa um item não defeituoso e 1 representa um item defeituoso:

{0, 0, 1, 0, 0, 0, 1, 0, 0, 0}

Logo, a proporção encontrada nessa amostra é:

$$\hat{p} = \frac{0 + 0 + 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0}{10} = \frac{2}{10} = 0{,}2$$

Note que o estimador $\hat{p}$ é calculado da mesma forma que a média amostra $\bar{\boldsymbol{X}}$ que vimos anteriormente. Logo, a **esperança** de $\hat{p}$ é calculada da mesma forma que para $\bar{X}$, utilizando $\boldsymbol{p}$ no lugar de μ:

$$E(\hat{p}) = \boldsymbol{p}$$

Ou seja, a **esperança** do estimador é igual à **proporção populacional**.

Em outras palavras, a proporção amostral **tende** à **proporção populacional**.

A **variância** de $\hat{p}$ também é calculada de forma análoga à de $\bar{X}$:

$$V(\hat{p}) = \frac{V(p)}{n}$$

Sabendo que a população segue distribuição de Bernoulli, temos $V(p) = p.q$, então:

$$\boxed{V(\hat{p}) = \frac{p.q}{n}}$$

E o **erro padrão** (ou desvio padrão) para $\hat{p}$, raiz quadrada da sua variância é dado por:

$$\boxed{EP(\hat{p}) = \sqrt{\frac{p.q}{n}}}$$

Se a proporção populacional p for **desconhecida**, utilizamos a proporção amostral $\hat{p}$ para **estimar** a variância populacional.

A estimativa da **variância populacional** é dada por:

$$V(p) = \hat{p}.\hat{q}$$

Em que $\hat{q} = 1 - \hat{p}$. E a estimativa da **variância do estimador $\hat{p}$** é:

$$V(\hat{p}) = \frac{\hat{p}.\hat{q}}{n}$$

Para o nosso exemplo, em que encontramos $\hat{p} = 0{,}2$ (logo, $\hat{q} = 1 - \hat{p} = 0{,}8$). A estimativa da **variância** da proporção **populacional** é:

$$V(p) = \hat{p}.\hat{q} = 0{,}2 \times 0{,}8 = 0{,}16$$

E a estimativa da **variância** da proporção **amostral** é:

$$V(\hat{p}) = \frac{\hat{p}.\hat{q}}{n} = \frac{0{,}2 \times 0{,}8}{10} = 0{,}016$$

Logo, a estimativa para o **erro padrão** (ou desvio padrão) da **proporção amostral** é:

$$EP(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{0{,}016} \cong 0{,}126$$

Considerando que cada elemento da população segue distribuição de Bernoulli, então o número de elementos com o atributo sucesso encontrados em uma **amostra** de tamanho n segue uma distribuição **binomial**, com parâmetros n e p. Para o nosso exemplo, temos uma distribuição binomial com $n = 10$ e proporção estimada $\hat{p} = 0{,}2$.

Porém, também é possível aproximar, pelo **Teorema Central do Limite**, a distribuição da proporção amostral a uma **normal**, com média $E(\hat{p}) = p$, variância $V(\hat{p}) = \frac{p.q}{n}$ e desvio (ou erro) padrão $EP(\hat{p}) = \sqrt{\frac{p.q}{n}}$, quando o tamanho da amostra n for suficientemente **grande**.

Assim, para calcular as probabilidades envolvendo a proporção amostral, utilizamos a **transformação** para a **normal padrão**:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}}$$

Por exemplo, supondo que a proporção de sucesso de uma população seja $p = 0{,}5$ (e proporção de fracasso $q = 1 - p = 0{,}5$), vamos calcular a probabilidade de observar uma proporção maior que $\hat{p} = 0{,}6$ em uma amostra de tamanho $n = 25$:

$$z = \frac{0{,}6 - 0{,}5}{\sqrt{\frac{0{,}5 \times 0{,}5}{25}}} = \frac{0{,}1}{\frac{0{,}5}{5}} = \frac{0{,}1}{0{,}1} = 1$$

Assim, a probabilidade $P(\hat{p} > 0{,}6)$ é aproximadamente igual à probabilidade $P(Z > 1)$, que pode ser calculada pela tabela da normal padrão.

Se a população for **finita** e a amostra for extraída **sem reposição**, será necessário aplicar o fator de **correção** para população finita, multiplicando a **variância** da proporção amostral por $\frac{N-n}{N-1}$:

$$V(\hat{p}) = \frac{p.q}{n} \times \frac{N-n}{N-1}$$

E o erro (ou desvio) padrão do estimador, com a correção para população finita, é igual à raiz quadrada:

$$EP(\hat{p}) = \sqrt{\frac{p.q}{n} \times \frac{N-n}{N-1}}$$

HORA DE PRATICAR!

(CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra {0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1} foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

A estimativa pontual da proporção de empresas da amostra com CNPJ regular é superior a 50%.

Comentários:

O estimador da proporção $\hat{p}$ pode ser calculado pela soma dos valores dos elementos, dividida pelo número de elementos na amostra:

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\hat{p} = \frac{0+1+1+0+0+1+0+1+0+1+1+0+0+1+1+0+1+1+1+1}{20} = \frac{12}{20} = 0{,}6$$

Logo, a proporção é de 60%.

Gabarito: Certo

(CESPE/2019 – TJ-AM) Para estimar a proporção de menores infratores reincidentes em determinado município, foi realizado um levantamento estatístico. Da população-alvo desse estudo, constituída por 10.050 menores infratores, foi retirada uma amostra aleatória simples sem reposição, composta por 201 indivíduos. Nessa amostra foram encontrados 67 reincidentes. Com relação a essa situação hipotética, julgue o seguinte item.

A estimativa do erro padrão da proporção amostral foi inferior a 0,04.

Comentários:

O erro padrão da proporção é dada pela relação:

$$EP(\hat{p}) = \sqrt{\frac{p \times q}{n}}$$

A questão nos diz que a amostra é composta por 201 indivíduos, sendo 67 deles reincidentes. Assim, temos $n = 201$ e $p = \frac{67}{201} = \frac{1}{3}$. Logo, $q = 1 - p = \frac{2}{3}$. Logo, o erro padrão é:

$$E = \sqrt{\frac{\frac{1}{3} \times \frac{2}{3}}{201}} = \frac{1}{3} \times \sqrt{\frac{2}{201}} \cong \frac{1}{3} \times \frac{1}{10} \cong 0{,}033$$

Gabarito: Certo.

RESUMINDO

Estimador para a média: $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$

Esperança: $E(\bar{X}) = \mu$; Variância: $V(\bar{X}) = \frac{V(X)}{n}$

Estimador para a proporção: $\hat{p} = \frac{X_1+X_2+\cdots+X_n}{n}$

Esperança: $E(\hat{p}) = p$; Variância: $V(\hat{p}) = \frac{p.q}{n}$

Distribuição Amostral da Variância

Quando a variância da população é desconhecida, precisamos estimá-la a partir da amostra, assim como fizemos com a média e a proporção.

O **estimador da variância** que utilizamos para uma amostra de tamanho n é:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{\boldsymbol{n-1}}$$

Utilizamos esse estimador, com a divisão por $\boldsymbol{n-1}$, porque a sua **esperança** é igual à variância populacional, como veremos posteriormente, o que **não** ocorre com o estimador com a divisão por n.

EXEMPLIFICANDO

Vamos supor que a variância da altura de determinado grupo de adultos seja desconhecida, assim como a sua média. Para estimar esses parâmetros, obtemos a seguinte amostra de 5 pessoas:

$$\{1{,}65;\ 1{,}75;\ 1{,}8;\ 1{,}85;\ 1{,}95\}$$

Primeiro, precisamos calcular a média da amostra:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1{,}65+1{,}75+1{,}8+1{,}85+1{,}95}{5} = \frac{9}{5} = 1{,}8$$

Agora, calculamos o estimador da variância, somando os desvios em relação à média, elevados ao quadrado, e dividindo o somatório por $n - 1$:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

$$s^2 = \frac{(1{,}65-1{,}8)^2+(1{,}75-1{,}8)^2+(1{,}8-1{,}8)^2+(1{,}85-1{,}8)^2+(1{,}95-1{,}8)^2}{4}$$

$$s^2 = \frac{(-0{,}15)^2+(-0{,}05)^2+(0)^2+(0{,}05)^2+(0{,}15)^2}{4}$$

$$s^2 = \frac{0{,}0225+0{,}0025+0{,}0025+0{,}0225}{4} = \frac{0{,}05}{4} = 0{,}0125$$

Uma maneira alternativa de calcular a variância **populacional** é:

$$\sigma^2 = E(X^2) - [E(X)]^2$$

Para a variância **amostral**, também podemos utilizar uma fórmula similar a essa, mas com as devidas adaptações. No lugar de $E(X)$, utilizamos a média amostral $\bar{X}$; e, no lugar de $E(X^2)$, utilizamos $\overline{X^2}$, que é a média dos valores elevados ao quadrado:

$$\overline{X^2} = \frac{\sum X_i^2}{n}$$

Para o exemplo anterior, teríamos:

$$\overline{X^2} = \frac{(1{,}65)^2+(1{,}75)^2+(1{,}8)^2+(1{,}85)^2+(1{,}95)^2}{5} = \frac{16{,}25}{5} = 3{,}25$$

Por fim, fazemos um ajuste. Para calcular a variância populacional, dividimos por n e para a variância amostral, dividimos por $n - 1$.

Logo, precisamos multiplicar o resultado por $\frac{n}{n-1}$ para obter a variância amostral:

$$s^2 = \left[\overline{X^2} - (\bar{X})^2\right] \times \frac{n}{n-1}$$

Para o nosso exemplo, em que $\overline{X^2} = 3{,}25$, $\bar{X} = 1{,}8$ e $n = 5$, a variância amostral pode ser calculada como:

$$s^2 = [3{,}25 - (1{,}8)^2] \times \frac{5}{4} = [3{,}25 - 3{,}24] \times \frac{5}{4} = \frac{0{,}01 \times 5}{4} = 0{,}0125$$

Para a variância amostral, a sua **esperança** é igual à **variância populacional**, analogamente ao que ocorreu com os demais estimadores. A sua variância e erro padrão são dados por:

$$E(s^2) = \sigma^2$$

$$V(s^2) = \frac{2.\sigma^4}{n-1}$$

$$EP(s^2) = \sqrt{\frac{2}{n-1}}\sigma^2$$

Se a variância populacional σ^2 for **desconhecida**, utilizamos, no lugar de σ^2, a própria estimativa s^2 para estimar a média $E(s^2)$, a variância $V(s^2)$ e o erro padrão $EP(s^2)$.

