Aula 00

Câmara dos Deputados (Analista Legislativo - Contador) Passo Estratégico de Estatística

Autor:

Allan Maux Santana

01 de Setembro de 2023

Índice

O QUE É O PASSO ESTRATÉGICO?

O Passo Estratégico é um material escrito e enxuto que possui dois objetivos principais:

a) orientar revisões eficientes;

b) destacar os pontos mais importantes e prováveis de serem cobrados em prova.

Assim, o Passo Estratégico pode ser utilizado tanto para **turbinar as revisões dos alunos mais adiantados nas matérias, quanto para maximizar o resultado na reta final de estudos por parte dos alunos que não conseguirão estudar todo o conteúdo do curso regular.**

Em ambas as formas de utilização, como regra, **o aluno precisa utilizar o Passo Estratégico em conjunto com um curso regular completo.**

Isso porque nossa didática é direcionada ao aluno que já possui uma base do conteúdo.

Assim, se você vai utilizar o Passo Estratégico:

a) **como método de revisão**, você precisará de seu curso completo para realizar as leituras indicadas no próprio Passo Estratégico, em complemento ao conteúdo entregue diretamente em nossos relatórios;

b) **como material de reta final**, você precisará de seu curso completo para buscar maiores esclarecimentos sobre alguns pontos do conteúdo que, em nosso relatório, foram eventualmente expostos utilizando uma didática mais avançada que a sua capacidade de compreensão, em razão do seu nível de conhecimento do assunto.

APRESENTAÇÃO

Olá! Sou o professor ***Allan Maux*** e serei o seu analista do ***Passo Estratégico*** nas matérias de ***EXATAS.***

Para que você conheça um pouco sobre mim, segue um resumo da minha ***experiência profissional***, acadêmica e como concurseiro:

*Sou, atualmente, Auditor Fiscal do Município de Petrolina – PE, **aprovado em 2° lugar** no concurso de 2011.*

*Sou formado em **matemática** e **pós-graduado em direito tributário municipal**.*

*Fui, por 05 anos, **Secretário de Fazenda do Município de Petrolina**, período no qual participei da comissão que elaborou o **novo Código Tributário da Cidade, vigente até o momento**, colocando a cidade entre as maiores arrecadações do Estado de Pernambuco.*

Lecionei, também, em cursos preparatórios para o ITA, em Recife-PE.

Fui aprovado e nomeado no concurso para Analista da Receita Federal, em 2012.

Aprovado e nomeado, em 2007, para o cargo de gestor de tributos da Secretaria da Fazenda do Estado de Minas Gerais.

Nossa carreira como Auditor Fiscal de Petrolina é bastante atraente e me fez refletir bastante por sua manutenção, nosso salário inicial beira aos 18k e, final de carreira, passa dos 35k, basicamente, esse salário me fez refletir por aposentar as chuteiras como concurseiro e permanecer no meu Pernambuco.

Atualmente, também, leciono matemática para concursos e vestibulares, presencialmente e com aulas em vídeo.

Estou extremamente feliz de ter a oportunidade de trabalhar na equipe do "Passo", porque tenho convicção de que nossos relatórios e simulados proporcionarão uma preparação diferenciada aos nossos alunos!

Bem, vamos ao que interessa!!

Prof. Allan Maux

ANÁLISE ESTATÍSTICA

Inicialmente, convém destacar os percentuais de incidência de todos os assuntos previstos em nosso curso – quanto maior o percentual de incidência de um determinado assunto, maior será sua importância para nosso certame.

Nossa análise será executada em concursos realizados de *2019 a 2023*, de *ESTATÍSTICA* da banca *FGV.*

- % de cobrança em provas anteriores	
PROBABILIDADE	*34,46%*
ESTATÍSTICA DESCRITIVA	*22,64%*
DISTRIBUIÇÃO DE PROBABILIDADE	*19,93%*
AMOSTRAGEM / ESTIMADORES	*11,82%*
REGRESSÃO LINEAR / CORRELAÇÃO	*8,45%*
ESTIMAÇÃO PONTUAL E INTERVALAR / TESTES DE HIPÓTESES	*2,70%*
TOTAL	*100%*

Vocês perceberão que nos cursos de exatas os perfis das questões das bancas são muito idênticos, portanto, treinem exaustivamente principalmente aquele assunto que possui uma maior incidência em nossa análise e que você tenha mais dificuldade.

Prof. Allan Maux

MEDIDAS DE POSIÇÃO / SEPARATRIZES

Sumário

O que é mais cobrado dentro do assunto

Vamos analisar agora como se comporta a incidência dos sub assuntos da nossa aula de hoje. Assim, você será melhor direcionado nos seus estudos, vejam:

MEDIDAS DE POSIÇÃO	Grau de incidência
MEDIANA	*39,0%*
MÉDIA	*36,2%*
MODA	*20,6%*
SEPARATRIZES	*4,2%*
TOTAL	*100,00%*

ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

A ideia desta seção é apresentar um roteiro para que você realize uma revisão completa do assunto e, ao mesmo tempo, destacar aspectos do conteúdo que merecem atenção.

Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

MEDIDAS DE TENDÊNCIA CENTRAL

Sabemos que essas medidas possuem como objetivo representar da melhor forma possível uma determinada população como um todo. A finalidade principal da Estatística é a tomada de decisões. Portanto, essas medidas de ***tendência central*** precisam ser confiáveis para que sua representatividade esteja muito próxima da realidade.

Iremos estudar as três medidas de tendência central a seguir:

Média, Moda e Mediana

Média Aritmética $\underline{X}$

Sabemos que esse conceito é bem simples e as provas não costuma complicar muito nas questões.

$$\underline{X} = \text{Média Aritmética} = \frac{Soma\ dos\ Valores}{Quantidade\ de\ Valores}$$

Uma observação importante é no cálculo da *Média Aritmética da Dados Agrupados em Classes*. Vamos pegar o nosso exemplo da aula 00. Vejam:

IDADES	FREQUÊNCIA ABSOLUTA (fi)	PONTO MÉDIO (PM)
0 ⊢ 10	12	5 = (0 + 10)/2
10 ⊢ 20	5	15 = (10 + 20)/2
20 ⊢ 30	13	25 = (20 + 30)/2
30 ⊢ 40	5	35 = (30 + 40)/2
40 ⊢ 50	6	45 = (40 + 50)/2
50 ⊢ 60	9	55 = (50 + 60)/2

É bem intuitivo a gente usar o *Ponto Médio* de cada classe como nosso valor de referência para o cálculo da Média Aritmética, correto? Logo, nossa Média Aritmética será dada por:

$$\underline{X} = \text{Média Aritmética} = \frac{12 \cdot 5 + 5 \cdot 15 + 13 \cdot 25 + 5 \cdot 35 + 6 \cdot 45 + 9 \cdot 55}{50} = 28$$

ou seja: *Média Arimética da Dados Agrupados em Classes é dada por:*

$$\underline{X} = \frac{\sum \quad f_i \cdot PM}{n}$$

A galera geralmente não gosta dessa simbologia toda...rsrsrs, muita gente odeia *ESTATÍSTICA* por conta desse excesso de símbolos. Mas, vejam apenas como uma maneira de resumir a fórmula. O importante aqui é você saber como faz, ok?

Pessoal, temos algumas *Propriedades* da *Média Aritmética* que são importantes para a nossa prova, ok?

Se você achar que são muitas propriedades e que não vai conseguir entender todas, então deem atenção, primeiramente, a essas *duas*:

Propriedades Importantes da Média Aritmética	
Ao somar, ou subtrair, uma constante "k" a cada elemento do conjunto, a média será aumentada, ou subtraída, de "k"	*Ao multiplicar, ou dividir, uma constante "k" a cada elemento do conjunto, a média será multiplicada, ou dividida, por "k"*
Dá para perceber que as duas propriedades acima poderiam se resumir a apenas uma?	

AO SOMAR, SUBTRAIR, MULTIPLICAR, OU DIVIDIR, uma constante "k" a cada elemento...

Vejam um exemplo simples:

A ***média*** entre 3 e 5 é ***4***, ok?

Somando ***1*** a cada elemento, temos a ***média*** entre 4 e 6 que é ***5***, ok?

Ou seja, somando 1 a cada elemento da média, a nova média passa a ser a anterior (4) + 1 = ***5***.

O mesmo vale para as demais operações, ok?

CUIDADO: essas propriedades não se aplicam à potenciação e à radiciação.

Uma terceira ***Propriedade*** simples de ser constatada é a da ***Soma dos Desvios***.

Ainda sobre o nosso exemplo anterior, sabemos que a Média foi de 4 para os elementos 3 e 5, ok?

Tomando a média como referência, vemos que:

$\textit{Desvio}_1 = 3 - 4 = -1$

$\textit{Desvio}_2 = 5 - 4 = 1$

Somando os Desvios, eles se anulam: -1 + 1 = 0

FIQUE ATENTO!

Desvio em relação à media nada mais é do que a *diferença* entre cada *elemento* de um conjunto de valores e sua *média aritmética*.

A soma do Desvios será NULA.

Uma quarta Propriedade, obviamente, decorrente do conceito de Média Aritmética, é a seguinte:

A Média entre 3 e 5 é ***4***, ok?

E se adicionarmos mais um elemento a esse conjunto de valores, o que ocorrerá?

Situação 1: Se o elemento for *igual* à média (4), *nada mudará* em relação à Média.

A Média entre {3, 4 e 5} é 4.

Situação 2: Se o elemento for *menor* do que a média, então a *média diminuirá*;

Situação 3: Se o elemento for *maior* do que a média, então a *média aumentará*.

Moda

A ***moda*** é uma medida de *posição* e de *tendência central*. Essa medida mostra o valor que mais se repete em um conjunto de observações. Sendo que, um conjunto de observações pode apresentar mais de uma ou até nenhuma moda. Logo, podemos ter um conjunto *unimodal* (uma moda), *bimodal* (duas modas), *multimodal* (mais de duas modas) e *amodal* (sem moda).

- Moda para Dados Não Agrupados

Quando temos um conjunto de *valores não agrupados*, a moda é o valor que aparece com maior frequência, mas como dito antes, pode haver mais de uma ou até nenhuma moda. Observem os exemplos abaixo.

Amodal	*Todos* os elementos do conjunto apresentam a *mesma* frequência. Ex: {1, 2, 3, 4, 5}
Unimodal	*Um* elemento do conjunto apresenta a maior frequência. Ex: {1, 2, 2, 3, 4, 5}
Bimodal	*Dois* elementos do conjunto com a *mesma* frequência de observações. Ex: {1, 2, 2, 3, 3, 4, 5}
Multimodal	*Três ou mais* elementos do conjunto com a *mesma* frequência de observações. Ex: {1, 2, 2, 3, 3, 4, 5, 5}

- Moda para Dados Agrupados sem intervalo de classe

Quando temos um conjunto de *valores agrupados*, a moda será o valor que apresentar a maior frequência absoluta. Observem a tabela abaixo.

IDADES	FREQUÊNCIA ABSOLUTA (fi)
10	12
20	5
30	13
40	5
50	6
60	9

Vejam que nesse exemplo, a moda será 30 anos, pois é o valor que apresenta a maior frequência absoluta (13).

- Moda para Dados Agrupados em classes

Quando os dados são agrupados em *classes de mesma amplitude*, a moda será o valor da classe que apresenta a maior frequência. Essa classe será chamada de classe modal. Existem alguns métodos de se calcular a moda em dados agrupados em classes. Esses métodos são os seguintes:

a) Moda de Czuber

A moda é calculada através da seguinte expressão:

$$M_o = l_{inf} + \left[\frac{\Delta_a}{\Delta_a + \Delta_p}\right].h$$

Onde:

$\Delta_a = diferença\ anterior = f_{Mo} - f_{ant}$

$f_{Mo} = frequência\ absoluta\ da\ classe\ modal$

$f_{ant} = frequência\ absoluta\ anterior\ a\ classe\ modal$

$\Delta_p = diferença\ posterior = f_{Mo} - f_{post}$

$f_{post} = frequência\ absoluta\ posterior\ a\ classe\ modal$

$h = amplitude\ da\ classe\ modal = l_{sup} - l_{inf}$

$l_{inf} = limite\ inferior\ da\ classe\ modal$

$l_{sup} = limite\ superior\ da\ classe\ modal$

Se a classe modal for a primeira ou a última. A f_{ant} será zero (se for a primeira) ou f_{post} será zero (se for a última).

Exemplo: Qual a moda das Idades de uma grupo de pessoas?

IDADES	FREQUÊNCIA ABSOLUTA (fi)
0 ⊢ 10	12
10 ⊢ 20	5
20 ⊢ 30	13
30 ⊢ 40	5
40 ⊢ 50	6
50 ⊢ 60	9

Nesse exemplo, temos que a frequência absoluta da classe modal é 13. Sabendo disso, a primeira coisa a ser feita é calcular as diferenças anteriores e posteriores.

$$\Delta_a = diferença\ anterior = f_{Mo} - f_{ant} = 13 - 5 = 8$$

$$\Delta_p = diferença\ posterior = f_{Mo} - f_{post} = 13 - 5 = 8$$

Além disso, temos que os limites e amplitude são os seguintes:

$$l_{inf} = 20$$

$$l_{sup} = 30$$

$$h = l_{sup} - l_{inf} = 30 - 20 = 10$$

Aplicando a fórmula de Czuber:

$$M_o = l_{inf} + \left[\frac{\Delta_a}{\Delta_a + \Delta_p}\right].h = 20 + \left[\frac{8}{8+8}\right].10 = 20 + \frac{8}{16}.10 = 20 + 5 = 25$$

b) Moda de King

A moda é calculada da seguinte forma:

$$M_o = l_{inf} + \left[\frac{f_{post}}{f_{ant} + f_{post}}\right].h$$

Vejam que o cálculo da moda de King é semelhante ao da moda de Czuber. Sendo que, ao invés de se utilizar as diferenças anterior e posterior, utilizamos as frequências.