Para o nosso exemplo, em que calculamos $s^2 = 0{,}0125$, para a amostra com $n = 5$ observações, o erro padrão de s^2 pode ser estimado como:

$$EP(s^2) = \sqrt{\frac{2}{n-1}}s^2 = \sqrt{\frac{2}{4}} \times 0{,}0125 \cong 0{,}009$$

Se a população seguir uma distribuição **normal**, então o estimador s^2, multiplicado pelo fator $\boldsymbol{\frac{n-1}{\sigma^2}}$, segue uma distribuição **qui-quadrado** com $\boldsymbol{n-1}$ graus de liberdade:

$$\mathcal{X}^2_{n-1} = \left(\frac{n-1}{\sigma^2}\right).s^2$$

Em outras palavras, o estimador s^2 é uma variável com distribuição qui-quadrado, com $n-1$ graus de liberdade, multiplicada por $\frac{\sigma^2}{n-1}$:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\mathcal{X}^2_{n-1}$$

Vamos supor que, para uma população normal, a variância populacional seja $\sigma^2 = 1$, e que vamos extrair amostras de tamanho $n = 5$. Nesse caso, a variância amostral s^2 terá a seguinte distribuição:

$$s^2 = \left(\frac{1}{5-1}\right).\mathcal{X}^2_{5-1} = \frac{\mathcal{X}^2_4}{4}$$

Ou seja, a variância amostral seguirá uma distribuição qui-quadrado com $n - 1 = 4$ graus de liberdade, dividida por 4.

A seguir, consta a tabela da distribuição qui-quadrado com 4 graus de liberdade, que apresenta os valores de probabilidade $P(\mathcal{X}^2_4 < x)$ e os respectivos valores de x. Como a variância amostral segue essa distribuição, dividida por 4, criamos uma terceira coluna, dividindo os valores de x por 4.

$P(\mathcal{X}_4^2 < x)$	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,25**	**0,5**	**0,75**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
x	0,21	0,30	0,48	0,71	1,06	1,92	3,36	5,39	7,78	9,49	11,14	13,28	14,86
$x/4$	0,05	0,07	0,12	0,18	0,27	0,48	0,84	1,35	1,94	2,37	2,79	3,32	3,72

Por exemplo, a probabilidade de a variância amostral observada ser inferior a 0,48 é:

$$P(s^2 < 0{,}48) = \mathbf{0,25}$$

A partir desse resultado, podemos calcular a esperança e a variância do estimador. A **esperança** é dada por:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\mathcal{X}_{n-1}^2$$

$$E[s^2] = E\left[\left(\frac{\sigma^2}{n-1}\right).\mathcal{X}_{n-1}^2\right] = \left(\frac{\sigma^2}{n-1}\right).E[\mathcal{X}_{n-1}^2]$$

Considerando que a média de uma distribuição qui-quadrado com k graus de liberdade é igual a k, então fazendo $k = n - 1$, temos:

$$E[\mathcal{X}_{n-1}^2] = k = n - 1$$

Substituindo no resultado anterior, temos:

$$E[s^2] = \left(\frac{\sigma^2}{n-1}\right).(n-1) = \sigma^2$$

Esse é o resultado que vimos no início da seção. A **variância** do estimador é:

$$V[s^2] = V\left[\left(\frac{\sigma^2}{n-1}\right)\mathcal{X}_{n-1}^2\right] = \left(\frac{\sigma^2}{n-1}\right)^2.V[\mathcal{X}_{n-1}^2] = \left(\frac{\sigma^4}{(n-1)^2}\right).V[\mathcal{X}_{n-1}^2]$$

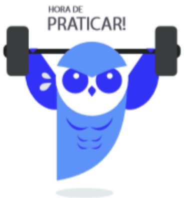
Considerando que a variância de uma distribuição qui-quadrado com k graus de liberdade é igual a $2k$, então fazendo $k = n - 1$, temos:

$$V[\mathcal{X}_{n-1}^2] = 2.k = 2.(n-1)$$

Substituindo no resultado anterior, temos:

$$V[s^2] = \left(\frac{\sigma^4}{(n-1)^2}\right).2.(n-1) = \frac{2.\sigma^4}{n-1}$$

Vale acrescentar que uma população normal depende dos dois parâmetros, variância e média, os quais são independentes. Consequentemente, os estimadores correspondentes também serão **independentes**.



(FGV/2016 – IBGE) Suponha que uma amostra de tamanho $\boldsymbol{n} = \mathbf{5}$ é extraída de uma população normal, com média desconhecida, obtendo as seguintes observações:

$$\boldsymbol{X_1} = \mathbf{3},\ \boldsymbol{X_2} = \mathbf{5},\ \boldsymbol{X_3} = \mathbf{6},\ \boldsymbol{X_4} = \mathbf{9} \text{ e } \boldsymbol{X_5} = \mathbf{12}$$

São dados ainda os seguintes valores, retirados da tabela da distribuição qui-quadrado:

- $P(\chi_4^2 < \mathbf{5}) \cong \mathbf{0{,}713}$
- $P(\chi_4^2 < \mathbf{12{,}5}) \cong \mathbf{0{,}986}$
- $P(\chi_5^2 > \mathbf{5}) \cong \mathbf{0{,}854}$
- $P(\chi_5^2 > \mathbf{12{,}5}) \cong \mathbf{0{,}971}$

Se a população tem variância verdadeira $\boldsymbol{\sigma}^{\mathbf{2}} = \mathbf{4}$, em nova amostra ($\boldsymbol{n} = \mathbf{5}$), a probabilidade de se observar uma variância amostral maior do que a anterior é de:

a) 0,014

b) 0,029

c) 0,146

d) 0,287

e) 0,713

Comentários:

Para resolver essa questão, vamos primeiro calcular a variância amostral obtida nessa primeira amostra:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{\boldsymbol{n} - \mathbf{1}}$$

Para isso, precisamos da média amostral:

$$\bar{X} = \frac{3+5+6+9+12}{5} = \frac{35}{5} = 7$$

Agora, podemos calcular a variância dessa primeira amostra:

$$s_1^2 = \frac{(3-7)^2 + (5-7)^2 + (6-7)^2 + (9-7)^2 + (12-7)^2}{4}$$

$$s_1^2 = \frac{(-4)^2 + (-2)^2 + (-1)^2 + (2)^2 + (5)^2}{4} = \frac{16+4+1+4+25}{4} = \frac{50}{4} = 12{,}5$$

Para calcular a probabilidade de a variância amostral ser $s^2 > 12{,}5$, consideramos que esse estimador segue uma distribuição qui-quadrado com n – 1 graus de liberdade, multiplicada por $\left(\frac{\sigma^2}{n-1}\right)$:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\mathcal{X}^2_{n-1}$$

Sabendo que a variância populacional é $\sigma^2 = 4$ e que o tamanho da amostra é $n = 5$, então:

$$s^2 = \left(\frac{4}{5-1}\right).\mathcal{X}^2_{5-1} = \mathcal{X}^2_4$$

Portanto, a variância amostral segue a mesma distribuição de $\mathcal{X}^2_4$. A probabilidade de $s^2 > 12{,}5$ é, portanto:

$$P(s^2 > 12{,}5) = P(\mathcal{X}^2_4 > 12{,}5)$$

O enunciado informa que $P(\mathcal{X}^2_4 < 12{,}5) = 0{,}986$. A probabilidade $P(\mathcal{X}^2_4 > 12{,}5)$ é complementar:

$$P(\mathcal{X}^2_4 > 12{,}5) = 1 - P(\mathcal{X}^2_4 < 12{,}5) = 1 - 0{,}986 = 0{,}014$$

Gabarito: A

Distribuições para Amostragem Estratificada

Para uma **amostragem estratificada**, com k estratos, a **média amostral** será calculada como:

$$\bar{X} = \sum_{i=1}^{k} \frac{N_i}{N}\bar{x_i}$$

Nessa expressão, N_i é o tamanho de cada **estrato**; N é o tamanho total da **população**; e $\bar{x_i}$, a **média amostral** observada para cada estrato. Ou seja, calculamos a média $\bar{x_i}$ para cada estrato i, multiplicamos pelo tamanho do estrato N_i e dividimos pelo tamanho total N.

Para ilustrar, vamos supor uma população dividida em 3 estratos, com os seguintes tamanhos N_i e os seguintes valores de média amostral $\bar{x_i}$ para cada estrato i:

Estrato	N_i	n_i	$\bar{x_i}$
1	50	5	2
2	30	3	3
3	20	2	4

A média amostral para toda a população corresponde às médias amostrais dos estratos, ponderadas pelos respectivos tamanhos dos estratos:

$$\bar{X} = \frac{50 \times 2 + 30 \times 3 + 20 \times 4}{50 + 30 + 20} = \frac{100 + 90 + 80}{100} = 2{,}7$$

A **razão** entre o tamanho do estrato N_i e o tamanho total da população N pode ser chamada de **peso** do estrato (W_i). Nessa situação, para calcular a média amostral, basta multiplicar a média de cada estrato pelo respectivo peso e somar todos os resultados:

$$\bar{X} = \sum_{i=1}^{k} W_i \times \bar{x_i}$$

Podemos calcular, ainda, a **variância da média amostral**:

$$V(\bar{X}) = V\left(\sum_{i=1}^{k} \frac{N_i}{N} \bar{x_i}\right)$$

Pelas propriedades da variância, temos:

$$V(\bar{X}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\bar{x_i})$$

Em que a variância da média amostral de cada estrato $V(\bar{x_i})$ é calculada pela **razão** entre a variância **populacional** do estrato e o **tamanho da amostra** do estrato:

$$V(\bar{x_i}) = \frac{V(X_i)}{n_i}$$

Em uma amostragem **proporcional**, o tamanho amostral do estrato n_i pode ser calculado pela **razão** entre o tamanho do estrato populacional e o tamanho total da população, **multiplicada** pelo tamanho total da amostra:

$$n_i = \frac{N_i}{N} . n$$

Vamos supor as seguintes variâncias para os estratos da população, que vimos anteriormente:

Estrato	N_i	n_i	$V(X_i)$
1	50	5	4
2	30	3	2
3	20	2	1

As variâncias das **médias amostrais** dos estratos são dadas por:

$$V(\overline{x_1}) = \frac{V(X_1)}{n_1} = \frac{4}{5} = 0{,}8$$

$$V(\overline{x_2}) = \frac{V(X_2)}{n_2} = \frac{2}{3} \cong 0{,}67$$

$$V(\overline{x_3}) = \frac{V(X_3)}{n_3} = \frac{1}{2} = 0{,}5$$

E a variância da média amostral global é dada por:

$$V(\bar{X}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\overline{x_i}) = \left(\frac{N_1}{N}\right)^2 . V(\overline{x_1}) + \left(\frac{N_2}{N}\right)^2 . V(\overline{x_2}) + \left(\frac{N_3}{N}\right)^2 . V(\overline{x_3})$$

$$V(\bar{X}) = \left(\frac{50}{100}\right)^2 \times 0{,}8 + \left(\frac{30}{100}\right)^2 \times 0{,}67 + \left(\frac{20}{100}\right)^2 \times 0{,}5 = 0{,}25 \times 0{,}8 + 0{,}09 \times 0{,}67 + 0{,}04 \times 0{,}5 = 0{,}28$$

A variância da média de uma amostra estratificada proporcional é **menor ou igual** do que a variância da média de uma amostra aleatória simples (AAS). Afinal, a amostragem estratificada proporcional é **mais precisa** do que a AAS[1].

[1] A variância da média de uma AAS é a razão entre a variância da população como um todo e o tamanho da amostra; e a variância de toda a população é **maior** (ou igual) às variâncias dos estratos ponderadas pelos respectivos tamanhos:

$$V(\bar{X}_{AAS}) = \frac{1}{n} \times V(X) \geq \frac{1}{n} \times \sum_{i=1}^{k} \frac{N_i}{N} . V(X_i)$$

A expressão à direita corresponde justamente à variância da média amostral global em uma amostragem estratificada proporcional. Sabendo que $V(\bar{X}_{Est}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\bar{x_i})$, que $V(\bar{x_i}) = \frac{V(X_i)}{n_i}$ e que, na alocação proporcional, $n_i = \frac{N_i}{N} . n$, então:

$$V(\bar{X}_{EP}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \frac{V(X_i)}{n_i} = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \frac{V(X_i). N}{N_i . n} = \sum_{i=1}^{k} \frac{N_i}{N} \frac{V(X_i)}{n}$$

Sendo a variância populacional de cada estrato **desconhecida**, precisamos estimá-la a partir da variância amostral de cada estrato $s_{\bar{x}_i}^2$. Substituindo, na fórmula acima, $V(\bar{X})$ por $s_{\bar{x}}^2$ e $V(\bar{x_\iota})$ por $s_{\bar{x}_i}^2$, temos:

$$s_{\bar{x}}^2 = \sum_{h=1}^{k} \left(\frac{N_i}{N}\right)^2 \times s_{\bar{x}_i}^2$$

Em que a estimativa da variância da média amostral para cada estrato $s_{\bar{x}_i}^2$, considerando uma amostra de tamanho n_i para esse estrato, com a **correção para população finita**, é dada por:

$$s_{\bar{x}_i}^2 = \frac{s_{x_i}^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$$

Vamos supor que as estimativas da variância para cada estrato sejam:

Estrato	N_i	n_i	$s_{x_i}^2$
1	50	5	2
2	30	3	1,5
3	20	2	1

Assim, as variâncias das médias amostrais para cada estrato, com o fator de correção, são:

$$s_{\bar{x}_1}^2 = \frac{2}{5}\left(\frac{50 - 5}{50 - 1}\right) \cong 0{,}367$$

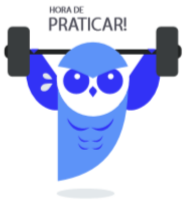
$$s_{\bar{x}_2}^2 = \frac{1{,}5}{3}\left(\frac{30 - 3}{30 - 1}\right) \cong 0{,}466$$

$$s_{\bar{x}_3}^2 = \frac{1}{2}\left(\frac{20 - 2}{20 - 1}\right) \cong 0{,}474$$

E a variância da média amostral global é dada por:

$$s_{\bar{x}}^2 = \left(\frac{50}{100}\right)^2 \times 0{,}367 + \left(\frac{30}{100}\right)^2 \times 0{,}466 + \left(\frac{20}{100}\right)^2 \times 0{,}474$$

$$s_{\bar{x}}^2 = 0{,}25 \times 0{,}367 + 0{,}09 \times 0{,}466 + 0{,}04 \times 0{,}474 = 0{,}092 + 0{,}042 + 0{,}019 = 0{,}153$$



(CESPE/2018 – STM) Um estudo acerca do tempo (x, em anos) de guarda de autos findos em determinada seção judiciária considerou uma amostragem aleatória estratificada. A população consiste de uma listagem de autos findos, que foi segmentada em quatro estratos, segundo a classe de cada processo (as classes foram estabelecidas por resolução de autoridade judiciária.)

A tabela a seguir mostra os tamanhos populacionais (N) e amostrais (n), a média amostral ($\bar{x}$) e a variância amostral dos tempos (s^2) correspondentes a cada estrato.

Estratos	Tamanhos Populacionais (N)	Tamanhos Amostra (n)	$\bar{x}$	s^2
A	30.000	300	20	3
B	40.000	400	15	16
C	50.000	500	10	5
D	80.000	800	5	8
Total	200.000	2.000	-	-

Considerando que o objetivo do estudo seja estimar o tempo médio populacional (em anos) de guarda dos autos findos, julgue os itens a seguir.

(CESPE/2018 – STM) A estimativa do tempo médio populacional da guarda dos autos findos é maior ou igual a 12 anos.

Comentários:

Em uma amostra aleatória por estratificação, a média amostral é dada por:

$$\bar{X} = \sum_{i=1}^{k} \frac{N_i}{N} \bar{x_i}$$

Em que k é a quantidade de estratos; N_i o tamanho de cada estrato; e $\bar{x_i}$ a média de cada estrato.

Pelos dados de N e $\bar{x}$ fornecidos na tabela, temos:

$$\bar{X} = \frac{N_A \times \overline{x_A} + N_B \times \overline{x_B} + N_C \times \overline{x_C} + N_D \times \overline{x_D}}{N}$$

$$\bar{X} = \frac{30000 \times 20 + 40000 \times 15 + 50000 \times 10 + 80000 \times 5}{200000}$$

$$\bar{X} = \frac{60 + 60 + 50 + 40}{20} = 10{,}5$$

Gabarito: Errado.

(CESPE/2018 – STM) Combinando-se todos os estratos envolvidos, a estimativa da variância do tempo médio amostral da guarda dos autos findos é inferior a $0{,}005\ ano^2$.

Comentários:

Em uma amostragem estratificada, a estimativa da variância da média amostral é dada por:

$$s_{\bar{x}}^2 = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \times s_{\bar{x}_i}^2$$

Em que $s_{\bar{x}_i}^2 = \frac{s_{\bar{x}_i}^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$. Pelos valores apresentados na tabela, temos:

$$s_{\bar{x}_A}^2 = \frac{3}{300}\left(\frac{30000 - 300}{30000 - 1}\right) = 0{,}0099$$

$$s_{\bar{x}_B}^2 = \frac{16}{400}\left(\frac{40000 - 400}{40000 - 1}\right) = 0{,}0396$$

$$s_{\bar{x}_C}^2 = \frac{5}{500}\left(\frac{50000 - 500}{50000 - 1}\right) = 0{,}0099$$

$$s_{\bar{x}_D}^2 = \frac{8}{800}\left(\frac{80000 - 800}{80000 - 1}\right) = 0{,}0099$$

Além disso, temos que:

$$\left(\frac{N_A}{N}\right)^2 = \left(\frac{30000}{200000}\right)^2 = 0{,}0225$$

$$\left(\frac{N_B}{N}\right)^2 = \left(\frac{40000}{200000}\right)^2 = 0{,}04$$

$$\left(\frac{N_C}{N}\right)^2 = \left(\frac{50000}{200000}\right)^2 = 0{,}0625$$

$$\left(\frac{N_D}{N}\right)^2 = \left(\frac{80000}{200000}\right)^2 = 0{,}16$$

Portanto:

$$s_{\bar{x}}^2 = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \times s_{\bar{x}_h}^2$$

$$s_{\bar{x}}^2 = (0{,}0225 \cdot 0{,}0099 + 0{,}04 \cdot 0{,}0396 + 0{,}0625 \cdot 0{,}0099 + 0{,}16 \cdot 0{,}0099)$$

$$s_{\bar{x}}^2 = (0{,}0225 + 0{,}0625 + 0{,}16) \cdot 0{,}0099 + 0{,}04 \cdot 0{,}0396 \cong 0{,}0041.$$

Gabarito: Certo.

QUESTÕES COMENTADAS – CESGRANRIO

Distribuição Binomial

1. (CESGRANRIO/2022 – Eletronuclear) Em uma fábrica automobilística, foi notado que os airbags de um certo lote não estavam funcionando tão bem quanto deveriam. Calcula-se que a probabilidade de um airbag desse lote não inflar era de 40%, valor muito acima do limite aceitável. Após o descarte de todo o lote, começou-se a fazer testes de colisão com airbags do lote seguinte. Supondo-se que os airbags do novo lote possuam a mesma probabilidade de não inflar que os do lote descartado e que, em cada colisão, apenas um airbag é testado, qual a probabilidade de 2 airbags não inflarem nas 5 primeiras colisões?

a) 12,96%

b) 20,16%

c) 31,25%

d) 32,75%

e) 34,56%

Comentários:

Precisamos calcular a probabilidade de 2 airbags não inflarem, dentre $n = 5$, sabendo que a probabilidade de um airbag não inflar (sucesso) é $p = 40\% = 0{,}4$, logo a probabilidade de um airbag inflar (fracasso) é complementar $q = 1 - p = 0{,}6$:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 2) = C_{5,2} \times 0,4^2 \times 0{,}6^3$$

A combinação $C_{5,2}$ é:

$$C_{5,2} = \frac{5!}{(5-2)! \times 2!} = \frac{5 \times 4 \times 3!}{3! \times 2} = 10$$

Assim, a probabilidade desejada é:

$$P(X = 2) = 10 \times 0{,}16 \times 0{,}216 = 0{,}3456 = 34{,}56\%$$

Gabarito: E

2. (CESGRANRIO/2021 – CEF - *Adaptada*) Por estudos estatísticos, estima-se que um cliente de um certo banco tem 75% de probabilidade de ir para atendimento de caixa eletrônico, e 25% de ir para um atendimento personalizado.