Exemplo: Qual a moda das Idades de um grupo de pessoas?

IDADES	FREQUÊNCIA ABSOLUTA (fi)

0 ⊢ 10	12
10 ⊢ 20	5
20 ⊢ 30	13
30 ⊢ 40	5
40 ⊢ 50	6
50 ⊢ 60	9

Nesse exemplo, temos que a frequência absoluta da classe modal é 13. Temos as seguintes informações:

$f_{ant} = 5$

$f_{post} = 5$

$l_{inf} = 20$

$h = l_{sup} - l_{inf} = 30 - 20 = 10$

Aplicando a fórmula de King, teremos o seguinte:

$$M_o = l_{inf} + \left[\frac{f_{post}}{f_{ant} + f_{post}}\right].h = 20 + \left[\frac{5}{5+5}\right].10 = 20 + \frac{5}{10}.10 = 25$$

Coincidentemente, como as frequências anterior e posterior foram iguais, os valores da moda de King foi igual a moda de Czuber. Mas isso nem sempre ocorre.

Se a classe modal for a primeira ou a última. A f_{ant} será zero (se for a primeira) ou f_{post} será zero (se for a última).

c) Moda de Pearson (*pouco explorado em prova*)

A moda é calculada através de uma *relação empírica*. Para isso, temos que conhecer o valor da *média* $(\underline{X})$ e da *mediana* (M_d). A expressão dessa relação é dada por:

$$M_o = 3 \,.\, M_d - 2 \,.\, \underline{X}$$

A fórmula de Pearson só poderá ser utilizada quando a questão expressamente pedir a *moda de Pearson*.

d) Moda bruta (*raro em prova*)

A moda é obtida através do ponto médio da classe modal, isto é, a média entre limite inferior (l_{inf}) e o limite superior (l_{sup}) da classe modal.

▪ $$M_o = \frac{l_{inf} + l_{sup}}{2}$$

Pessoal, assim como a ***Média Aritmética,*** também temos duas importantes ***Propriedades para Moda.***

Propriedades Importantes da Moda	
Ao somar, ou subtrair, uma constante "k" a cada elemento do conjunto, a moda será aumentada, ou subtraída, de "k"	*Ao multiplicar, ou dividir, uma constante "k" a cada elemento do conjunto, a moda será multiplicada, ou dividida, por "k"*
Dá para perceber que as duas propriedades acima poderiam se resumir a apenas uma? *AO SOMAR, SUBTRAIR, MULTIPLICAR, OU DIVIDIR, uma constante "k" a cada elemento...*	

Mediana

A ***Mediana*** é a medida de posição que se encontra *no centro* de um conjunto de observações. Sendo que, os valores do conjunto devem estar ordenados.

- Mediana para Dados em Rol

Considere os seguintes dados bruto:

1, 2, 9,12, 5, 2, 6, 3, 5, 9, 4

A primeira coisa a ser feita é organizar esses dados em um Rol.

{1, 2, 2, 3, 4, 5, 5, 6, 9, 9, 12 }

Nesse caso, temos um *número ímpar*. E a mediana será exatamente o valor central (5). Mas nem sempre é fácil encontrar esse valor. Logo, podemos encontrar valor central da seguinte forma:

$$\frac{n+1}{2}\ (posição\ da\ mediana)$$

No nosso conjunto de dados temos 11 elementos.

$$\frac{11+1}{2} = \frac{12}{2} = 6°\ posição$$

Se tivermos um *número par* de elementos, a mediana será a média dos dois valores centrais.

{10, 20, 30, 40, 50, 60}

Nesse caso, fazendo a média dos elementos centrais temos que a mediana é 35. Sendo que, também podemos utilizar uma expressão para encontrar os elementos centrais.

Primeira posição central: $\frac{n}{2}$

Segunda posição central: $\frac{n}{2} + 1$

No nosso exemplo, temos o seguinte:

$$\frac{n}{2} = \frac{6}{2} = 3°\ posição$$

$$\frac{n}{2} + 1 = \frac{6}{2} + 1 = 4°\ posição$$

Fazendo a média dessas duas posições encontramos o valor da mediana:

$$M_d = \frac{30+40}{2} = \frac{70}{2} = 35$$

- Mediana para Dados agrupados por valor

Para encontrar a mediana temos que utilizar a frequência acumulada.

IDADES	FREQUÊNCIA ABSOLUTA (fi)	FREQUÊNCIA ACUMULADA (fac)
10	12	12
20	5	17
30	13	30
40	5	35
50	6	41
60	9	50
TOTAL	50	

Para encontrar a posição da mediana, basta utilizar a seguinte expressão:

$$\frac{n}{2}$$

Também precisamos nos preocupar com número par ou ímpar de elementos. Aqui o "n" é 50 (par). Logo,

$$\frac{n}{2} = \frac{50}{2} = 25° \; posição$$

$$\frac{n}{2} + 1 = \frac{50}{2} + 1 = 26° \; posição$$

Portanto, observando a coluna das frequência acumuladas, podemos ver que as posições 25 e 26 correspondem a 30 anos cada. Fazendo a média teremos a mediana igual 30 anos.

$$\frac{30+30}{2} = \frac{60}{2} = 30$$

- Mediana para Dados Agrupados em Classes

Para encontrar a mediana também utilizamos a frequência acumulada.

IDADES	FREQUÊNCIA ABSOLUTA (fi)	FREQUÊNCIA ACUMULADA (fac)
0 ⊢ 10	12	12

10 ⊢ 20	5	17
20 ⊢ 30	13	30
30 ⊢ 40	5	35
40 ⊢ 50	6	41
50 ⊢ 60	9	50
TOTAL	50	

De posse da coluna da frequência acumulada, a primeira coisa a ser feita é encontrar a classe mediana. Para isso, não precisamos nos preocupar com elementos par ou ímpar. Basta utilizar a seguinte expressão:

$$\frac{n}{2}$$

No nosso exemplo, o "n" é igual a 50. Logo, teremos a posição 25 e essa posição fica na classe de 20 a 30. Sabendo a classe modal, basta calcular a mediana através da fórmula.

A fórmula da mediana é a seguinte:

$$M_d = l_{inf} + \left[\frac{\left(\frac{n}{2}\right) - f_{ac_{ant}}}{f_i}\right].h$$

Onde:

$n = número\ de\ elementos$

$l_{inf} = limite\ inferior\ da\ classe\ mediana$

$f_{ac_{ant}} = frequência\ acumulada\ anterior\ à\ classe\ mediana$

$f_i = frequência\ absoluta\ simples\ da\ classe\ mediana$

$h = amplitude\ da\ classe\ mediana = l_{sup} - l_{inf}$

Aplicando a fórmula no exemplo, teremos o seguinte:

$$M_d = l_{inf} + \left[\frac{\left(\frac{n}{2}\right) - f_{ac_{ant}}}{f_i}\right].h = 20 + \left[\frac{\left(\frac{50}{2} - 17\right)}{13}\right].10 = 20 + \left[\frac{25 - 17}{13}\right].10 = 20 + \frac{8}{13}.10$$

$$M_d = 20 + 6{,}15 = 26{,}15$$

Pessoal, assim como a ***Média*** ***Aritmética,*** também temos duas importantes ***Propriedades para Mediana.***

Propriedades Importantes da Mediana	
Ao somar, ou subtrair, uma constante "k" a cada elemento do conjunto, a mediana será aumentada, ou subtraída, de "k"	*Ao multiplicar, ou dividir, uma constante "k" a cada elemento do conjunto, a mediana será multiplicada, ou dividida, por "k"*
Dá para perceber que as duas propriedades acima poderiam se resumir a apenas uma? *AO SOMAR, SUBTRAIR, MULTIPLICAR, OU DIVIDIR, uma constante "k" a cada elemento...*	

Apenas a ***Média Aritmética*** é afetada por *valores extremos*.

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.

Questões FGV

Q.01 (FGV / SEFAZ-ES / 2022)

As notas de nove candidatos num certo exame foram:

54, 48, 46, 51, 38, 50, 44, 58, 32.

A mediana dessas notas é igual a

a) 44.
b) 46.
c) 48.
d) 50.
e) 51.

Comentários:

O erro mais frequente cometido pelo candidato, nas questões que envolvem a mediana, é não organizar os dados brutos em Rol, ou seja, em ordem crescente ou decrescente. Felizmente, nessa questão, o examinador não colocou como opção nas alternativas o 38, senão alguns candidatos mais desatenciosos poderiam marcar essa opção, visto que o *termo central é o 5º*.

*32, 38, 44, 46, **48**, 50, 51, 54, 58*

CONSELHO: Sugiro que o candidato escreva todos os termos e não apenas até o 5° porque você pode esquecer de algum deles e perder a questão.

Gabarito: C

Q.02 (FGV / SEFAZ-ES / 2022)

Uma amostra de idades de usuários de determinado serviço forneceu os seguintes dados:

23; 34; 30; 22; 34; 53; 34; 28; 30; 22

A soma dos valores da média, da moda e da mediana desses dados é igual a

a) 93.
b) 94.
c) 95.
d) 96.
e) 97.

Comentários:

Vamos colocá-los em ordem crescente:

22, 22, 23, 28, 30, 30, 34, 34, 34, 53

Para o cálculo da ***Mediana***, precisaremos calcular a média aritmética entre o 5° e o 6° termos, pois a quantidade de termos é par, logo não existe um termo central.

*22, 22, 23, 28, **30, 30,** 34, 34, 34, 53*

Mediana: 30

Moda: 34 (o termo mais frequente)

Média: 310/10 = 31

A soma da Mediana, da Moda e da Média: 30 + 34 + 31 = 95

Gabarito: C

Q.03 (FGV / Pref. Salvador - BA /2019)

Em uma pequena empresa, a média salarial dos 12 funcionários era de R$2400,00. Lúcio Mauro, que ganhava R$3000,00, se aposentou e para ocupar sua vaga foi contratado Felipe, com um salário de R$1800,00.

Assinale a opção que indica a nova média salarial dos 12 funcionários dessa empresa.

a) *R$2350,00.*
b) *R$2300,00.*
c) *R$2280,00.*
d) *R$2250,00.*
e) *R$2200,00.*

Comentários:

Pessoal, a soma de todos os salários é igual a:

Soma dos Salários = 12 x 2400,00 =R$ 28800,00

Após a aposentadoria de Lúcio Mauro, que ganhava R$ 3000,00, a soma passou a ser de R$ 28.800,00 – R$ 3.000,00 = R$ 25.800,00.

No entanto, entrou Felipe com um salário de R$ 1.800,00 que será somado aos R$ 25.800,00, perfazendo um total de R$ 27.600,00. Logo, a nova média será de:

$$\underline{X} = \frac{27.600,00}{12} = R\$\ 2.300,00$$

Gabarito: B

Q.04 (FGV/Analista de Patologia Clínica (FunSaúde CE) /2021)

Em um conjunto de 12 números, a média de 4 deles é 15 e a média dos outros 8 é 18.

A média dos 12 números é

a) 17.

b) 16,8.

c) 16,5.

d) 16.

e) 15,5.

Comentários:

Pessoal, a banca quer saber a média do conjunto de 12 números. Para tanto, ela dar a média de 4 e 8 deles. Podemos resolver essa questão de duas formas.

Primeira forma:

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

Encontrar o somatório do conjunto de 4.

$\bar{X}_4 = 15$

$$\bar{X}_4 = \frac{X_1 + X_2 + X_3 + X_4}{n}$$

$$15 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

$$X_1 + X_2 + X_3 + X_4 = 15\,.\,4$$

$$X_1 + X_2 + X_3 + X_4 = 60$$

Encontrar o somatório do conjunto de 8.

$\bar{X}_8 = 18$

$$\bar{X}_8 = \frac{X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12}}{n}$$

$$18 = \frac{X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12}}{8}$$

$$X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} = 18\,.\,8$$

$$X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} = 144$$

Por fim, utilizar os dois somatórios para encontras a média do conjunto de 12 números.

$$\bar{X}_{12} = \frac{(X_1 + X_2 + X_3 + X_4) + (X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12})}{n}$$

$$\bar{X}_{12} = \frac{60 + 144}{12} = \frac{204}{12}$$

$$\bar{X}_{12} = 17$$

<u>*Segunda forma:*</u> É a que faria na prova.

Utilizaria o conceito de média global.

$$\bar{X}_G = \frac{n_1\overline{X_1} + n_2\overline{X_2} + \cdots + n_n\overline{X_n}}{n_1 + n_2 + \cdots + n_n}$$

Para nossa questão podemos escrever a equação da seguinte forma:

$$\bar{X}_{12} = \frac{n_4\overline{X_4} + n_8\overline{X_8}}{n_4 + n_8}$$

Onde,

$n_4 = 4$ (conjunto de 4 números);

$n_8 = 8$ (conjunto de 4 números);

$\overline{X_4} = 15$ (média do conjunto de 4 números);

$\overline{X_8} = 18$ (média do conjunto de 8 números);

$$\bar{X}_{12} = \frac{4\,.\,15 + 8\,.\,18}{4 + 8} = \frac{60 + 144}{12} = \frac{204}{12}$$

$$\bar{X}_{12} = 17$$

Gabarito: A

Q.05 (FGV/Auxiliar Técnico Administrativo (IMBEL)/Almoxarife/2021)

A lista a seguir representa a quantidade de itens de certo produto vendidos por uma loja nos 6 dias de certa semana, organizados em ordem crescente:

N, 14, 15, 17, 20, 2N.