Em uma amostra de quatro clientes entrando no banco, qual é a probabilidade de que, no mínimo, a maioria deles se dirija ao atendimento personalizado?

a) 1/64

b) 5/256

c) 3/64

d) 13/256

e) 27/64

Comentários:

Precisamos calcular a probabilidade de que, dentre os 4 clientes, ao menos 3 se dirijam ao atendimento personalizado, o que corresponde à soma:

$$P(X \geq 3) = P(X = 3) + P(X = 4)$$

Essas probabilidades são dadas pela distribuição binomial, considerando que o tamanho da amostra é $n = 4$, que a probabilidade de um cliente se dirigir ao atendimento personalizado é $p = 25\% = \frac{1}{4}$ e que a probabilidade de um cliente não se dirigir a esse atendimento é $q = 75\% = \frac{3}{4}$:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 3) = C_{4,3} \times \left(\frac{1}{4}\right)^3 \times \left(\frac{3}{4}\right)^1 = 4 \times \frac{3}{4^4} = \frac{12}{4^4}$$

$$P(X = 4) = C_{4,4} \times \left(\frac{1}{4}\right)^4 \times \left(\frac{3}{4}\right)^0 = 1 \times \frac{1}{4^4} = \frac{1}{4^4}$$

E a soma é:

$$P(X \geq 3) = \frac{12}{4^4} + \frac{1}{4^4} = \frac{13}{256}$$

Gabarito: D

3. (CESGRANRIO/2018 – Liquigás) Sabe-se por estudos estatísticos que a eficiência de uma certa vacina para uma dada doença é de 80%.

Vacinando-se três indivíduos, qual a probabilidade de que apenas um deles não fique imunizado à doença?

a) 12,8%

b) 16%

c) 32%

d) 36%

e) 38,4%

Comentários:

O enunciado informa que a probabilidade de um indivíduo ser imunizado após tomar determinada vacina é p = 80%.

Sabendo que há apenas 2 possibilidades para cada indivíduo (ser imunizado ou não), e que serão vacinados n = 3 indivíduos, então temos uma distribuição binomial com n = 3 e p = 80% = 0,8 (logo, q = 1 – p = 1 – 0,8 = 0,2).

Assim, a probabilidade de ter 1 indivíduo não imunizado, consequentemente, 2 indivíduos imunizados é:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 2) = C_{3,2} \times 0,8^2 \times 0,2^1$$

A combinação $C_{3,2}$ é:

$$C_{3,2} = \frac{3!}{(3-2)! \times 2!} = \frac{3 \times 2!}{1! \times 2!} = 3$$

Assim, a probabilidade desejada é:

$$P(X = 2) = 3 \times 0,64 \times 0,2 = 0,384 = 38,4\%$$

Gabarito: E

4. (CESGRANRIO/2016 – IBGE) Uma pesquisa está interessada em estudar os eleitores de determinado candidato. Sabe-se que 50% da população alegam votar no candidato em questão.

Se 6 pessoas forem abordadas aleatoriamente, a probabilidade de que exatamente 3 pessoas sejam eleitoras do candidato em questão é aproximadamente

a) 51%

b) 50%

c) 31%

d) 21%

e) 11%

Comentários:

O enunciado informa que 50% das pessoas alegam votar em determinado candidato e que serão abordadas 6 pessoas.

Considerando que há somente 2 respostas possíveis (votar ou não no candidato) para cada pessoa (ou seja, cada pessoa corresponde a um ensaio de Bernoulli) e que serão abordadas n = 6 pessoas, então temos uma distribuição binomial com n = 6 e p = 50% = 0,5 (logo, q = 1 – p = 1 – 0,5 = 0,5).

Assim, a probabilidade de encontrar exatamente k = 3 pessoas que votam no candidato é:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

$$P(X = 3) = C_{6,3} \times 0,5^3 \times 0,5^{6-3}$$

A combinação $C_{6,3}$ é:

$$C_{6,3} = \frac{6!}{(6-3)! \times 3!} = \frac{6 \times 5 \times 4 \times 3!}{3! \times 3!} = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 5 \times 4 = 20$$

Assim, a probabilidade desejada é:

$$P(X = 3) = 20 \times 0,5^3 \times 0,5^3 = 20 \times 0,125 \times 0,125 = 0,3125 \cong 31\%$$

Gabarito: C

5. (CESGRANRIO/2012 – Petrobras) 10% dos parafusos produzidos por uma máquina são defeituosos. A probabilidade de que, entre 4 parafusos, pelo menos 3 não sejam defeituosos é de

a) 29,16%

b) 65,61%

c) 94,77%

d) 98,37%

e) 99,99%

Comentários:

O enunciado informa que 10% dos parafusos são **defeituosos** e que serão analisados 4 parafusos. Considerando que há somente 2 resultados possíveis (defeituoso ou não) para cada parafuso (ou seja, cada parafuso corresponde a um ensaio de Bernoulli) e que serão analisados n = 4 parafusos, então temos uma distribuição binomial com n = 4 e p = 10% = 0,1 (logo, q = 1 – p = 1 – 0,1 = 0,9).

A probabilidade P(X = k) para a distribuição binomial é calculada como:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

O enunciado deseja a probabilidade de encontrar pelo menos 3 **não** sejam defeituosos, ou seja, **no máximo 1 defeituoso**, isto é, 0 ou 1 parafuso defeituoso, dentre 4.

A probabilidade de não encontrar parafuso defeituoso algum é:

$$P(X = 0) = C_{4,0} \times 0,1^0 \times 0,9^4 = 1 \times 1 \times 0,6561 = 0,6561$$

A probabilidade de encontrar 1 parafuso defeituoso é:

$$P(X = 1) = C_{4,1} \times 0,1^1 \times 0,9^3 = 4 \times 0,1 \times 0,729 = 0,2916$$

Sabendo que buscamos a probabilidade de encontrar 0 **OU** 1 parafuso defeituoso, temos a **união de eventos excludentes**, dada pela **soma** das probabilidades:

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0,6561 + 0,2916 = 0,9477 = 94,77\%$$

Gabarito: C

6. (CESGRANRIO/2011 – Petrobras) Uma carta tem 2/3 de chances de chegar ao destino correto. Se seis cartas são enviadas de forma independente, a probabilidade de que pelo menos duas cheguem a o destino correto é

a) 4/9

b) 68/81

c) 113/162

d) 230/243

e) 716/729

Comentários:

O enunciado informa a probabilidade de uma carta chegar ao destino correto é de 2/3 e que serão enviadas 6 cartas.

Considerando que há apenas 2 resultados possíveis para cada carta (chegar ou não ao destino correto), ou seja, cada carta representa um ensaio de Bernoulli, e que serão enviadas n = 6 cartas, então temos uma distribuição binomial com n = 6 e p = 2/3. Logo, a probabilidade de a carta **não** chegar no destino correto é:

$$q = 1 - p = 1 - \frac{2}{3} = \frac{3}{3} - \frac{2}{3} = \frac{1}{3}$$

A probabilidade P(X = k) para a distribuição binomial é calculada como:

$$P(X = k) = C_{n,k} \times p^k \times q^{n-k}$$

O enunciado deseja a probabilidade de encontrar **pelo menos 2** que cheguem ao destino correto, dentre as 6 cartas enviadas. Podemos calcular essa probabilidade como o **complementar** da probabilidade de chegar **menos de 2** cartas ao destino correto:

$$P(X \geq 2) = 1 - P(X < 2)$$

Menos de 2 cartas corresponde a 0 ou 1 carta.

A probabilidade de não chegar carta alguma ao destino correto é:

$$P(X = 0) = C_{6,0} \times \left(\frac{2}{3}\right)^0 \times \left(\frac{1}{3}\right)^6 = 1 \times 1 \times \frac{1}{3^6} = \frac{1}{729}$$

A probabilidade de chegar 1 carta ao destino correto é:

$$P(X=1)=C_{6,1}\times\left(\frac{2}{3}\right)^{1}\times\left(\frac{1}{3}\right)^{5}=6\times\frac{2}{3}\times\frac{1}{3^5}=\frac{12}{3^6}=\frac{12}{729}$$

A probabilidade de chegar 0 **OU** 1 carta ao destino correto corresponde à **união de eventos excludentes**, que é dada pela **soma** das probabilidades:

$$P(X<2)=P(X=0)+P(X=1)=\frac{1}{729}+\frac{12}{729}=\frac{13}{729}$$

Portanto, a probabilidade de chegar pelo menos 2 cartas ao destino correto é complementar:

$$P(X\geq 2)=1-P(X<2)=1-\frac{13}{729}=\frac{729}{729}-\frac{13}{729}=\frac{716}{729}$$

Gabarito: E

7. (CESGRANRIO/2009 – BNDES) Em um dado com seis faces numeradas de 1 a 6, a probabilidade de que cada um dos resultados ocorra é a mesma. Esse dado será lançado até que se obtenha o resultado 6. A probabilidade de que isso aconteça em, no máximo, 2 lançamentos é

a) 1/36

b) 5/36

c) 6/36

d) 7/36

e) 11/36

Comentários:

O dado será lançado até que seja obtida a face 6 (que podemos chamar de sucesso). Sabendo que a probabilidade de obter a face 6 é p = 1/6, temos uma distribuição geométrica com parâmetro p = 1/6. E a probabilidade de **não** obter a face 6 (fracasso) é:

$$q=1-p=1-\frac{1}{6}=\frac{6}{6}-\frac{1}{6}=\frac{5}{6}$$

A probabilidade de ter que lançar o dado no máximo 2 vezes é a probabilidade de obter sucesso na 1ª ou na 2ª tentativa.