Sabe-se que a média desses números é 1 unidade maior que a mediana deles.

O número N é elemento do conjunto

a) {1, 2, 3}.

b) {4, 5, 6}.

c) {7, 8, 9}.

d) {10, 11}.

e) {12, 13}.

Comentários:

Essa questão traz a lista quantidade de itens de certo produto vendidos em 6 dias. Além disso, diz que a lista está em ondem crescente.

N, 14, 15, 17, 20, 2N

Na questão também foi informado que a média é uma unidade maior que a mediana.

$$\bar{X} = M_d + 1$$

Vejam que temos um ROL com um número par de elementos. Logo, temos que encontrar os dois números centrais e fazer a média deles para descobrir a M_d.

N, 14, 15, 17, 20, 2N

$$M_d = \frac{15 + 17}{2} = \frac{32}{2} = 16$$

Agora basta encontrar a média.

$$\bar{X} = 16 + 1 = 17$$

De posse da média podemos encontrar o valor do "N" através do somatório da fórmula da média.

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

$$17 = \frac{N + 14 + 15 + 17 + 20 + 2N}{6}$$

$$3N + 66 = 17 \,.\, 6$$

$$3N = 102 - 66$$

$$3N = 36$$

$$N = 12$$

A questão quer saber qual conjunto o "N" pertence. Portanto, resposta alternativa "E" ***{12, 13}***.

Gabarito: E

Questões CEBRASPE

Q.01 (CESPE - CEBRASPE / Financiadora de Estudos e Projetos / 2009)

Um levantamento efetuado entre os 100 jovens inscritos em um projeto de inclusão social desenvolvido por uma instituição mostra a seguinte distribuição etária.

idade (X, em anos)	frequência
16	40
17	30
18	20
19	10

Com base nessas informações, assinale a opção incorreta.

a) *A mediana da distribuição etária é igual a 17,5 anos.*
b) *A variável X apresentada na tabela de frequências é uma variável discreta.*
c) *A média das idades dos jovens observados no levantamento é igual a 17 anos.*
d) *A moda da distribuição etária é igual a 16 anos.*
e) *Dos jovens inscritos no referido projeto de inclusão social, 30% possuem idades maiores ou iguais a 18 anos.*

Comentários:

Alternativa A: Errada

A mediana da distribuição etária é igual a 17,5 anos.

Nossa população amostral é de 100 elementos, a mediana é a média da posição 50 e 51.

O elemento 50° é 17 e o 51° é 17 então a mediana é:

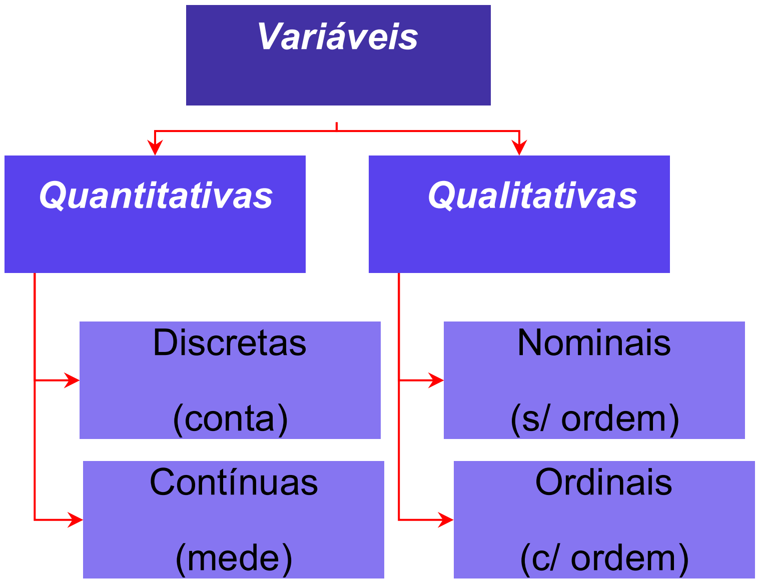
$$Md = \frac{17+17}{2}$$

$$Md = 17 \text{ anos}$$

Alternativa B: Correta

A variável X apresentada na tabela de frequências é uma variável discreta.

Vamos colocar nossa tabelinha para lembrarmos as diferenças entre as variáveis:



A idade além de ser uma variável quantitativa, também, é discreta.

Alternativa C: Correta

A média das idades dos jovens observados no levantamento é igual a 17 anos.

$$\text{Média Aritmética} = \frac{40 \cdot 16 + 30 \cdot 17 + 20 \cdot 18 + 10 \cdot 19}{100} = 17$$

Alternativa D: Correta

A moda da distribuição etária é igual a 16 anos.

idade (*X*, em anos)	frequência
16	40
17	30
18	20
19	10

Observando nossa tabela de frequências absolutas, vemos, facilmente, que a idade 16 anos é a nossa *moda*, visto que é o elemento com a maior frequência.

Alternativa E: Correta

Dos jovens inscritos no referido projeto de inclusão social, 30% possuem idades maiores ou iguais a 18 anos.

Maiores ou iguais a 18 anos são alunos com 18 ou 19 anos, que nesse caso somam (20 + 10) = 30 anos. Logo, 30 de um total de 100, equivale a 30%.

Gabarito: A

Q.02 (CESPE - CEBRASPE / ABIN / 2010)

Sabendo que X é variável aleatória discreta que pode assumir valores inteiros não negativos, julgue o próximo item.

A média de X é não negativa.

C - Certo.

E – Errado.

Comentários:

Não é possível representar quantidade de filhos com números fracionários ou negativos, logo temos uma variável quantitativa discreta. Nessa linha, a média de uma variável discreta não pode assumir valores negativos.

Gabarito: Certa

Q.03 (CESPE - CEBRASPE / Tribunal de Contas - RJ / 2021)

X	frequência absoluta
0	5
1	10
2	20
3	15
total	50

Considerando que a tabela precedente mostra a distribuição de frequências de uma variável quantitativa X, julgue o item a seguir.

A moda e a mediana da variável X são, respectivamente, iguais a 2 e 1,5

C - Certo.

E - Errado.

Comentários:

Pessoal, de imediato, sabemos que, dentre todas as medidas de *tendência central*, a mais simples de ser calcula é a Moda.

Na nossa questão a *Moda (mo) é 2*, visto que é o valor com maior frequência *(20)*, logo, até então, nosso item está correto.

Vamos agora atrás da *Mediana (md)*.

Um *erro* bastante frequente no cálculo da *Mediana (md)* é o de não organizar os dados em ordem crescente ou decrescente, mas que não será nosso caso, visto que as informações já estão na tabela de frequências.

A quantidade de dados é par (50), logo precisaremos determinar a *média aritmética* dos *valores centrais*, ou seja: o *25°* e *26°* termos.

Para o cálculo das posições dos valores centrais, basta dividirmos a quantidade de termos por 2, sendo o primeiro termo central *(25°)*, e, em seguida, para determinar o 2° termo, somar 1 *(26°)*.

Portanto, na tabela de frequências, vamos encontrar o 25° e 26° termos.

Se você não se sente seguro, para fazer a sequencia de forma mental, reescreva os termos até chegar no 25° e 26°, o que vale é ganhar o ponto da questão. Mas, se você está de boa com sua mente, então, para ser mais rápido, vamos fazer assim:

0 – 05 vezes / *1* – 10 vezes / *2* – 20 vezes

Como as variáveis 0 e 1 já somam 15 termos, e há 20 termos na 2, logo o 25° e 26° termos serão o número 2. Logo, sua Md será a Média Aritmética entre eles, *Md = 2*.

Gabarito: Errado

Q.04 (CESPE - CEBRASPE / SEFAZ - DF / 2020)

A partir de uma amostra aleatória simples de tamanho "n", sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue os itens que se seguem.

A distribuição da variável X é simétrica em torno da sua média amostral.

C - Certo.

E - Errado.

Comentários:

A questão faz parte de uma sequência de 4 assertivas. Por isso, precisamos de informações constantes nas questões anteriores.

Considere a seguinte tabela:

Valores de X	*Frequência Absoluta*
-1	*0,2**n***
4	*0,8**n***
Total	***n***

Sabemos que uma distribuição será ***simétrica*** quando a Média ***Aritmética***, a ***Moda*** e a ***Mediana*** forem ***iguais***, ok?

Mas, Allan, não sabemos o valor de "n", e agora?

Meus caros, para ficar mais simples a resolução, devemos sugerir um valor qualquer para "n", lógico que faremos isso de forma que a solução fique mais simples, por isso irei considerar n = 10, ok?

Logo:

Valores de X	Frequência Absoluta
-1	2
4	8
Total	10

- *Mediana* = $\frac{4+4}{2} = 4$
- *Moda* = 4
- *Média* = $\frac{2 \cdot (-1) + 8 \cdot 4}{10} = 3$

Sendo assim, nossa *distribuição NÃO é simétrica*.

Gabarito: ERRADO

Q.05 (CESPE - CEBRASPE/ SEFAZ-DF / 2020)

A partir de uma amostra aleatória simples de tamanho "n", sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue o item que se segue.

A mediana amostral da variável X foi igual a 2,5.

C - Certo.

E - Errado.

Comentários:

A questão faz parte de uma sequência de 4 assertivas. Por isso, precisamos de informações constantes nas questões anteriores.

Considere a seguinte tabela:

Valores de X	Frequência Absoluta
-1	0,2**n**
4	0,8**n**
Total	n

Meus caros, para ficar mais simples a resolução, devemos sugerir um valor qualquer para "n", lógico que faremos isso de forma que a solução fique mais simples, por isso irei considerar n = 10, ok?

Logo:

Valores de X	*Frequência Absoluta*
-1	*2*
4	*8*
Total	*10*

- *Mediana* $= \frac{4+4}{2} = 4$

Gabarito: ERRADO

Q.06 (CEBRASPE / Tribunal de Justiça do Pará / Analista Judiciário / 2019)

Uma amostra aleatória dos registros de furto no município de Abaetetuba, no ano de 2017, apresenta os valores 245, 247, 238, 282 e 261. Uma estimativa não tendenciosa e eficiente para a média de furtos ocorridos em Abaetetuba no ano de 2017, considerando os dados apresentados na amostra, é

a) *238,0.*
b) *254,6.*
c) *260,0.*
d) *282,7.*
e) *308,5.*

Comentários:

Uma questão que trata do assunto Média Aritmética, ok?

$$X = \frac{245 + 247 + 238 + 282 + 261}{5} = \frac{1273}{5} = 254{,}6$$

Pessoal, existe um método prático para efetuarmos uma ***divisão por 5***.

1°: Basta ***dobrar*** o número a ser dividido: 1273 x 2 = ***2546***

2°: Agora, divida o resultado obtido na multiplicação por 10: ***254,6***

Gabarito: B

Q.07 (CEBRASPE / IPHAN / Analista / 2018)

Define-se estatística descritiva como a etapa inicial da análise utilizada para descrever e resumir dados. Em relação às medidas descritivas, julgue o item a seguir.

A mediana é o valor que ocupa a posição central da série de observações de uma variável, dividindo-se o conjunto de valores ordenados em partes assimétricas desiguais.

C - Certo.

E - Errado.

Comentários:

A mediana, conforme vimos em aula, é o valor que ocupa a posição central da série de observações de uma variável. Porém a mesma divide o conjunto de valores em partes simétricas iguais.

Gabarito: Errado

Q.08 (CEBRASPE / IPHAN / Analista / 2018)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

A mediana do conjunto é igual a 3.

C - Certo.

E - Errado.

Comentários:

Meus caros, mediana é o valor que separa a metade maior da metade menor de uma amostra, uma população ou uma distribuição de probabilidade.

A mediana pode ser o valor do meio de um conjunto de dados, caso a quantidade de dados seja um valor ímpar. No entanto, se houver um número par de observações, não há um único valor do meio. Portanto, a mediana é definida como a média dos dois valores centrais.

Como temos 11 observações em nossa questão, a mediana da amostra estará no 6° termo, sendo igual a 3, por representar o termo central.

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

Gabarito: Certo

Q.09 (CEBRASPE / Prof. Pref São Cristóvão /2019)

A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

Idade (em anos)	9	10	11	12	13	14
Quantidade de estudantes	6	22	0	1	0	1

A partir dessa tabela, julgue o item.

Se, em outra turma B, as frequências das idades fossem respectivamente iguais ao dobro das frequências da turma A, então a média aritmética das idades da turma B seria igual ao dobro da média da turma A.

C - CERTO

E - ERRADO

Comentários:

Pessoal, a banca forneceu uma tabela como os valores das idades e a quantidade (frequência) de um grupo de alunos da turma A e questiona sobre a média aritmética de outra turma, a turma B. Para isso, diz que a frequência será o dobro da turma A e que nesse caso a média aritmética de B seria o dobro de A.

A tabela apresentada da questão é a seguinte.

X	f	$f \cdot X$
9	6	54
10	22	220
11	0	0
12	1	12

13	0	0
14	1	14
Total	30	300

Onde "f" é a quantidade de alunos e "X" é a idade. A primeira coisa a ser feita é calcular a média da turma A, para isso temos que construir uma coluna f · X (ver tabela acima) e em seguida fazer o somatório. Temos que a média pode ser obtida da seguinte forma:

$$\underline{X} = \frac{\sum f \cdot X_i}{n}$$

$$\underline{X_A} = \frac{300}{30} = 10$$

De posse dessa média, calculamos a média da turma B (com as frequências em dobro) e fazemos a comparação para saber se será o dobro como afirma a banca.

X	f	f · X
9	12	108
10	44	440
11	0	0
12	2	24
13	0	0
14	4	28
Total	60	600

$$\underline{X_B} = \frac{600}{60} = 10$$

Vejam que as médias são iguais. Nessa questão nem precisaríamos fazer conta alguma, pois se a frequência for dobrada em consequência será dobrada a quantidade de alunos e o valor de f · X. Desta forma, os valores se anulam e as médias ficam iguais. Portanto, errada a questão.

A banca tentou enganar o candidato com as propriedades da média. Como sabemos a média é afetada pelas quatros operações. Se a banca afirmasse que ao invés da frequência fosse dobrada as idades, aí sim teríamos a média das idades em dobro. Senão vejamos.

Tabela com as idades em dobro:

X	f	f · X
18	6	108
20	22	440

22	0	0
24	1	24
26	0	0
28	1	28
Total	30	600

$$\underline{X} = \frac{600}{30} = 20$$

Gabarito: Errado

Questões FCC

Q.01 (FCC / TRT – 18ª / Oficial de Justiça / 2023)

Em uma turma de 60 alunos, 10 foram reprovados. Sabendo-se que a média dos alunos aprovados foi 8,5 e a média dos alunos reprovados foi de 3,4, a média da turma foi

a) *8,35*
b) *7,65*
c) *7,95*
d) *6,95*
e) *7,05*

Comentários:

- Total de Alunos: 60
- Reprovados: 10
- Aprovados: 50

Para determinarmos a média da turma basta somarmos todas as notas dos aprovados com as dos reprovados e dividirmos pelo total de alunos.

Apesar de não termos as notas de cada um, o enunciado nos informou as respectivas médias, logo:

- Somatório dos ***Aprovados:*** 50 x 8,5 = ***425***
- Somatório dos ***Reprovados:*** 10 x 3,4 = ***34***

Logo, a **média** será dada por:

$$\bar{X} = \frac{425 + 34}{60} = 7{,}65$$

Gabarito: B

Q.02 (FCC / COPERGÁS / Analista / 2023)

Quatro trabalhadores executam uma tarefa em tempos diferentes. Os tempos gastos para realizar essa tarefa foram 1h35min, 1h40min, 1h33min e 1h43min. Um novo trabalhador, sabendo do tempo de seus colegas, garante que o tempo médio para realizar essa tarefa será de 1h35min com a sua participação. O tempo desse novo trabalhador é

a) *1h22min.*
b) *1h34min.*
c) *1h24min.*
d) *1h20min.*
e) *1h30min*

Comentários:

Tempo do novo trabalhador: "x"

Vamos transformar todos os intervalos de tempo para minutos, assim ficará mais fácil:

$$\bar{X} = \frac{95 + 100 + 93 + 103 + X}{5} = 95$$

$$391 + X = 95 \cdot 5$$

$$X = 475 - 391$$

X = 84 minutos que equivale a 1h24min

Gabarito: C

Q.03 (FCC / Procuradoria – PGE-AM / Analista / 2022)

Uma ginasta executa três vezes uma determinada prova. Suas notas, na primeira e segunda tentativas foram, respectivamente, metade e dois terços da nota da terceira tentativa. A média aritmética das notas das três tentativas foi de 32,5 pontos. A nota da primeira prova foi

a) *20,5 pontos.*
b) *30,0 pontos.*
c) *22,5 pontos.*
d) *45,0 pontos.*

e) *20,0 pontos.*

Comentários:

A ideia aqui é representar a terceira nota por uma incógnita, ok?

Geralmente a chamamos de "X". Mas, dessa vez, vamos fazer diferente:

- ***3ª Nota:*** 6X
- ***1ª Nota:*** 1/2 de 6X = 3X
- ***2ª Nota:*** 2/3 de 6X = 4X

Ao determinarmos o valor de "6X" para a 3ª nota evitamos frações indesejáveis para efetuarmos as nossas contas, beleza?

$$\bar{X} = \frac{6X + 3X + 4X}{3} = 32{,}5$$

$$13\ X = 3 \cdot 32{,}5$$

$$13\ X = 97{,}5$$

$$X = 7{,}5$$

A NOTA DA 1ª PRIMEIRA PROVA É DADA POR:

$$= 3X =$$

$$= 3 \cdot 7{,}5 =$$

$$= 22{,}5 =$$

Gabarito: C

LISTA DE QUESTÕES ESTRATÉGICAS

Questões FGV

Q.01 (FGV / SEFAZ-ES / 2022)

As notas de nove candidatos num certo exame foram:

54, 48, 46, 51, 38, 50, 44, 58, 32.

A mediana dessas notas é igual a

a) 44.
b) 46.
c) 48.
d) 50.
e) 51.

Q.02 (FGV / SEFAZ-ES / 2022)

Uma amostra de idades de usuários de determinado serviço forneceu os seguintes dados:

23; 34; 30; 22; 34; 53; 34; 28; 30; 22

A soma dos valores da média, da moda e da mediana desses dados é igual a

a) 93.
b) 94.
c) 95.
d) 96.
e) 97.

Q.03 (FGV / Pref. Salvador - BA /2019)

Em uma pequena empresa, a média salarial dos 12 funcionários era de R$2400,00. Lúcio Mauro, que ganhava R$3000,00, se aposentou e para ocupar sua vaga foi contratado Felipe, com um salário de R$1800,00.

Assinale a opção que indica a nova média salarial dos 12 funcionários dessa empresa.

a) R$2350,00.
b) R$2300,00.
c) R$2280,00.
d) R$2250,00.
e) R$2200,00.

Q.04 (FGV/Analista de Patologia Clínica (FunSaúde CE) /2021)

Em um conjunto de 12 números, a média de 4 deles é 15 e a média dos outros 8 é 18.

A média dos 12 números é

a) 17.
b) 16,8.
c) 16,5.
d) 16.
e) 15,5.

Q.05 (FGV/Auxiliar Técnico Administrativo (IMBEL)/Almoxarife/2021)

A lista a seguir representa a quantidade de itens de certo produto vendidos por uma loja nos 6 dias de certa semana, organizados em ordem crescente:

N, 14, 15, 17, 20, 2N.

Sabe-se que a média desses números é 1 unidade maior que a mediana deles.

O número N é elemento do conjunto

a) {1, 2, 3}.

b) {4, 5, 6}.

c) {7, 8, 9}.

d) {10, 11}.

e) {12, 13}.

Gabarito - FGV

1	2	3	4	5
C	C	A	C	E

Questões CEBRASPE

Q.01 (CESPE - CEBRASPE / Financiadora de Estudos e Projetos / 2009)

Um levantamento efetuado entre os 100 jovens inscritos em um projeto de inclusão social desenvolvido por uma instituição mostra a seguinte distribuição etária.

idade (*X*, em anos)	frequência
16	40
17	30
18	20
19	10

Com base nessas informações, assinale a opção incorreta.

a) *A mediana da distribuição etária é igual a 17,5 anos.*
b) *A variável X apresentada na tabela de frequências é uma variável discreta.*
c) *A média das idades dos jovens observados no levantamento é igual a 17 anos.*
d) *A moda da distribuição etária é igual a 16 anos.*
e) *Dos jovens inscritos no referido projeto de inclusão social, 30% possuem idades maiores ou iguais a 18 anos.*

Q.02 (CESPE - CEBRASPE / ABIN / 2010)

Sabendo que X é variável aleatória discreta que pode assumir valores inteiros não negativos, julgue o próximo item.

A média de X é não negativa.

C - Certo.

E – Errado.

Q.03 (CESPE - CEBRASPE / Tribunal de Contas - RJ / 2021)

X	frequência absoluta
0	5
1	10
2	20
3	15
total	50

Considerando que a tabela precedente mostra a distribuição de frequências de uma variável quantitativa X, julgue o item a seguir.

A moda e a mediana da variável X são, respectivamente, iguais a 2 e 1,5

C - Certo.

E - Errado.

Q.04 (CESPE - CEBRASPE / SEFAZ - DF / 2020)

A partir de uma amostra aleatória simples de tamanho "n", sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue os itens que se seguem.

A distribuição da variável X é simétrica em torno da sua média amostral.

C - Certo.

E - Errado.

Q.05 (CESPE - CEBRASPE/ SEFAZ-DF / 2020)

A partir de uma amostra aleatória simples de tamanho "n", sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue o item que se segue.

A mediana amostral da variável X foi igual a 2,5.

C - Certo.

E - Errado.

Q.06 (CEBRASPE / Tribunal de Justiça do Pará / Analista Judiciário / 2019)

Uma amostra aleatória dos registros de furto no município de Abaetetuba, no ano de 2017, apresenta os valores 245, 247, 238, 282 e 261. Uma estimativa não tendenciosa e eficiente para a média de furtos ocorridos em Abaetetuba no ano de 2017, considerando os dados apresentados na amostra, é

a) *238,0.*
b) *254,6.*
c) *260,0.*
d) *282,7.*
e) *308,5.*

Q.07 (CEBRASPE / IPHAN / Analista / 2018)

Define-se estatística descritiva como a etapa inicial da análise utilizada para descrever e resumir dados. Em relação às medidas descritivas, julgue o item a seguir.

A mediana é o valor que ocupa a posição central da série de observações de uma variável, dividindo-se o conjunto de valores ordenados em partes assimétricas desiguais.

C - Certo.

E - Errado.

Q.08 (CEBRASPE / IPHAN / Analista / 2018)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

A mediana do conjunto é igual a 3.

C - Certo.

E - Errado.

Q.09 (CEBRASPE / Prof. Pref São Cristóvão /2019)

A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

Idade (em anos)	9	10	11	12	13	14
Quantidade de estudantes	6	22	0	1	0	1

A partir dessa tabela, julgue o item.

Se, em outra turma B, as frequências das idades fossem respectivamente iguais ao dobro das frequências da turma A, então a média aritmética das idades da turma B seria igual ao dobro da média da turma A.

C - CERTO

E - ERRADO

Gabarito - CEBRASPE

1	2	3	4	5	6	7	8	9
E	E	B	E	C	E	B	A	E

Questões FCC

Q.01 (FCC / TRT – 18ª / Oficial de Justiça / 2023)

Em uma turma de 60 alunos, 10 foram reprovados. Sabendo-se que a média dos alunos aprovados foi 8,5 e a média dos alunos reprovados foi de 3,4, a média da turma foi

a) 8,35
b) 7,65
c) 7,95
d) 6,95
e) 7,05

Q.02 (FCC / COPERGÁS / Analista / 2023)

Quatro trabalhadores executam uma tarefa em tempos diferentes. Os tempos gastos para realizar essa tarefa foram 1h35min, 1h40min, 1h33min e 1h43min. Um novo trabalhador, sabendo do tempo de seus colegas, garante que o tempo médio para realizar essa tarefa será de 1h35min com a sua participação. O tempo desse novo trabalhador é

a) 1h22min.
b) 1h34min.
c) 1h24min.
d) 1h20min.
e) 1h30min

Q.03 (FCC / Procuradoria – PGE-AM / Analista / 2022)

Uma ginasta executa três vezes uma determinada prova. Suas notas, na primeira e segunda tentativas foram, respectivamente, metade e dois terços da nota da terceira tentativa. A média aritmética das notas das três tentativas foi de 32,5 pontos. A nota da primeira prova foi

a) 20,5 pontos.
b) 30,0 pontos.
c) 22,5 pontos.
d) 45,0 pontos.
e) 20,0 pontos.

Gabarito - FCC

1	2	3
B	C	C

MEDIDAS SEPARATRIZES

Sabemos que as ***medidas separatrizes*** servem para dividir o conjunto de dados ordenados (crescente ou decrescente) em duas ou mais partes. O nome da medida separatriz é definido de acordo com a quantidade de partes em que o conjunto de dados é dividido.

As principais medidas separatrizes as seguintes: mediana, quartis, decis e percentis.

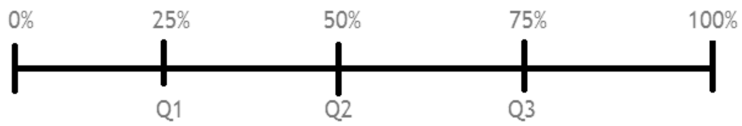
Iremos estudar as três medidas de separatrizes a seguir:

Quartil, Decil e Percentil

A ***mediana*** já foi estudada em aula anterior.

Quartil

Os **quartis** dividem os dados em *quatro partes* de mesma frequência. Desta forma, teremos sempre *três quartis* e cada um deles corresponderá a **25%** do conjunto de dados. Observem a figura abaixo.