A probabilidade de obter sucesso na 1ª tentativa é:

$$P(X = 1) = p = \frac{1}{6}$$

E a probabilidade de obter sucesso na 2ª tentativa corresponde à probabilidade de obter fracasso na 1ª tentativa e sucesso na 2ª:

$$P(X = 2) = q \times p = \frac{5}{6} \times \frac{1}{6} = \frac{5}{36}$$

Então, a probabilidade de ter que lançar o dado no máximo 2 vezes, isto é, 1 **OU** 2 vezes, corresponde à união desses eventos mutuamente excludentes, dada pela soma das probabilidades:

$$P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{5}{36} = \frac{6}{36} + \frac{5}{36} = \frac{11}{36}$$

Gabarito: E

8. (CESGRANRIO/2008 – Petrobras) Um estudante marca, ao acaso, as respostas de um teste de 10 questões de múltipla escolha, com 4 alternativas por questão. O número mais provável de acertos é

a) 1,5

b) 2,0

c) 2,5

d) 3,0

e) 3,5

Comentários:

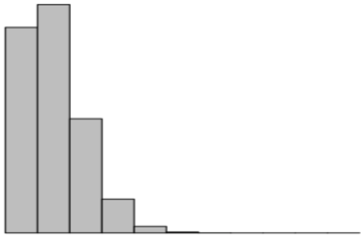
O enunciado informa que serão marcadas ao acaso 10 questões com 4 alternativas cada e pergunta pelo número mais provável de acertos, isto é, pela moda da distribuição.

Dentre alternativas, apenas duas são resultados possíveis para o número de acertos, quais sejam, 2 e 3, pois as demais não são números inteiros (e não há como acertar meia questão). Então, podemos resolver a questão, calculando as probabilidades P(X = 2) e P(X = 3), para saber qual valor é mais provável.

Os parâmetros da distribuição são n = 10 e a p = probabilidade de acertar uma questão. Sabendo que há 4 alternativas, a probabilidade de acertar uma questão ao acaso é:

$$p = \frac{1}{4}$$

Alternativamente, podemos perceber que a probabilidade p é menor que 0,5, logo, temos uma distribuição assimétrica positiva, conforme ilustrado a seguir:



Nesse tipo de distribuição, temos Moda < Média. A média da distribuição é:

$$E(X) = n \times p = 10 \times \frac{1}{4} = 2{,}5$$

Logo, a moda será necessariamente menor que a média, isto é, Moda = 2.

Gabarito: B

QUESTÕES COMENTADAS – CESGRANRIO

Distribuição Normal

1. (CESGRANRIO/2023 – BB) A distribuição das alturas dos atletas de vôlei de uma determinada seleção é normal. Sabe-se que 5% dos atletas têm altura superior a 200 cm, e 2,5% têm altura inferior a 192,8 cm. Considere que:

- **A variável Z tem distribuição normal padrão $(Z \sim N(0,1))$;**
- $Prob(Z > 1,64) = 5\%$; **e**
- $Prob(Z > 1,96) = 2,5\%$.

O desvio padrão, em centímetros, dessa distribuição é de, aproximadamente

a) 2

b) 4

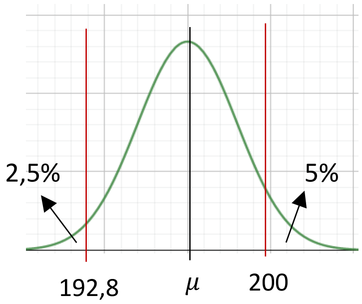
c) 8

d) 16

e) 64

Comentários:

Precisamos calcular os parâmetros dessa distribuição, considerando que 5% da distribuição é superior a 200 e 2,5% é inferior a 192,8, conforme ilustrado a seguir:



O enunciado informa que P(Z > 1,64) = 5%, logo, z = 1,64 está associado a x = 200. Assim, a transformação para a normal padrão para esse valor é:

$$z = \frac{x - \mu}{\sigma}$$

$$1{,}64 = \frac{200 - \mu}{\sigma}$$

$$1{,}64.\sigma = 200 - \mu$$

Como precisamos do desvio padrão, vamos isolar μ:

$$\mu = 200 - 1{,}64.\sigma$$

O enunciado também informa que P(Z > 1,96) = 2,5%. Pela simetria da normal padrão, temos:

$$P(Z < -1{,}96) = 2{,}5\%$$

Logo, z = -1,96 está associado a x = 192,8. Assim, a transformação para a normal padrão para esse valor é:

$$-1{,}96 = \frac{192{,}8 - \mu}{\sigma}$$

$$-1{,}96.\sigma = 192{,}8 - \mu$$

$$\mu = 192{,}8 + 1{,}96.\sigma$$

Igualando ambas as expressões, temos:

$$200 - 1{,}64.\sigma = 192{,}8 + 1{,}96.\sigma$$

$$200 - 192{,}8 = 1{,}96.\sigma + 1{,}64.\sigma$$

$$7{,}2 = 3{,}6.\sigma$$

$$\sigma = \frac{7{,}2}{3{,}6} = 2$$

Gabarito: A

2. (CESGRANRIO/2016 – IBGE) O tempo de atendimento (em minutos) para chamadas de emergência do SAMU no Rio de Janeiro segue uma distribuição normal com média 12 e variância 25.

Qual a probabilidade de que o tempo de atendimento para uma dada chamada exceda a 20 minutos?

a) 1,79%

b) 2,87%

c) 5,48%

d) 25,50%

e) 37,45%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

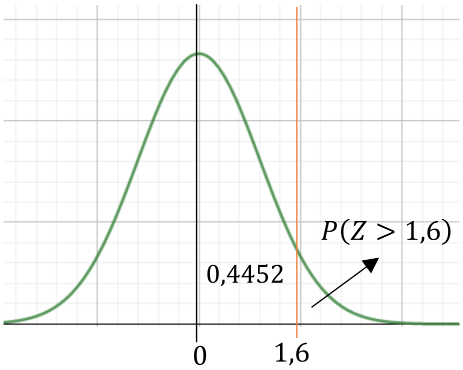
O enunciado informa que a variável segue distribuição normal com média $\mu = 12$ e variância $\sigma^2 = 25$. O desvio padrão é, então:

$$\sigma = \sqrt{\sigma^2} = \sqrt{25} = 5$$

Portanto, para calcular a probabilidade de uma chamada exceder x = 20 minutos, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{20 - 12}{5} = \frac{8}{5} = 1{,}6$$

Pela tabela normal padrão, observamos que P(0 < Z < 1,6) = 0,4452, como ilustrado a seguir.



Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5, então a probabilidade P(Z > 1,6) é:

$$P(Z > 1{,}6) = 0{,}5 - 0{,}4452 = 0{,}0548 = 5{,}48\%$$

Gabarito: C

3. (CESGRANRIO/2012 – Petrobras) Considere que, no verão, o valor das vendas de determinada loja, nas sextas-feiras que sejam dias úteis, tem uma distribuição de probabilidades normal, com média igual a R$ 10.000,00 e desvio padrão de R$ 2.000,00.

Com essa hipótese, a probabilidade de que, em uma dessas sextas-feiras úteis do verão, o valor das vendas seja inferior a R$ 8.000,00 é

a) menor que 20%

b) igual a 20%

c) maior que 20% e menor que 25%

d) igual a 25%

e) maior que 25%

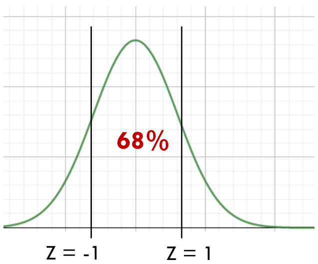
Comentários:

O enunciado informa que a variável segue distribuição normal com média $\mu = 10.000$ reais e desvio padrão $\sigma = 2000$ reais. Então, para calcular a probabilidade de o valor das vendas ser inferior a x = 8000 reais, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{8000 - 10.000}{2000} = \frac{-2000}{2000} = -1$$

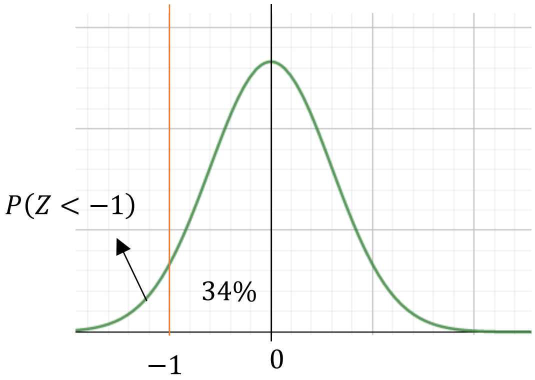
Pela regra empírica, sabemos que:

$$P(-1 < Z < 1) \cong 68\%$$



Pela simetria da normal padrão, a probabilidade P(-1 < Z < 0) é a metade desse valor:

$$P(-1 < Z < 0) \cong \frac{68\%}{2} = 34\%$$



Sabendo que a área à esquerda da média tem probabilidade P(Z < 0) = 50%, então a probabilidade P(Z < -1) é:

$$P(Z < -1) \cong 50\% - 34\% = 16\%$$

Que é menor que 20%.

Gabarito: A

4. (CESGRANRIO/2012 – Innova) Suponha que o tempo de vida de baterias de celular tenha distribuição normal com média de 120 minutos e variância de 100 minutos. Qual é a probabilidade aproximada de uma bateria durar menos que 100 minutos?

a) 0,15%

b) 2,5%

c) 5%

d) 10%

e) 16%

Comentários:

O enunciado informa que a variável segue distribuição normal com média $\mu = 120$ e variância $\sigma^2 = 100$. Logo, o desvio padrão, raiz quadrada da variância é:

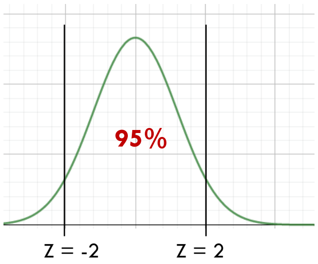
$$\sigma = \sqrt{\sigma^2} = \sqrt{100} = 10$$

Então, para calcular a probabilidade de a bateria durar menos de x = 100 minutos, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{100 - 120}{10} = \frac{-20}{10} = -2$$

Pela regra empírica, sabemos que:

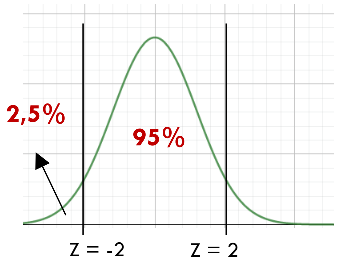
$$P(-2 < Z < 2) \cong 95\%$$



Então, a probabilidade associada aos valores externos ao intervalo (-2; 2) é:

100% - 95% = 5%

Pela simetria da normal padrão, a probabilidade P(Z < -2) é a metade desse valor:



$$P(Z < -2) \cong \frac{5\%}{2} = 2{,}5\%$$

Que é a probabilidade de a bateria durar menos de 100 minutos.