Observações importantes:

1) Q_1 = *primeiro quartil* = valor situado de tal forma que uma quarta parte (25%) dos dados é menor que ele e as três quartas partes restantes (75%) são os maiores;

2) Q_2 = *segundo quartil* = coincide com a mediana e separa os 50% menores dos 50% maiores;

3) Q_3 = *terceiro quartil* = valor situado de tal forma que as três quartas partes (75%) dos dados é menor que ele e uma quarta parte restante (25%) é maior;

4) Q_1 será a mediana dos valores que sobraram à esquerda da mediana (Q_2);

5) Q_3 será a mediana dos valores que sobraram à direita da mediana (Q_2);

6) A amplitude interquartílica é dada por: $Q_3 - Q_1$

7) A amplitude semi-interquartílica ou desvio quartílico é dado por: $\frac{Q_3 - Q_1}{2}$

Considere o seguinte o conjunto de dados {1, 2, 2, 3, 4, 5, 5, 6, 9, 9, 12}. Calcule os quartis?

Primeira forma: calculando primeiro a mediana e depois disso utilizar as observações 4 e 5.

Nesse caso, temos um *número ímpar*. Logo, podemos encontrar valor central da seguinte forma:

$$\frac{n+1}{2}\ (posição\ da\ mediana)$$

No nosso conjunto de dados temos 11 elementos.

$$\frac{11+1}{2} = \frac{12}{2} = 6°\ posição$$

{1, 2, 2, 3, 4, **5**, 5, 6, 9, 9, 12}

Portanto, a mediana é 5 e consequentemente Q_2 também será **5**.

Utilizando a observação (4) podemos encontrar Q_1, achando a mediana dos elementos que sobraram à esquerda mediana que encontramos. Nesse caso, temos 5 elementos no conjunto de dados.

{1, 2, 2, 3, 4, ~~5, 5, 6, 9, 9, 12~~}

Temos um *número ímpar* novamente.

$$\frac{5+1}{2} = \frac{6}{2} = 3°\ posição$$

{1, 2, **2**, 3, 4}

Portanto, Q_1 será **2**.

Utilizando a observação (5) podemos encontrar Q_3, achando a mediana dos elementos que sobraram à direita mediana que encontramos. Neste caso, temos 5 elementos no conjunto de dados.

{~~1, 2, 2, 3, 4, 5,~~ 5, 6, 9, 9, 12}

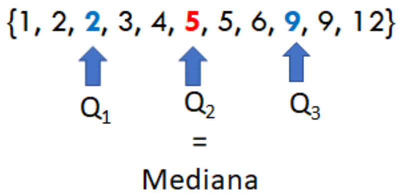
Temos um *número ímpar* novamente.

$$\frac{5+1}{2} = \frac{6}{2} = 3°\ posição$$

{5, 6, 9, 9, 12}

Portanto, Q_3 será 9.

Desta forma, para Rol em análise temos as seguintes quantis.



Observação: se o número de elementos for par, temos que utilizar as seguintes expressões para encontrar a mediana.

Primeira posição central: $\frac{n}{2}$

Segunda posição central: $\frac{n}{2} + 1$

Depois é só fazer a média das duas posições.

Segunda forma: aqui utilizamos a seguinte expressão para calcular a posição dos quatis:

$$P_{Qk} = \frac{k}{4} . n$$

Onde:

k pode assumir o valor 1, 2 ou 3. Vai depender do quartil que se deseja calcular.

n é o número de elementos do conjunto de dados

Pessoal, se o valor da posição não der um valor exato, temos que arredondar para cima.

Considere o seguinte o conjunto de dados {1, 2, 2, 3, 4, 5, 5, 6, 9, 9, 12}. Calcule a posição dos quantis?

1) Cálculo de P_{Q1}

$$P_{Q1} = \frac{1}{4}\,.11 = 2{,}75 \;\approx 3°\,posição$$

2) Cálculo de P_{Q2}

$$P_{Q2} = \frac{2}{4}\,.11 = 5{,}5 \;\approx 6°\,posição$$

2) Cálculo de P_{Q3}

$$P_{Q3} = \frac{3}{4}\,.11 = 8{,}25 \;\approx 9°\,posição$$

Superada essa parte, temos que aprender a calcular os quartis para dados agrupados e agrupados em classes. A forma de se calcular é semelhante ao cálculo da mediana (vista em aula anterior).

- Quartil para Dados agrupados por valor

Para encontrar os quartis temos que utilizar a frequência acumulada.

idades	frequência Absoluta (fi)	Frequência Acumulada (fac)
10	12	12
20	5	17
30	13	30
40	5	35
50	6	41
60	9	50
Total	**50**	

Para encontrar a posição dos quartis, basta utilizar a seguinte expressão:

$$P_{Qk} = \frac{k}{4}\,.n$$

Onde:

k pode assumir o valor 1, 2 ou 3. Vai depender do quartil que se deseja calcular.

n é o número de elementos do conjunto de dados.

Para Q_2 temos o seguinte:

$$P_{Q1} = \frac{1}{4}.50 = 12{,}5 \approx 13°posição$$

Portanto, observando a coluna das frequências acumuladas, podemos ver que a posição 13 corresponde a 20 anos. Logo, Q_1 corresponde a 20 anos.

Para Q_2 temos o seguinte:

$$P_{Q2} = \frac{2}{4}.50 = 25°\ posição$$

Portanto, observando a coluna das frequências acumuladas, podemos ver que a posição 25 corresponde a 30 anos. Logo, Q_2 corresponde a 30 anos.

Para Q_3 temos o seguinte:

$$P_{Q3} = \frac{3}{4}.50 = 37{,}5 \approx 38°\ posição$$

Portanto, observando a coluna das frequências acumuladas, podemos ver que a posição 38 corresponde a 50 anos. Logo, Q_2 corresponde a 50 anos.

- Quartil para Dados Agrupados em Classes

Para encontrar os quartis também utilizamos a frequência acumulada.

idades	frequência Absoluta (fi)	Frequência Acumulada (fac)
0 ⊢ 10	12	12
10 ⊢ 20	5	17
20 ⊢ 30	13	30
30 ⊢ 40	5	35
40 ⊢ 50	6	41
50 ⊢ 60	9	50
Total	**50**	

De posse da coluna das frequências acumulada, a primeira coisa a ser feita é encontrar a posição da classe do quartil desejado. Para isso, basta utilizar a seguinte expressão:

$$P_{Qk} = \frac{k}{4}.n$$

A fórmula do quartil é a seguinte:

$$Q_k = l_{inf} + \left[\frac{\left(\frac{k}{4} \cdot n\right) - f_{ac_{ant}}}{f_i}\right] . h$$

Onde:

$n = número\ de\ elementos$

$l_{inf} = limite\ inferior\ da\ classe\ do\ quartil$

$f_{ac_{ant}} = frequência\ acumulada\ anterior\ à\ classe\ do\ quartil$

$f_i = frequência\ absoluta\ simples\ da\ classe\ do\ quartil$

$h = amplitude\ da\ classe\ do\ quartil = l_{sup} - l_{inf}$

$k = (1, 2\ ou\ 3), vai\ depender\ do\ quartil\ desejado$

Para exemplificar, iremos calcular o Q_2. No nosso exemplo, o "n" é igual a 50 e "h" é 10.

A posição de Q_2 será a seguinte:

$$P_{Q2} = \frac{2}{4} . 50 = 25°\ posição$$

Desta forma, a classe do Q_2 fica entre 20 e 30. Sendo, portanto, 13 o valor da frequência absoluta simples e 17 a frequência acumulada anterior a classe do quartil. Aplicando a fórmula, ficamos com o seguinte:

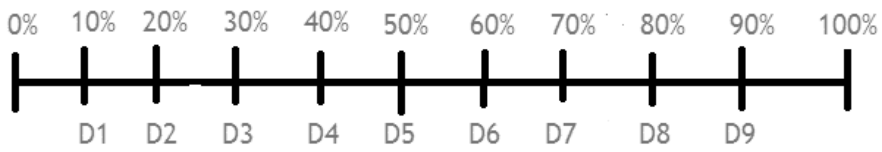
$$Q_2 = l_{inf} + \left[\frac{\left(\frac{2}{4} . n\right) - f_{ac_{ant}}}{f_i}\right] . h = 20 + \left[\frac{\left(\frac{2}{4} . 50 - 17\right)}{13}\right] . 10 = 20 + \left[\frac{25 - 17}{13}\right] . 10Q$$

$$Q_2 = 20 + \frac{8}{13} . 10$$

$$M_d = 20 + 6{,}15 = 26{,}15$$

Decil

Os ***decis*** dividem os dados em *dez partes* de mesma frequência. Desta forma, teremos sempre *nove decis* e cada um deles corresponderá a **10%** do conjunto de dados. Observem a figura abaixo.



D_5 = *quinto decil* = coincide com a mediana e separa os 50% menores dos 50% maiores;

- Decil para Dados em Rol ou Agrupados

Os cálculos dos decis são semelhantes aos dos quartis. Logo, para calcular a posição do decil, utilizamos a seguinte expressão:

$$P_{Dk} = \frac{k}{10} . n$$

Onde:

k pode assumir o valor 1, 2, 3, ..., 9. Vai depender do decil que se deseja calcular.

n é o número de elementos do conjunto de dados.

- Quartil para Dados agrupados em classe

Os cálculos são semelhantes aos do quantis. Sendo a posição do decil encontrada pela seguinte expressão:

$$P_{Dk} = \frac{k}{10} . n$$

k pode assumir o valor 1, 2, 3, ..., 9. Vai depender do decil que se deseja calcular.

n é o número de elementos do conjunto de dados.

De posse, da posição do decil podemos localizar a classe do decil e com isso calcular ele através da seguinte fórmula:

$$D_k = l_{inf} + \left[\frac{\left(\frac{k}{10} \cdot n\right) - f_{ac_{ant}}}{f_i} \right] . h$$

Onde:

$n = número\ de\ elementos$

$l_{inf} = limite\ inferior\ da\ classe\ do\ decil$

$f_{ac_{ant}} = frequência\ acumulada\ anterior\ à\ classe\ do\ decil$

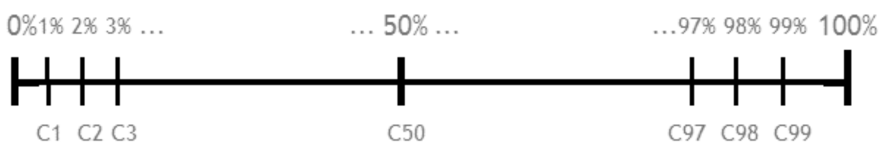
$f_i = frequência\ absoluta\ simples\ da\ classe\ do\ decil$

$h = amplitude\ da\ classe\ do\ decil = l_{sup} - l_{inf}$

$k = (1, 2, 3, \ldots, 9), vai\ depender\ do\ decil\ desejado$

Percentil

Os ***percentis (ou centil)*** dividem os dados em *100 partes* de mesma frequência. Desta forma, teremos sempre *noventa e nove percentis* e cada um deles corresponderá a **1%** do conjunto de dados. Observem a figura abaixo.



P$_{50}$ = *quinquagésimo decil* = coincide com a mediana e separa os 50% menores dos 50% maiores;

- Percentil para Dados em Rol ou Agrupados

Os cálculos dos decis são semelhantes aos dos quartis. Logo, para calcular a posição do decil, utilizamos a seguinte expressão:

$$P_{Pk} = \frac{k}{100} . n$$

Onde:

k pode assumir o valor 1, 2, 3, ..., 99. Vai depender do decil que se deseja calcular.

n é o número de elementos do conjunto de dados.

- Percentil para Dados agrupados em classe

Os cálculos são semelhantes aos do quantis. Sendo a posição do decil encontrada pela seguinte expressão:

$$P_{Pk} = \frac{k}{100} . n$$

k pode assumir o valor 1, 2, 3, ..., 99. Vai depender do decil que se deseja calcular.

n é o número de elementos do conjunto de dados.

De posse, da posição do decil podemos localizar a classe do decil e com isso calcular ele através da seguinte fórmula:

$$P_k = l_{inf} + \left[\frac{\left(\frac{k}{100} . n \right) - f_{ac_{ant}}}{f_i} \right] . h$$

Onde:

$n = número\ de\ elementos$

$l_{inf} = limite\ inferior\ da\ classe\ do\ percentil$

$f_{ac_{ant}} = frequência\ acumulada\ anterior\ à\ classe\ do\ percentil$

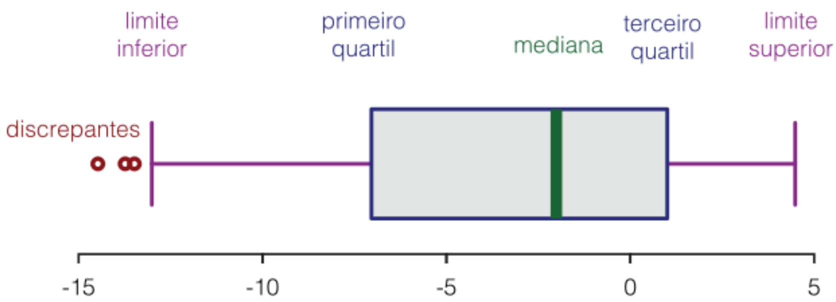
$f_i = frequência\ absoluta\ simples\ da\ classe\ do\ percentil$

$h = amplitude\ da\ classe\ do\ percentil = l_{sup} - l_{inf}$

$k = (1, 2, 3, \ldots, 99), vai\ depender\ do\ percentil\ desejado$

BOX PLOT

Pessoal, o *box plot* (diagrama de caixas) é um gráfico que utiliza os quartis para representar o conjunto de dados. Sua representação pode ser tanto horizontal como vertical. Observem a figura abaixo.



De posse de uma box plot, podemos encontrar as seguintes medidas:

- Limite inferior = valor mínimo;
- Limite superior = valor máximo;
- Valores discrepantes = outliers;
- Os quatis (Q_1, Q_2=M_d e Q_3);
- Amplitude interquartílica = $Q_3 - Q_1$
- Amplitude semi- interquartílica ou desvio quartílico = $\frac{Q_3 - Q_1}{2}$

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.

Q.01 (IBFC/Supervisor de Pesquisas (IBGE)/Suporte Gerencial/2021)

Marcos pretende determinar a mediana referente aos dados brutos coletados e relacionados abaixo:

23 - 22 - 21 - 22 - 32 - 33

41 - 21 - 20 - 32 - 42 - 38

De acordo com os dados, o resultado encontrado por Marcos é igual a:

a) 37.

b) 33.

c) 41.

d) 27,5.

e) 28.

Comentários:

Pessoal, aqui a banca dar dados brutos é pede a mediana. A primeira coisa a ser feita é organizar os dados em um Rol.

{20, 21, 21, 22, 22, 23, 32, 32, 33, 38, 41, 42}

Temos 12 elementos nesse Rol, logo um número par de elementos. Desta forma, temos que encontrar as duas posições centrais.

Primeira posição central: $\frac{n}{2} = \frac{12}{2} = 6°\ posição$

Segunda posição central: $\frac{n}{2} + 1 = \frac{12}{2} + 1 = 7°\ posição$

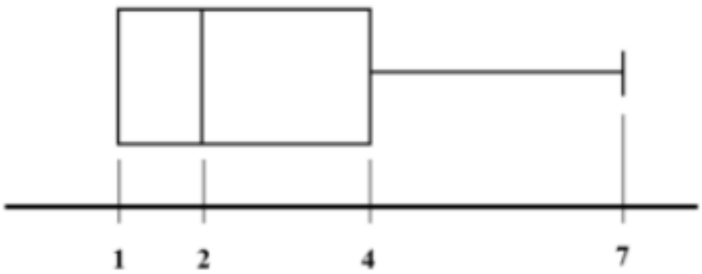
Portanto, a mediana será a média dessas duas posições centrais.

{20, 21, 21, 22, 22, 23, 32, 32, 33, 38, 41, 42}

$$M_d = \frac{23 + 32}{2} = \frac{55}{2} = 27{,}5$$

Gabarito: D

Q.02 (CEBRASPE - Analista de Controle Externo (TCE-RJ)/Controle Externo/Tecnologia da Informação/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

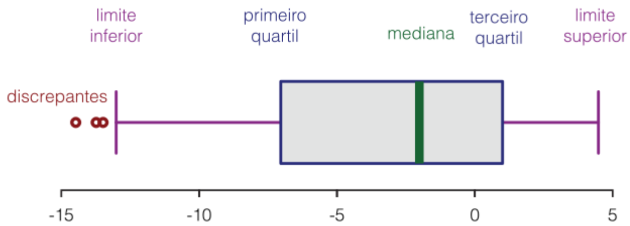
A mediana da variável X é igual a 4.

C – CERTO

E - ERRADO

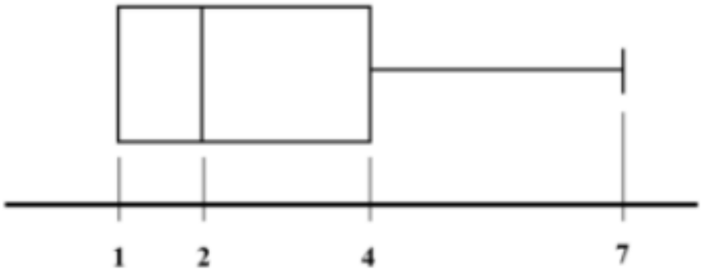
Comentários:

Pessoal, nessa questão temos que conhecer a estrutura do box plot. Conforme estudado na aula, podemos facilmente observar que está errado. Pois, o valor da mediana corresponde a 2. O valor 4 corresponde ao Q_3. Façam uma comparação da figura apresentada pela banca com a que vimos na aula.



Gabarito: Errado

Q.03 (CEBRASPE - Analista de Controle Externo (TCE-RJ)/Controle Externo/Tecnologia da Informação/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

O diagrama boxplot indica que o intervalo interquartil (ou interquartílico) da distribuição da variável X é igual a 3.

C – CERTO

E - ERRADO

Comentários:

A amplitude interquartílica é dada pela diferença entre Q_3 e Q_1. Logo,

Q1 = 1

Q3 = 4

amplitude interquartílica = $Q_3 - Q_1 = 4 - 1 = 3$

Portanto, correto o item.

Gabarito: Certo

Q.04 (COMPERVE (UFRN) - Profissional Analista Superior (CRECI 17 RN)/Agente Fiscal/2021)

O quadro abaixo descreve os cinco imóveis comerciais de um mesmo dono, mostrando os valores, em reais, dos aluguéis de apenas quatro deles.

Imóvel 1	Imóvel 2	Imóvel 3	Imóvel 4	Imóvel 5
2.500	4.300	1.250	1.700	x

Se a mediana dos valores dos aluguéis desses cinco imóveis é 1,7 mil reais, conclui-se que

a) o valor do aluguel do Imóvel 5 é 3 mil reais.

b) o Imóvel 5 é aquele cujo aluguel tem maior valor.

c) o valor do aluguel do Imóvel 5 é, no máximo, 1,7 mil reais.

d) o aluguel do Imóvel 5 custa mais de 2,5 mil reais.

Comentários:

Pessoal, nessa questão a banca dá um quadro com os valores de aluguéis de 5 imóveis e quer saber o valor do aluguel do imóvel 5. Além disso, diz que o valor da média é 1.700. Com base nessas informações, podemos organizar os dados da seguinte forma:

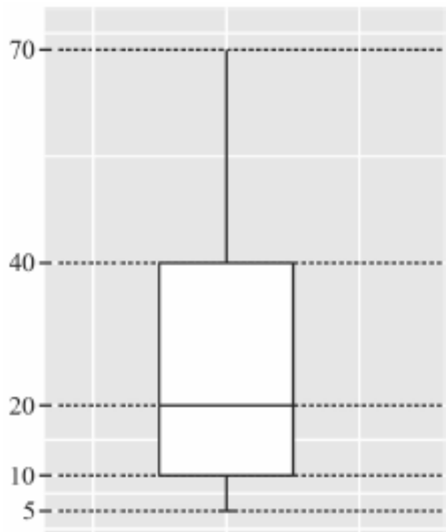
		1.700	2.500	4.300
		Medina = imóvel 4	Imóvel 1	Imóvel 2

Vejam que com as informações que temos não podemos afirmar que o imóvel 3 (1.250) irá ocupar a posição 1 ou 2.

Fazendo uma análise nas alternativas, podemos observar que a única resposta possível é a Letra C, pois o valor do imóvel 5, com certeza, será um valor, no máximo, igual a 1.700.

Gabarito: C

Q.05 (CEBRASPE (CESPE) - Analista Judiciário (TJ PA)/Estatística/2020)



Considerando que o desenho esquemático (boxplot) antecedente se refere a uma variável quantitativa X, assinale a opção correta.

a) O intervalo interquartil é igual a 65.

b) Metade da distribuição da variável X se encontra entre os valores 20 e 40.

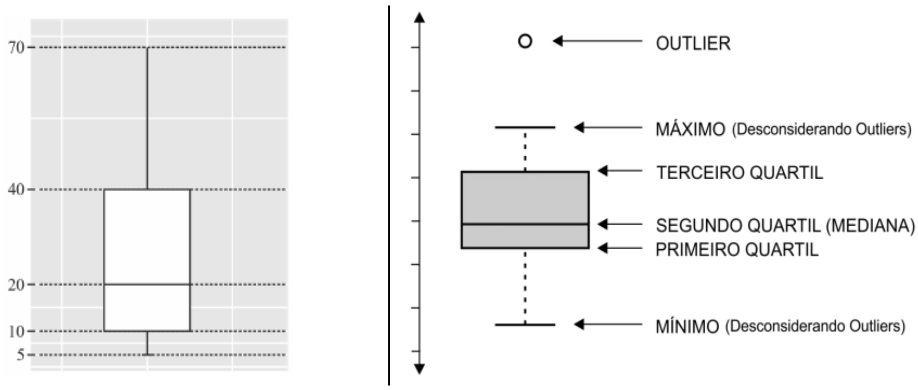
c) Os valores da variável X que se encontram no intervalo [5;10] representam 5% da distribuição de X.

d) A mediana de X é igual a 25.

e) O primeiro quartil da distribuição de X é igual a 10.

Comentários:

Pessoal, nessa questão iremos analisar cada alternativa tendo como base o boxplot dado. Além disso, essa questão exige o conhecimento da estrutura de uma boxplot. Logo, iremos comparar a gráfico dado na questão como o gráfico abaixo:



Letra (A) – Errada

A amplitude interquartílica é dada pela diferença entre Q_3 e Q_1. Logo,

Q1 = 10

Q3 = 40

amplitude interquartílica = $Q_3 - Q_1 = 40 - 10 = 30$

Letra (B) – Errada, pois como a mediana é 20, metade encontra-se entre 5 e 20 e outra metade entre 20 e 70.

Letra (C) – Errada. Pois, entre 5 e 10 temos 25% dos dados. Nesse intervalo, temos os valores inferiores ao Q_1.

Letra (D) – Errada. Podemos observar que a mediana é 20. $M_d = Q_2$.

Letra (E) – Correta. Pois, o Q_1 é igual a 10.

Gabarito: E

Q.06 (CEBRASPE / Prefeitura de Imbé / 2019)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

O valor do primeiro quartil do conjunto de dados (Q1/4) é igual a 3.

C – CERTO

E - ERRADO

Comentários:

Md = $\frac{11+1}{2}$ = 6° termo {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}

O elemento de 6ª posição é o número 3, o qual será o nosso *2° quartil*.

Logo, o nosso primeiro quartil será representado pelo número 2:

{*1, 2, 2, 3, 3, 3*, 4, 4, 4, 4, 4}

Ou:

Posição do quartil inferior (Q_1) = $\frac{n+1}{4}$

O conjunto de dados do enunciado é:

{1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}

O conjunto possui 11 elementos, portanto n = 11. Assim,

Posição do quartil inferior (Q_1) = $\frac{11+1}{4}$ = *3° termo* {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}

Portanto, Q_1 é o elemento cuja posição no conjunto (ordenado) é 3.

Logo, Q(1/4) = 2

Gabarito: Errado

Q.07 (CEBRASPE / IPHAN / Analista / 2018)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

O valor do terceiro quartil do conjunto de dados (Q3/4) é igual a 4.

C – CERTO

E - ERRADO

Comentários:

Pessoal, se você conhece o conceito de Quartil, não será necessário nem fazer conta.

Percebe-se, visualmente, que o (Q3/4) será o número 4, ok?

Mas, vamos à solução técnica:

O segundo quartil é a mediana da sequência apresentada.

Sequência = {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}

Como o número de elementos é ímpar (11), para descobrir a posição do elemento que será a mediana, adicionamos 1 e dividimos o resultado por dois.

Md = $11+12$ **= 6° termo {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}**

O elemento de 6ª posição é o número 3, o qual será o nosso *2° quartil*.

O 3° quartil pode ser obtido pela mediana dos números que estiverem à direita do 2° quartil.

{1, 2, 2, 3, 3, 3, *4, 4, 4, 4, 4*}

Como há 5 elementos:

Md = $5+12$ **= 3 {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}**

Logo, Q (3/4) = 4

Gabarito: Certo

LISTA DE QUESTÕES ESTRATÉGICAS

Q.01 (IBFC/Supervisor de Pesquisas (IBGE) /Suporte Gerencial/2021)

Marcos pretende determinar a mediana referente aos dados brutos coletados e relacionados abaixo:

23 - 22 - 21 - 22 - 32 - 33

41 - 21 - 20 - 32 - 42 - 38

De acordo com os dados, o resultado encontrado por Marcos é igual a:

a) 37.

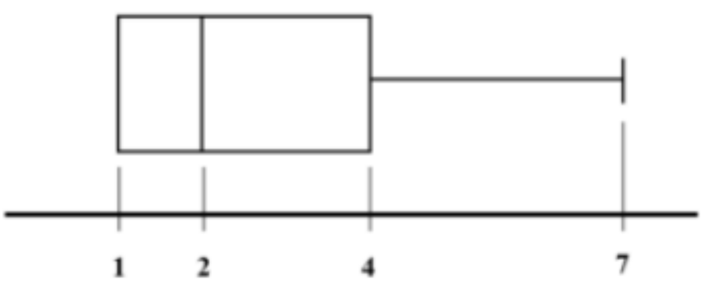
b) 33.

c) 41.

d) 27,5.

e) 28.

Q.02 (CEBRASPE - Analista de Controle Externo (TCE-RJ)/Controle Externo/Tecnologia da Informação/2021)



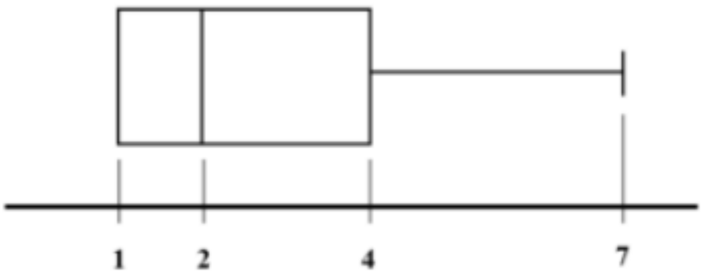
Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

A mediana da variável X é igual a 4.