Gabarito: B

5. (CESGRANRIO/2012 – Petrobras) Os salários de técnicos de uma empresa se distribuem normalmente com média de R$ 3.200,00 e desvio padrão de R$ 800,00.

Selecionando-se aleatoriamente dois salários de técnicos dessa empresa, qual a probabilidade de pelo menos um deles ser superior a R$ 3.880,00?

a) 72,25%

b) 15,86%

c) 3,91%

d) 19,77%

e) 35,63%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

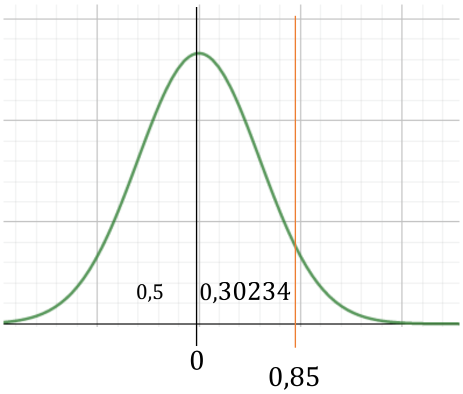
A probabilidade de pelo menos um, dentre os dois salários selecionados aleatoriamente, ser superior a R$ 3.880, pode ser calculada pelo complementar da probabilidade de ambos os salários serem inferiores a esse valor:

$$P(pelo\ menos\ um\ superior)\ =\ 1 - P(ambos\ inferiores)$$

Para calcular a probabilidade de ambos os salários serem inferiores a tal valor, precisamos primeiro calcular a probabilidade de um salário ser inferior. Para isso, o enunciado informa que os salários seguem distribuição normal com média $\mu = 3.200$ e desvio padrão $\sigma = 800$. Portanto, para calcular a probabilidade de um salário ser inferior a x = 3.880 minutos, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x-\mu}{\sigma} = \frac{3880 - 3200}{800} = \frac{680}{800} = 0{,}85$$

Pela tabela normal, observamos que P(0 < Z < 0,85) = 0,30234, como ilustrado a seguir.



Sabendo que a área à esquerda da média tem probabilidade P(Z < 0) = 0,5, então a probabilidade P(Z < 0,85) é:

$$P(Z < 0{,}85) = 0{,}5 + 0{,}30234 = 0{,}80234$$

Então, a probabilidade de os dois salários selecionados aleatoriamente serem inferiores ao valor (interseção dos eventos) é o produto das probabilidades:

$$P(ambos\ inferiores) = 0{,}80234 \times 0{,}80234 \cong 0{,}6437$$

E a probabilidade de ter pelo menos um salário superior ao citado valor é o complementar:

$$P(pelo\ menos\ um\ superior) \cong 1 - 0{,}6437 = 0{,}3563 = 35{,}63\%$$

Gabarito: E

6. (CESGRANRIO/2012 – Petrobras) O peso de lotes produzidos por uma certa indústria segue uma distribuição normal, com média de 10 kg e desvio padrão de 0,2 kg.

Em um lote dessa indústria, selecionado aleatoriamente, qual a probabilidade de o peso do lote não se afastar por mais de 1% do peso médio?

a) 50%

b) 38,30%

c) 17,42%

d) 7,96%

e) 0%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

O enunciado pede a probabilidade de o peso de um lote selecionado aleatoriamente não se afaste da média em mais de 1% = 0,01. Sabendo que a média é de 10kg, a distância máxima desejada é:

$$10 \times 0{,}01 = 0{,}1$$

Então, buscamos a probabilidade de o lote estar no intervalo:

$$P(9{,}9 < X < 10{,}1)$$

Sabendo que a média é $\mu = 10$ e desvio padrão $\sigma = 0{,}2$, calculamos a probabilidade de o lote ser superior x = 10,1, pela seguinte transformação para a curva normal padrão:

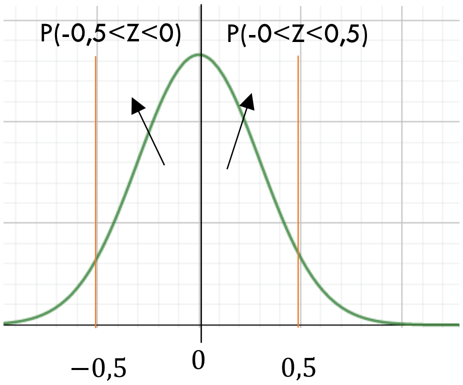
$$z = \frac{x-\mu}{\sigma} = \frac{10{,}1 - 10}{0{,}2} = \frac{0{,}1}{0{,}2} = 0{,}5$$

E a transformação para x = 9,9 é:

$$z = \frac{x-\mu}{\sigma} = \frac{9{,}9 - 10}{0{,}2} = \frac{-0{,}1}{0{,}2} = -0{,}5$$

Logo, a probabilidade buscada é igual a:

$$P(9{,}9 < X < 10{,}1) = P(-0{,}5 < Z < 0{,}5) = P(-0{,}5 < Z < 0) + P(0 < Z < 0{,}5)$$



Pela tabela normal, observamos que P(0 < Z < 0,5) = 0,19146. Pela simetria, temos:

P(0 < Z < 0,5) = P(-0,5 < Z < 0) = 0,19146

Logo:

$$P(-0{,}5 < Z < 0{,}5) = 0{,}19146 + 0{,}19146 = 0{,}38292 \cong 38{,}3\%$$

Gabarito: B

7. (CESGRANRIO/2010 – IBGE) Suponha que as notas dos candidatos de um concurso público, em uma certa prova, sigam distribuição normal com média 7 e desvio padrão 1. A relação candidato/vaga é de 40 para 1. A nota mínima necessária para aprovação nessa prova é

a) 8,65

b) 8,96

c) 9,37

d) 9,58

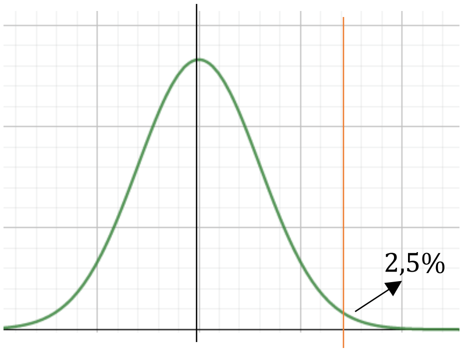
e) 9,75

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

O enunciado informa que as notas em um concurso seguem distribuição normal e que a relação candidato-vaga é de 40 para 1. Então, para que o candidato passe nesse concurso, ele deverá tirar a melhor nota dentre 40 candidatos, ou seja, encaixar-se na seguinte proporção de melhores notas:

$$\frac{1}{40} = 2{,}5\% = 0{,}025$$



Ou seja, precisamos do valor de z tal que P(Z > z) = 0,025. Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5, então precisamos do valor de z tal que:

$$P(0 < Z < z) = 0{,}5 - 0{,}025 = 0{,}475$$

Pela tabela, observamos que z = 1,96, pois P(Z < 1,96) = 0,475.

Considerando que a média das notas é $\mu = 7$ e que o desvio padrão é $\sigma = 1$, utilizamos a seguinte transformação para a normal padrão:

$$z = \frac{x - \mu}{\sigma}$$

$$1{,}96 = \frac{x - 7}{1}$$

$$x - 7 = 1{,}96$$

$$x = 8{,}96$$

Gabarito: B

8. (CESGRANRIO/2010 – IBGE) Seja H a variável aleatória que representa as alturas dos cidadãos de certo país. Sabe-se que H tem distribuição normal com média 1,70 m e desvio padrão 0,04 m. A probabilidade de que um cidadão desse país tenha mais do que 1,75 m de altura é, aproximadamente,

a) 9,9%

b) 10,6%

c) 22,2%

d) 39,4%

e) 40,6%

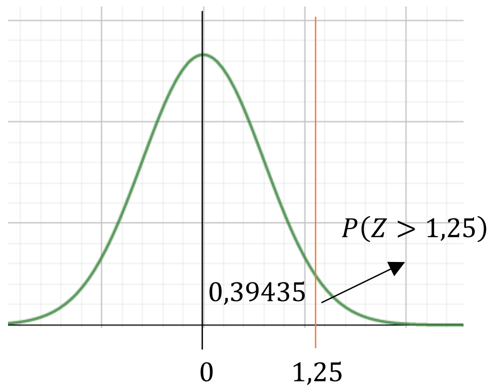
Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

O enunciado informa que a variável segue distribuição normal com média $\mu = 1{,}70$ e desvio padrão $\sigma = 0{,}04$. Então, para calcular a probabilidade de um cidadão ter mais de x = 1,75 m de altura, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{1{,}75 - 1{,}7}{0{,}04} = \frac{0{,}05}{0{,}04} = 1{,}25$$

Pela tabela normal, observamos que P(0 < Z < 1,25) = 0,39435, como ilustrado a seguir.



Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5, então a probabilidade P(Z > 1,25) é:

$$P(Z > 1{,}25) = 0{,}5 - 0{,}39435 = 0{,}10565 \cong 10{,}6\%$$

Gabarito: B

9. (CESGRANRIO/2010 – BACEN) Estima-se que os retornos de um determinado mercado tenham distribuição normal, com média 20% e desvio padrão 10%. A probabilidade de perdas financeiras é de, aproximadamente,

a) 1%

b) 2,5%

c) 5%

d) 10%

e) 20%

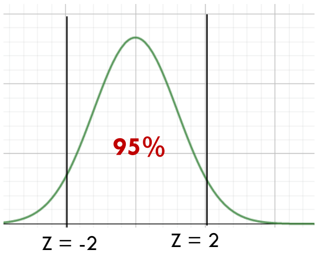
Comentários:

O enunciado informa que a variável segue distribuição normal com média $\mu = 20\% = 0{,}2$ e desvio padrão $\sigma = 10\% = 0{,}1$. Então, para calcular a probabilidade de haver perdas financeiras (retorno menor que x = 0%), usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{0 - 0{,}2}{0{,}1} = \frac{-0{,}2}{0{,}1} = -2$$

Pela regra empírica, sabemos que:

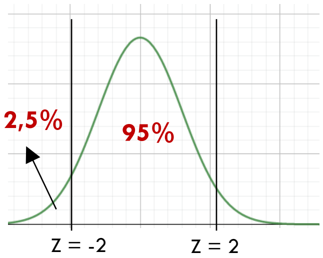
$$P(-2 < Z < 2) \cong 95\%$$



Então, a probabilidade associada aos valores externos ao intervalo (-2; 2) é:

100% - 95% = 5%

Pela simetria da normal padrão, a probabilidade P(Z < -2) é a metade desse valor:



$$P(Z < -2) \cong \frac{5\%}{2} = 2{,}5\%$$

Que é a probabilidade de perdas financeiras.