C – CERTO

E - ERRADO

Q.03 (CEBRASPE - Analista de Controle Externo (TCE-RJ)/Controle Externo/Tecnologia da Informação/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

O diagrama boxplot indica que o intervalo interquartil (ou interquartílico) da distribuição da variável X é igual a 3.

C – CERTO

E - ERRADO

Q.04 (COMPERVE (UFRN) - Profissional Analista Superior (CRECI 17 RN)/Agente Fiscal/2021)

O quadro abaixo descreve os cinco imóveis comerciais de um mesmo dono, mostrando os valores, em reais, dos aluguéis de apenas quatro deles.

Imóvel 1	Imóvel 2	Imóvel 3	Imóvel 4	Imóvel 5
2.500	4.300	1.250	1.700	x

Se a mediana dos valores dos aluguéis desses cinco imóveis é 1,7 mil reais, conclui-se que

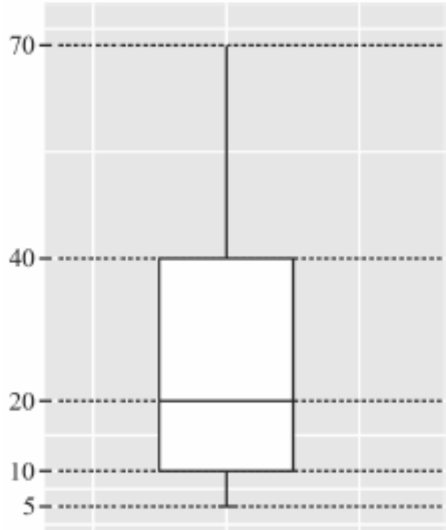
a) o valor do aluguel do Imóvel 5 é 3 mil reais.

b) o Imóvel 5 é aquele cujo aluguel tem maior valor.

c) o valor do aluguel do Imóvel 5 é, no máximo, 1,7 mil reais.

d) o aluguel do Imóvel 5 custa mais de 2,5 mil reais.

Q.05 (CEBRASPE (CESPE) - Analista Judiciário (TJ PA)/Estatística/2020)



Considerando que o desenho esquemático (boxplot) antecedente se refere a uma variável quantitativa X, assinale a opção correta.

a) O intervalo interquartil é igual a 65.

b) Metade da distribuição da variável X se encontra entre os valores 20 e 40.

c) Os valores da variável X que se encontram no intervalo [5;10] representam 5% da distribuição de X.

d) A mediana de X é igual a 25.

e) O primeiro quartil da distribuição de X é igual a 10.

Q.06 (CEBRASPE / Prefeitura de Imbé / 2019)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

O valor do primeiro quartil do conjunto de dados (Q1/4) é igual a 3.

C – CERTO

E – ERRADO

Q.07 (CEBRASPE / IPHAN / Analista / 2018)

Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

O valor do terceiro quartil do conjunto de dados (Q3/4) é igual a 4.

C – CERTO

E - ERRADO

Gabarito

1	2	3	4	5	6	7	8	9	10
D	E	C	C	E	E	C	*	*	*

Variáveis Aleatórias Discretas

Sumário

ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

A ideia desta seção é apresentar um roteiro para que você realize uma revisão completa do assunto e, ao mesmo tempo, destacar aspectos do conteúdo que merecem atenção.

Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

VARIÁVEIS ALEATÓRIAS DISCRETAS

O estudo das ***variáveis aleatórias*** é fundamental na estatística inferencial. Essas variáveis podem ser discretas ou contínuas. As discretas podem assumir apenas certos valores, e são fruto da contagem. Já a contínua resulta de uma medida e pode assumir qualquer valor dentro de um intervalo.

Na aula de hoje, iremos estudar a ***variável aleatória discreta***. Esse tipo de variável estar associada a uma distribuição de probabilidade.

Esperança Matemática

A ***esperança matemática*** é também chamada de valor médio ou simplesmente média. Sua representação é dada por:

$$E(X) = \mu = \sum_{i=1}^{n} X_i \,.\, P(X_i)$$

Portanto, a esperança de uma variável aleatória discreta é dada pela multiplicação de cada valor da variável pela sua probabilidade. Depois disso, basta fazer o somatório.

Para exemplificar o cálculo da esperança, iremos utilizar o exemplo clássico de um lançamento de um dado honesto. Um dado honesto tem 6 faces e as probabilidades são iguais a 1/6.

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

X_i	$P(X_i)$	$X_i \,.\, P(X_i)$
1	1/6	$1 \,.\, \frac{1}{6} = \frac{1}{6}$

2	1/6	$2\,.\frac{1}{6}=\frac{2}{6}$
3	1/6	$3\,.\frac{1}{6}=\frac{3}{6}$
4	1/6	$4\,.\frac{1}{6}=\frac{4}{6}$
5	1/6	$4\,.\frac{1}{6}=\frac{5}{6}$
6	1/6	$6\,.\frac{1}{6}=\frac{6}{6}$
Total	1	21/6

Portanto, a E(X) será

$$E(X)=\mu=\sum_{i=1}^{n} X_i\,.\,P(X_i)=\frac{21}{6}$$

Propriedades da Esperança Matemática

As propriedades descritas a seguir são aplicadas tanto para variáveis aleatórias discretas quanto para contínua. O conhecimento dessas propriedades é fundamental para resolver de forma rápida uma questão. Além disso, elas podem ser utilizadas de maneira conjunta.

1) $E(k\,.\,X)=k\,.E(X)$

ou

$E\left(\frac{X}{k}\right)=\frac{E(X)}{k}$

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

ou

$E(X+k)=E(X)-k$

As propriedades (1) e (2) são iguais às vistas quando estudamos as propriedades da média aritmética.

- Se for multiplicada/dividida uma variável aleatória "k", a esperança ficará multiplicada/dividida por "k".
- Se for somada/subtraída uma variável aleatória "k", a esperança ficará somada/subtraída a "k".

Exemplo!

Considere a seguinte tabela:

X	P(X)	X.P(X)
1	0,25	0,25
2	0,10	0,20
3	0,35	1,05
4	0,30	1,20
Total	**1,00**	**2,70**

$$E(X) = \sum_{i=1}^{n} X_i \, . \, P(X_i) = 2{,}70$$

a) Se for multiplicada uma variável aleatória "**2**" a E(X) também ficará multiplicada por essa variável.

$$E(k \, . \, X) = k \, . \, E(X) = 2 \, . \, 2{,}70 = 5{,}40$$

Provando:

X	P(X)	X.P(X)
1.2=2	0,25	0,50
2.2=4	0,10	0,40
3.2=6	0,35	2,10
4.2=8	0,30	2,40
Total	**1,00**	**5,40**

$$E(X) = \sum_{i=1}^{n} X_i \, . \, P(X_i) = 5{,}40$$

b) Se for adicionada uma variável aleatória "**2**" a E(x) também ficará somada a essa variável.

$$E(X + k) = E(X) + k = 2{,}70 + 2 = 4{,}70$$

Provando:

X	P(X)	X.P(X)
1+2=3	0,25	0,75
2+2=4	0,10	0,40
3+2=5	0,35	1,75

4+2=6	0,30	1,80
Total	**1,00**	**4,70**

$$E(X) = \sum_{i=1}^{n} \quad X_i \, . \, P(X_i) = 4{,}70$$

3) $E(X + Y) = E(X) + E(Y)$

ou

$E(X - Y) = E(X) - E(Y)$

A esperança da soma/subtração de duas variáveis é igual a soma/subtração das esperanças das variáveis.

Exemplo!

Considere as seguintes tabelas: Observem que as variáveis possuem a mesma probabilidade.

X	P(X)	X.P(X)
1	0,25	0,25
2	0,10	0,20
3	0,35	1,05
4	0,30	1,20
Total	**1,00**	**2,70**

$$E(X) = \sum_{i=1}^{n} \quad X_i \, . \, P(X_i) = 2{,}70$$

Y	P(Y)	Y.P(Y)
5	0,25	1,25
6	0,10	0,60
7	0,35	2,45
8	0,30	2,40
Total	**1,00**	**6,70**

$$E(Y) = \sum_{i=1}^{n} \quad Y_i \, . \, P(Y_i) = 6{,}70$$

Se for feita a soma das variáveis a esperança será igual a soma das esperanças das variáveis individualmente.

$E(X) = 2{,}70$

$E(Y) = 6{,}70$

$$E(X + Y) = E(X) + E(Y) = 2{,}70 + 6{,}70 = 9{,}40$$

Provando:

(X+Y)	P(X)	X.P(X)
1+5=6	0,25	1,50
2+6=8	0,10	0,80
3+7=10	0,35	3,50
4+8=12	0,30	3,60
Total	**1,00**	**9,40**

$$E(X) = \sum_{i=1}^{n} X_i \,.\, P(X_i) = 9{,}40$$

4) $E(k) = k$

A esperança de uma constante é a própria constante.

5) Se X e Y são variáveis aleatórias independentes, então $E(X.Y) = E(X) \,.\, E(Y)$

Essa propriedade só vale se as variáveis forem independentes. Porém, o fato de se saber $E(X.Y) = E(X) \,.\, E(Y)$ não é garantia que as variáveis são independentes. Isto é, elas podem ser dependentes ou independentes.

Moda

Sabemos que a ***moda*** é o valor com maior frequência em uma distribuição. Fazendo uma analogia para a distribuição de probabilidade, a ***moda*** será o maior valor de probabilidade. Mas, se os valores das probabilidades forem iguais, temos uma distribuição ***amodal***.

Uma moda, **5**. Pois, a maior probabilidade foi 0,30 ou 30%.

X_i	$P(X_i)$
1	0,10
2	0,25
3	0,15
4	0,03
5	0,30
6	0,17

Duas modas, **2** e **5**. Pois, a maior probabilidade foi 0,25 ou 25%.

X_i	$P(X_i)$
1	0,10
2	0,25
3	0,15
4	0,13
5	0,25
6	0,12

Amodal, pois todas as probabilidades foram iguais.

X_i	$P(X_i)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Total	1

Total	1

Total	1

Função de Distribuição Acumulada

A ***função de distribuição acumulada*** de uma variável aleatória discreta mostra a probabilidade acumulada de todos os valores menores ou iguais a determinado valor. Essa função é dada por:

$$F(X) = P(X \leq x)$$

Considerem a seguinte distribuição de frequência:

x	P(X)
1	0,25
2	0,10
3	0,35
4	0,30

$F(1) = P(X \leq 1) = P(X = 1) = 0{,}25$

$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0{,}25 + 0{,}10 = 0{,}35$

$F(3) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0{,}25 + 0{,}10 + 0{,}35 = 0{,}70$

$F(4) = P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0{,}25 + 0{,}10 + 0{,}35 + 0{,}30 = 1$

Portanto, se for incluída uma coluna para a frequência acumulada, a tabela ficara da seguinte forma:

x	$P(X)$	$F(X) = P(X \leq x)$
1	0,25	0,25
2	0,10	0,35
3	0,35	0,70
4	0,30	1,00

Nas questões de prova a frequência de distribuição acumulada vem da seguinte forma:

$$F(x) = \begin{cases} 0, se\ \ x < 1 \\ 0{,}25, se\ 1 \leq x < 2 \\ 0{,}35, se\ 2 \leq x < 3 \\ 0{,}70, se\ 3 \leq x < 4 \\ 1, se\ x \leq 4 \end{cases}$$

Mediana

A ***mediana***, como já sabemos, divide a distribuição em duas partes iguais. Portanto, na distribuição acumulada de frequência ela é igual a ***50%***. Além disso, temos os quartis que dividem a destruição em quatro partes iguais de ***25%*** cada. O que dá origem ao Q_1, Q_2 e Q_3.

Na ***função de distribuição acumulada*** de uma variável aleatória discreta, podemos encontrar os valores de X para os quais a função de distribuição acumulada apresenta valores maiores ou iguais aos percentuais da mediana e quartis.

$$F(x_{Q1}) \geq 25\%$$

$$F(x_{Q2}) = F(x_{mediana}) \geq 50\%$$

$$F(x_{Q3}) \geq 75\%$$

No cálculo tanto a <u>mediana</u> quanto os <u>quartis</u> temos que calcular a coluna de distribuição acumulada de probabilidade. Sendo que, existiram casos em que teremos exatamente a porcentagem desses parâmetros e outros em que não teremos essa coincidência. Como definimos acima, a $F(x) \geq \%$, logo temos duas formas de encontrar a mediana e os quartis:

Se tivermos exatamente a %, teremos que fazer a média do valor de X correspondente a % e o valor do próximo X;

Se não tivermos exatamente a %, basta considerar o valor da mediana ou quartis o valor de X que ultrapassar a %.

Como os exemplos que faremos a seguir ficará melhor essa observação.

Exemplo (1): Considere a seguinte distribuição:

x	$P(X)$	$F(X) = P(X \leq x)$
1	0,20	0,20
2	0,15	0,35
3	0,35	0,70
4	0,30	1,00

$M_d = Q_2 = 2$

$Q_1 = 3$

$Q_3 = 4$

Vejam que nesse exemplo, na coluna das frequências acumuladas ultrapassaram os valores dos parâmetros que queríamos encontrar.

Exemplo (2): Considere a seguinte distribuição:

x	$P(X)$	$F(X) = P(X \leq x)$
1	0,15	0,15
2	0,10	0,25
3	0,25	0,50
4	0,30	0,80
5	0,20	1,00

$M_d = Q_2 = \frac{3+4}{2} = \frac{7}{2} = 3{,}5$

$Q_1 = \frac{1+2}{2} = \frac{3}{2} = 1{,}5$

$Q_3 = 4$

Nesse exemplo, as porcentagens da mediana/Q_2 e Q_1 foram exatamente as suas respectivas porcentagens, logo tivemos que fazer a média do valor de X correspondente a porcentagem e o X próximo.