Gabarito: B

10. (CESGRANRIO/2010 – Petrobras – *Adaptada*) A distribuição normal é uma das mais utilizadas para inferência estatística da probabilidade de ocorrência de diversos fenômenos em engenharia. Nela, qualquer variável aleatória normal X é convertida em uma variável normal padronizada Z, tal que Z = $\frac{X-\mu}{\hat{\sigma}}$, onde $\hat{\sigma}$ é o desvio padrão e μ é a média aritmética. Na distribuição normal padronizada,

a) a variável Z é contínua e representa o número de desvios em relação à média.

b) a variável Z possui média não nula e desvio padrão igual a 1.

c) um ponto da curva da distribuição indica a probabilidade da variável aleatória assumir um valor real entre 0 e 1.

d) valores idênticos acima e abaixo da média têm probabilidades complementares, pois a curva é simétrica.

Comentários:

A questão trata da distribuição normal padrão Z ~ N(0,1).

Em relação à alternativa A, a variável Z é contínua e representa os valores de $\frac{X-\mu}{\hat{\sigma}}$, ou seja, dos desvios em relação à média μ (relativizados pelo desvio padrão). A alternativa A está correta.

Em relação à alternativa B, a variável Z possui média **nula** e desvio padrão igual a 1. A alternativa B está incorreta pois diz que a média é **não nula**.

Em relação à alternativa C, um ponto qualquer da distribuição normal padrão indica a probabilidade de a variável assumir um valor na reta real $(-\infty, \infty)$, e não entre 0 e 1, logo, a alternativa C está incorreta.

Em relação à alternativa D, a curva normal padrão é simétrica em torno de $\mu = 0$, logo valores idênticos acima e abaixo da média possuem as **mesmas** probabilidades. Por exemplo, P(Z < -1) = P(Z > 1). A alternativa D está incorreta.

Gabarito: A

11. (CESGRANRIO/2007 – TCE/RO) O gasto médio dos clientes de um posto de gasolina é uma variável aleatória normal com média R$ 100,00 e desvio padrão R$ 25,00. Os 10% dos que mais consomem recebem um tratamento VIP, incluindo lavagem de carroceria, calibragem nos pneus e verificação do óleo e da água. Quanto você precisa gastar nesse posto de gasolina, em reais, para obter tratamento VIP?

a) 158,00

b) 149,00

c) 141,00

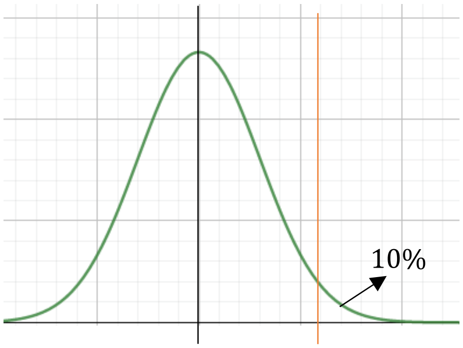
d) 132,00

e) 128,00

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

O enunciado informa que o gasto dos clientes segue distribuição normal e que os 10% que mais consomem recebem tratamento VIP, como ilustrado a seguir:



Ou seja, para sabermos o valor do gasto necessário para obter tratamento VIP, primeiro precisamos do valor de z tal que P(Z > z) = 0,1. Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5, então precisamos do valor de z tal que:

$$P(0 < Z < z) = 0{,}5 - 0{,}1 = 0{,}4$$

Pela tabela, observamos que z = 1,28, pois P(Z < 1,28) = 0,39973 ≅ 0,4.

Considerando que a média dos gastos é $\mu = 100$ e que o desvio padrão é $\sigma = 25$, utilizamos a seguinte transformação para a normal padrão:

$$z = \frac{x - \mu}{\sigma}$$

$$1{,}28 = \frac{x - 100}{25}$$

$$x - 100 = 32$$

$$x = 132$$

Gabarito: D

12. (CESGRANRIO/2006 – EPE) Se uma distribuição segue um padrão normal, é correto afirmar que:

a) 98% dos números estão a dois desvios padrão da média.

b) 95% dos números estão a 1,5 desvio padrão da média.

c) 95% dos números estão a um desvio padrão da média.

d) 86% dos números estão a um desvio padrão da média.

e) 68% dos números estão a um desvio padrão da média.

Comentários:

A questão trabalha com a regra empírica da curva normal. Segundo essa regra, aproximadamente 68% dos números estão a 1 desvio padrão da média; 95% dos números estão a 2 desvios padrão da média; e 99,7% dos números estão a 3 desvios padrão da média.

Gabarito: E

13. (CESGRANRIO/2006 – EPE) Suponha os pesos dos pacotes de arroz normalmente distribuídos com média 1kg e desvio padrão 20g. Escolhendo um pacote ao acaso, qual é a probabilidade de ele pesar mais de 1 030g?

a) 13,4%

b) 11,6%

c) 10,0%

d) 8,4%

e) 6,7%

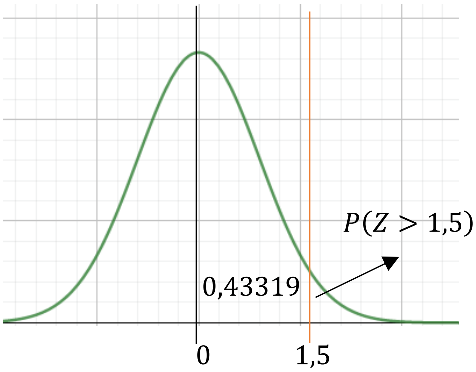
Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Comentários:

O enunciado informa que a variável segue distribuição normal com média $\mu = 1kg = 1000g$ e desvio padrão $\sigma = 20g$. Então, para calcular a probabilidade de um pacote de arroz pesar mais de x = 1030g, usamos a seguinte transformação para a curva normal padrão:

$$z = \frac{x - \mu}{\sigma} = \frac{1030 - 1000}{20} = \frac{30}{20} = 1{,}5$$

Pela tabela normal, observamos que P(0 < Z < 1,5) = 0,43319, como ilustrado a seguir.



Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5, então a probabilidade P(Z > 1,5) é:

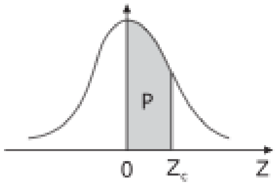
$$P(Z > 1{,}5) = 0{,}5 - 0{,}43319 = 0{,}06681 \cong 6{,}7\%$$

Gabarito: E

Distribuição Normal Padrão

$Z \sim N(0,1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0<Z<Z_c)$



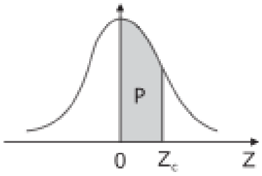
parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

QUESTÕES COMENTADAS – CESGRANRIO

Soma de Variáveis e o Teorema Central do Limite

Para a próxima questão, utilize a tabela normal a seguir, constante na respectiva prova da CESGRANRIO.

Distribuição Normal Padrão
$Z \sim N(0,1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0<Z<Z_c)$



parte inteira e primeira decimal de Z_c	0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

1. (CESGRANRIO/2011 – BNDES) Um produto tem massa normalmente distribuída com média 60 gramas e desvio padrão 8 gramas e é agrupado por dúzias. A probabilidade de a massa de uma dúzia ser superior a 750 gramas é, aproximadamente, de

a) 86%

b) 50%

c) 38%

d) 32%

e) 14%

Comentários:

O enunciado informa que a massa de um produto segue distribuição normal, com média $\mu_0 = 60$ e desvio padrão $\sigma_0 = 8$. Sabendo que o produto é agrupado por dúzias (12), então a média (esperança) do grupo é:

$$E(Dúzia) = 12 \times E(Produto)$$

$$\mu = 12 \times 60 = 720$$

Para calcular o desvio padrão da dúzia, primeiro calculamos a variância. Sabendo que a variância é o quadrado do desvio padrão, então a variância é 8^2 e a variância da dúzia é:

$$V(Dúzia) = 12 \times V(Produto) = 12 \times 8^2$$

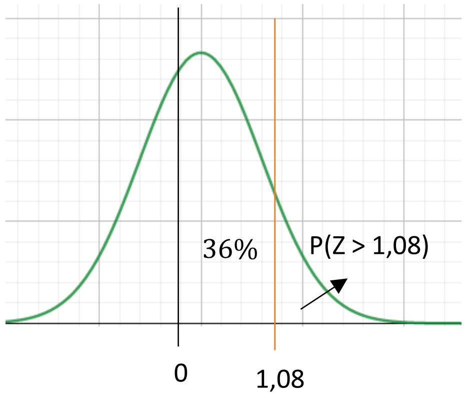
E o desvio padrão da dúzia é a raiz quadrada:

$$\sigma = \sqrt{12 \times 8^2} = 8 \times \sqrt{12} \cong 27{,}7$$

Conhecendo a média e o desvio padrão da dúzia, utilizamos a seguinte transformação para x = 750:

$$z = \frac{x - \mu}{\sigma} \cong \frac{750 - 720}{27{,}7} = \frac{30}{27{,}7} \cong 1{,}08$$

Pela tabela da distribuição normal padrão, observamos que P(0 < Z < 1,08) = 0,36 = 36%, como ilustrado a seguir:



Sabendo que a área à direita da média tem probabilidade P(Z > 0) = 0,5 = 50%, então a probabilidade desejada é:

$$P(Z > 1{,}08) = 50\% - 36\% = 14\%$$

Gabarito: E

LISTA DE QUESTÕES – CESGRANRIO

Distribuição Binomial

1. (CESGRANRIO/2022 – Eletronuclear) Em uma fábrica automobilística, foi notado que os airbags de um certo lote não estavam funcionando tão bem quanto deveriam. Calcula-se que a probabilidade de um airbag desse lote não inflar era de 40%, valor muito acima do limite aceitável. Após o descarte de todo o lote, começou-se a fazer testes de colisão com airbags do lote seguinte.

Supondo-se que os airbags do novo lote possuam a mesma probabilidade de não inflar que os do lote descartado e que, em cada colisão, apenas um airbag é testado, qual a probabilidade de 2 airbags não inflarem nas 5 primeiras colisões?

a) 12,96%

b) 20,16%

c) 31,25%

d) 32,75%

e) 34,56%

2. (CESGRANRIO/2021 – CEF - *Adaptada*) Por estudos estatísticos, estima-se que um cliente de um certo banco tem 75% de probabilidade de ir para atendimento de caixa eletrônico, e 25% de ir para um atendimento personalizado.

Em uma amostra de quatro clientes entrando no banco, qual é a probabilidade de que, no mínimo, a maioria deles se dirija ao atendimento personalizado?

a) 1/64

b) 5/256

c) 3/64

d) 13/256

e) 27/64

3. (CESGRANRIO/2018 – Liquigás) Sabe-se por estudos estatísticos que a eficiência de uma certa vacina para uma dada doença é de 80%. Vacinando-se três indivíduos, qual a probabilidade de que apenas um deles não fique imunizado à doença?

a) 12,8%

b) 16%

c) 32%

d) 36%

e) 38,4%

4. (CESGRANRIO/2016 – IBGE) Uma pesquisa está interessada em estudar os eleitores de determinado candidato. Sabe-se que 50% da população alegam votar no candidato em questão.

Se 6 pessoas forem abordadas aleatoriamente, a probabilidade de que exatamente 3 pessoas sejam eleitoras do candidato em questão é aproximadamente

a) 51%

b) 50%

c) 31%

d) 21%

e) 11%

5. (CESGRANRIO/2012 – Petrobras) 10% dos parafusos produzidos por uma máquina são defeituosos. A probabilidade de que, entre 4 parafusos, pelo menos 3 não sejam defeituosos é de

a) 29,16%

b) 65,61%

c) 94,77%

d) 98,37%

e) 99,99%

6. (CESGRANRIO/2011 – Petrobras) Uma carta tem 2/3 de chances de chegar ao destino correto. Se seis cartas são enviadas de forma independente, a probabilidade de que pelo menos duas cheguem a o destino correto é

a) 4/9

b) 68/81

c) 113/162

d) 230/243

e) 716/729

7. (CESGRANRIO/2009 – BNDES) Em um dado com seis faces numeradas de 1 a 6, a probabilidade de que cada um dos resultados ocorra é a mesma. Esse dado será lançado até que se obtenha o resultado 6. A probabilidade de que isso aconteça em, no máximo, 2 lançamentos é

a) 1/36

b) 5/36

c) 6/36

d) 7/36

e) 11/36

8. (CESGRANRIO/2008 – Petrobras) Um estudante marca, ao acaso, as respostas de um teste de 10 questões de múltipla escolha, com 4 alternativas por questão. O número mais provável de acertos é

a) 1,5

b) 2,0

c) 2,5

d) 3,0

e) 3,5

GABARITO

1. LETRA E
2. LETRA D
3. LETRA E
4. LETRA C
5. LETRA C
6. LETRA E
7. LETRA E
8. LETRA B

LISTA DE QUESTÕES – CESGRANRIO

Distribuição Normal

1. (CESGRANRIO/2023 – BB) A distribuição das alturas dos atletas de vôlei de uma determinada seleção é normal. Sabe-se que 5% dos atletas têm altura superior a 200 cm, e 2,5% têm altura inferior a 192,8 cm.

Considere que:

- **A variável Z tem distribuição normal padrão $(Z \sim N(0,1))$;**
- **$Prob(Z > 1,64) = 5\%$; e**
- **$Prob(Z > 1,96) = 2,5\%$.**

O desvio padrão, em centímetros, dessa distribuição é de, aproximadamente

a) 2

b) 4

c) 8

d) 16

e) 64

2. (CESGRANRIO/2016 – IBGE) O tempo de atendimento (em minutos) para chamadas de emergência do SAMU no Rio de Janeiro segue uma distribuição normal com média 12 e variância 25.

Qual a probabilidade de que o tempo de atendimento para uma dada chamada exceda a 20 minutos?

a) 1,79%

b) 2,87%

c) 5,48%

d) 25,50%

e) 37,45%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

3. (CESGRANRIO/2012 – Petrobras) Considere que, no verão, o valor das vendas de determinada loja, nas sextas-feiras que sejam dias úteis, tem uma distribuição de probabilidades normal, com média igual a R$ 10.000,00 e desvio padrão de R$ 2.000,00.

Com essa hipótese, a probabilidade de que, em uma dessas sextas-feiras úteis do verão, o valor das vendas seja inferior a R$ 8.000,00 é

a) menor que 20%

b) igual a 20%

c) maior que 20% e menor que 25%

d) igual a 25%

e) maior que 25%

4. (CESGRANRIO/2012 – Innova) Suponha que o tempo de vida de baterias de celular tenha distribuição normal com média de 120 minutos e variância de 100 minutos. Qual é a probabilidade aproximada de uma bateria durar menos que 100 minutos?

a) 0,15%

b) 2,5%

c) 5%

d) 10%

e) 16%

5. (CESGRANRIO/2012 – Petrobras) Os salários de técnicos de uma empresa se distribuem normalmente com média de R$ 3.200,00 e desvio padrão de R$ 800,00. Selecionando-se aleatoriamente dois salários de técnicos dessa empresa, qual a probabilidade de pelo menos um deles ser superior a R$ 3.880,00?

a) 72,25%

b) 15,86%

c) 3,91%

d) 19,77%

e) 35,63%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

6. (CESGRANRIO/2012 – Petrobras) O peso de lotes produzidos por uma certa indústria segue uma distribuição normal, com média de 10 kg e desvio padrão de 0,2 kg. Em um lote dessa indústria, selecionado aleatoriamente, qual a probabilidade de o peso do lote não se afastar por mais de 1% do peso médio?

a) 50%

b) 38,30%

c) 17,42%

d) 7,96%

e) 0%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

7. (CESGRANRIO/2010 – IBGE) Suponha que as notas dos candidatos de um concurso público, em uma certa prova, sigam distribuição normal com média 7 e desvio padrão 1. A relação candidato/vaga é de 40 para 1. A nota mínima necessária para aprovação nessa prova é

a) 8,65

b) 8,96

c) 9,37

d) 9,58

e) 9,75

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

8. (CESGRANRIO/2010 – IBGE) Seja H a variável aleatória que representa as alturas dos cidadãos de certo país. Sabe-se que H tem distribuição normal com média 1,70 m e desvio padrão 0,04 m.

A probabilidade de que um cidadão desse país tenha mais do que 1,75 m de altura é, aproximadamente,

a) 9,9%

b) 10,6%

c) 22,2%

d) 39,4%

e) 40,6%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

9. (CESGRANRIO/2010 – BACEN) Estima-se que os retornos de um determinado mercado tenham distribuição normal, com média 20% e desvio padrão 10%.

A probabilidade de perdas financeiras é de, aproximadamente,

a) 1%

b) 2,5%

c) 5%

d) 10%

e) 20%

10. (CESGRANRIO/2010 – Petrobras – *Adaptada*) A distribuição normal é uma das mais utilizadas para inferência estatística da probabilidade de ocorrência de diversos fenômenos em engenharia. Nela, qualquer variável aleatória normal X é convertida em uma variável normal padronizada Z, tal que Z = $\frac{X-\mu}{\hat{\sigma}}$, onde $\hat{\sigma}$ é o desvio padrão e μ é a média aritmética.

Na distribuição normal padronizada,

a) a variável Z é contínua e representa o número de desvios em relação à média.

b) a variável Z possui média não nula e desvio padrão igual a 1.

c) um ponto da curva da distribuição indica a probabilidade da variável aleatória assumir um valor real entre 0 e 1.

d) valores idênticos acima e abaixo da média têm probabilidades complementares, pois a curva é simétrica.

11. (CESGRANRIO/2007 – TCE/RO) O gasto médio dos clientes de um posto de gasolina é uma variável aleatória normal com média R$ 100,00 e desvio padrão R$ 25,00. Os 10% dos que mais consomem recebem um tratamento VIP, incluindo lavagem de carroceria, calibragem nos pneus e verificação do óleo e da água.

Quanto você precisa gastar nesse posto de gasolina, em reais, para obter tratamento VIP?

a) 158,00

b) 149,00

c) 141,00

d) 132,00

e) 128,00

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

12. (CESGRANRIO/2006 – EPE) Se uma distribuição segue um padrão normal, é correto afirmar que:

a) 98% dos números estão a dois desvios padrão da média.

b) 95% dos números estão a 1,5 desvio padrão da média.

c) 95% dos números estão a um desvio padrão da média.

d) 86% dos números estão a um desvio padrão da média.

e) 68% dos números estão a um desvio padrão da média.

13. (CESGRANRIO/2006 – EPE) Suponha os pesos dos pacotes de arroz normalmente distribuídos com média 1kg e desvio padrão 20g. Escolhendo um pacote ao acaso, qual é a probabilidade de ele pesar mais de 1 030g?

a) 13,4%

b) 11,6%

c) 10,0%

d) 8,4%

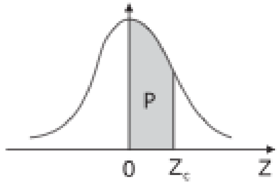
e) 6,7%

Para resolver essa questão, utilize a tabela apresentada ao final desta seção de questões, fornecida na prova.

Distribuição Normal Padrão

$Z \sim N(0,1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0<Z<Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

GABARITO

1. LETRA A
2. LETRA C
3. LETRA A
4. LETRA B
5. LETRA E
6. LETRA B
7. LETRA B
8. LETRA B
9. LETRA B
10. LETRA A
11. LETRA D
12. LETRA E
13. LETRA E

LISTA DE QUESTÕES – CESGRANRIO

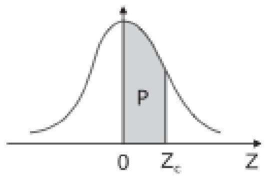
Soma de Variáveis e o Teorema Central do Limite

Para a próxima questão, utilize a tabela normal a seguir, constante na respectiva prova da CESGRANRIO.

Distribuição Normal Padrão

$Z \sim N(0,1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0<Z<Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

1. (CESGRANRIO/2011 – BNDES) Um produto tem massa normalmente distribuída com média 60 gramas e desvio padrão 8 gramas e é agrupado por dúzias. A probabilidade de a massa de uma dúzia ser superior a 750 gramas é, aproximadamente, de

a) 86%

b) 50%

c) 38%

d) 32%

e) 14%

GABARITO

1. LETRA E

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.

1. Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.

2. Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).

3. Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).

4. Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).

5. Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).

6. Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).

7. Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.

8. O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.

Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.