Variância e Desvio Padrão

A ***variância*** é uma medida de dispersão que mede o grau de dispersão da distribuição em torno da média. Desta foram, a variância de uma variável aleatória X, para uma população finita, é dada por:

$$\sigma^2 = V(X) = Var(X) = E(X - E(X))^2$$

Já o ***desvio padrão*** é a raiz quadrada da variância.

$$\sigma = \sqrt{Var(X)}$$

Além disso, E(X) pode ser representado por μ.

$$\sigma^2 = E(X - \mu)^2$$

Se for feito o desenvolvimento dessa equação, chegamos a seguinte expressão:

$$\sigma^2 = E(X^2) - \mu^2$$

Exemplo!

Considere a seguinte tabela:

X	P(X)
1	0,25
2	0,10
3	0,35
4	0,30

A primeira a ser feita é construir a coluna de X.P(X) e em seguida calcular a esperança.

X	P(X)	X.P(X)
1	0,25	0,25
2	0,10	0,20
3	0,35	1,05
4	0,30	1,20
Tota l	1,00	2,70

$$\mu = E(X) = \sum_{i=1}^{n} X_i \, . \, P(X_i) = 2{,}70$$

Para calcular a variância utilizaremos a seguinte fórmula.

$$\sigma^2 = E(X^2) - \mu^2$$

Desta forma, teremos que construir uma coluna para o $E(X^2)$.

X	P(X)	X.P(X)	X^2	X^2.P(X)
1	0,25	0,25	$1^2 = 1$	$0{,}25 \, . \, 1 = 0{,}25$
2	0,10	0,20	$2^2 = 4$	$0{,}10 \, . \, 4 = 0{,}40$
3	0,35	1,05	$3^2 = 9$	$0{,}35 \, . \, 9 = 3{,}15$
4	0,30	1,20	$4^2 = 16$	$0{,}30 \, . \, 16 = 4{,}80$
Total	1,00	2,70		8,60

Aplicando a fórmula,

$$\sigma^2 = 8{,}60 - (2{,}70)^2 = 8{,}60 - 7{,}29 = 1{,}31$$

A desvio padrão será

$$\sigma = \sqrt{\sigma^2} = \sqrt{1{,}31}$$

Propriedades

1) $Var(X + k) = Var(X)$

ou

$Var(X - k) = Var(X)$

2) $Var(k\,.X) = k^2\,.Var(X)$

ou

$Var\left(\frac{1}{k}\,.X\right) = \frac{1}{k^2}\,.Var(X)$

As propriedades (1) e (2) podem ser descritas da seguinte forma:

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, variância e o desvio padrão não se alteram;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, a variância fica multiplicada ou dividida pelo quadrado dessa constante.

3) $Var(k) = 0$

A variância de uma constante "k" é zero.

4) Se X e Y são variáveis aleatórias independentes, então $Var(X + Y) = Var(X) + \ Var(Y)$

Essa propriedade só vale se as variáveis forem independentes. Porém, o fato de se saber $Var(X + Y) = Var(X) + \ Var(Y)$ não é garantia que as variáveis são independentes. Isto é, elas podem ser dependentes ou independentes. Além disso, se X e Y forem independentes, a diferença X – Y também será a soma das variâncias.

$$Var(X - Y) = Var(X) + \ Var(Y)$$

Covariância e Correlação

Sabemos que duas variáveis aleatórias podem ser dependentes ou independentes. Sendo que é necessário também saber como se dá essa relação, isto é, se é forte ou fraca. Aqui entra em cena a ***covariância*** e a ***correlação***.

A ***covariância*** é uma medida de variação entre duas variáveis aleatórias. Por exemplo, se os maiores/menores valores de uma variável corresponde principalmente aos maiores/menores valores da outra, essas variáveis tendem a mostrar um comportamento semelhante. E nesse caso, teremos uma covariância positiva. No entanto, se os maiores valores de uma variável corresponde principalmente aos menores valores da outra, essas variáveis tendem a mostrar um comportamento oposto. E nesse caso, teremos uma covariância negativa.

Vejam que, o sinal da variância é muito importante para mostrar a tendência na relação linear entre duas variáveis aleatórias.

- Se o sinal for positivo, temos uma relação positiva, isto é, quando uma variável aumenta, a outra também aumenta;
- Se o sinal for negativo, temos uma relação negativa, isto é, quando uma variável aumenta, a outras diminui.

Sendo que, o aumento ou diminuição não necessariamente ocorre na mesma magnitude, pois a covariância pode assumir qualquer valor e isso torna difícil a interpretação.

A ***covariância*** entre duas variáveis aleatórias é dada por:

$$cov(X,Y) = E(X - \mu_X).(Y - \mu_Y)$$

Onde,

μ_X é a média (esperança de X)

$\mu_X = E(X)$

μ_Y é a média (esperança de Y)

$\mu_Y = E(Y)$

Pessoal, a fórmula da covariância pode ser desenvolvida para chegar à seguinte expressão:

$$cov(X,Y) = E(XY) - \mu_X\mu_Y$$

ou

$$cov(X,Y) = E(XY) - E(X).E(Y)$$

Sendo, $E(XY)$ correspondente ao seguinte:

$$E(XY) = \frac{\sum \quad X \, . \, Y}{N}$$

Onde, N é a soma dos elementos de X e Y.

Desta forma, a covariância procura mostrar se existe um comportamento de interdependência linear entre duas variáveis, mas é uma medida dimensional, sendo, portanto, afetada pelas unidades de medida de X e Y. Para corrigir esse problema, utiliza-se a medida de ***correlação***. Uma vez que, a correlação é uma medida adimensional e só assume valores entre -1 e 1.

A ***correlação*** (ou coeficiente linear de Pearson) entre as variáveis aleatórias é dada por:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X . \sigma_y}$$

Onde,

σ_X é o desvio padrão de X

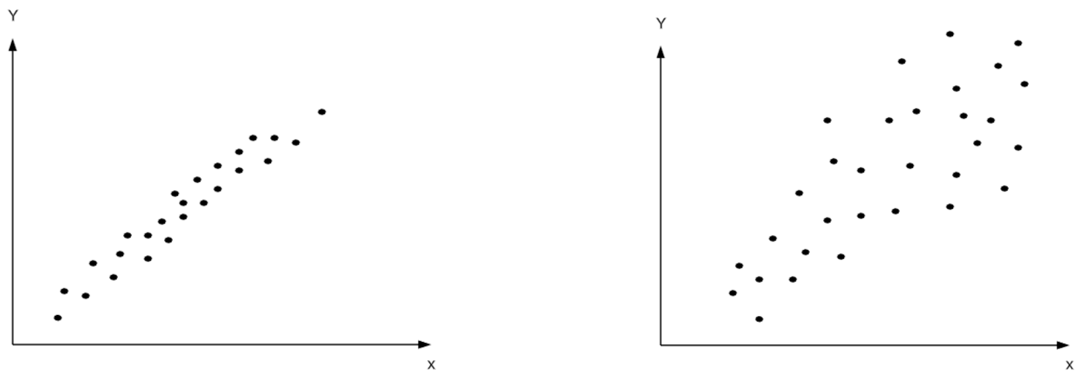
σ_Y é o desvio padrão de Y

A interpretação do coeficiente de correlação é dada da seguinte forma:

- Se $\rho(X,Y) = -1$, significa que as variáveis possuem correlação negativa perfeita. Se X aumentar em uma unidade, Y diminui na mesma magnitude;
- Se $\rho(X,Y) = 0$, significa que as variáveis não são linearmente correlacionadas. Uma variação em X não está associada a uma variação em Y de forma linear;
- Se $\rho(X,Y) = 1$, significa que as variáveis possuem correlação positiva perfeita. Se X aumentar em uma unidade, X aumenta na mesma magnitude;

Para ilustrar, as figuras abaixo mostram a correlação forte de positiva ***(A)*** e a correlação fraca e positiva ***(B)***.

(A) Correlação Forte e Positiva ***(B) Correlação Fraca e Positiva***



$\rho(X, Y) = 1$, relação linear perfeita e positiva;

$\rho(X, Y) = 0$, relação linear inexistente;

$\rho(X, Y) = -1$, relação linear perfeita e negativa;

$\rho(X, Y) > 0$, relação linear positiva;

$\rho(X, Y) < 0$, relação linear negativa.

Se X e Y são variáveis aleatórias independentes, então $E(XY) = E(X) \, . \, E(Y)$. Logo, $cov(X, Y) = \rho(X, Y) = 0$. Mas se $cov(X, Y) = \rho(X, Y) = 0$ não significa que as variáveis são independentes ou dependentes.

Propriedades da Covariância

Considere que X, Y e Z são variáveis aleatórias e que "k" é uma constante. É possível as seguintes relações:

1) $cov(X, Y) = cov(Y, X)$

Aqui a covariância entre X e Y é igual à covariância entre Y e X.

2) $cov(X, X) = Var(X)$

Na realidade, a variância de X é igual a covariância entre X e X.

3) $cov(k, Y) = cov(X, k) = 0$

A covariância entre uma variável e uma constante é sempre zero.

4) $cov(X+Y,Z) = cov(X,Z) + cov(Y,Z)$

Aqui temos o desmembramento de uma soma dentro da covariância.

5) $cov(kX,Y) = cov(X,kY) = k\,.\,cov(X,Y)$

Se uma constante estiver multiplicando uma das variáveis, ela pode sair multiplicando a variância.

Variância da Soma e da Diferença

Quando estudamos a esperança, vimos que a soma/diferença das esperanças de duas variáveis aleatórias era dada por:

$$E(X+Y) = E(X) + E(Y)$$

$$E(X-Y) = E(X) - E(Y)$$

Para a variância, a soma/diferença entre duas variáveis aleatórias é dada por:

$$Var(X+Y) = Var(X) + Var(Y) + 2\,.\,cov(X,Y)$$

$$Var(X-Y) = Var(X) + Var(Y) - 2\,.\,cov(X,Y)$$

Vejam que essas expressões são parecidas com os produtos notáveis.

$$(x+y)^2 = x^2 + y^2 + 2xy$$

$$(x-y)^2 = x^2 + y^2 - 2xy$$

Sendo assim, como a variância pode ser dada por:

$$Var(X) = E(X-\mu_X)^2$$

Podemos dizer que,

$$Var(X+Y) = E[(X+Y) - (\mu_X + \mu_Y)]^2$$

ou

$$Var(X+Y) = E[(X-\mu_X) + (Y-\mu_Y)]^2$$

Aplicando o produto notável, podemos chegar a seguinte expressão:

$$Var(X+Y) = E(X-\mu_X)^2 + E(Y-\mu_Y)^2 + 2\,.\,E[(X-\mu_X).(Y-\mu_Y)]$$

Sendo,

$E(X-\mu_X)^2 = Var(X)$

$E(Y-\mu_Y)^2 = Var(Y)$

$E[(X-\mu_X).(Y-\mu_Y)] = cov(X,Y)$

Portanto,

$$Var(X+Y) = Var(X) + Var(Y) + 2\,.\,cov(X,Y)$$

Além disso, se as variáveis aleatórias forem multiplicadas por constantes. Por exemplo, "a" e "b", respectivamente. Teremos o seguinte:

$$Var(aX+bY) = a^2.Var(X) + b^2.Var(Y) + 2ab\,.\,cov(X,Y)$$

$$Var(aX-bY) = a^2.Var(X) + b^2.Var(Y) - 2ab\,.\,cov(X,Y)$$

Coeficiente de Variação e Variância Relativa

O ***Coeficiente de variação*** de uma variável aleatória X relaciona o desvio-padrão com a média.

$$CV = \frac{\sigma}{\mu}$$

Vejam que é a mesma fórmula que estudamos na estatística descritiva. O coeficiente de variação representa uma normalização do desvio padrão pela média. Essa normalização é feita para transformar determinados valores em escala comum e com isso comparar dispersões de variáveis com médias distintas.

Coeficiente de Variação

É uma medida adimensional, pois o desvio-padrão e média possuem a mesma unidade;
Normalmente é expresso em porcentagem;
Quanto menor o CV, mais homogêneo será o conjunto de dados, ou seja, menor será a dispersão em torno da média.

Outra medida de dispersão relativa é a ***Variância Relativa***. Essa medida é simplesmente o quadrado do CV.

$$VR = \left(\frac{\sigma}{\mu}\right)^2 = \frac{\sigma^2}{\mu^2}$$

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.

Q.01 (IBFC - Analista Administrativo (EBSERH)/Estatística/2020)

Numa distribuição de probabilidade discreta a variável X representa a quantidade de relógios que uma pessoa(adulta) possui:

X = quantidade de relógios	P(X)
0	27%
1	36%
2	18%
3	16%
4	3%

De acordo com a tabela, assinale a alternativa que apresenta o valor esperado para a quantidade de relógios.

a) 1,23.

b) 1,18.

c) 1,46.

d) 1,52.

e) 1,32.

Comentários:

Nessa questão, pede-se o valor esperado para a quantidade de relógio e fornece uma tabela com as quantidades de relógios (X) e a P(X). A esperança é dada por:

$$E(X) = \sum_{i=1}^{n} X_i \, . \, P(X_i)$$

Desta forma, basta construir uma coluna de X.P(X) e depois disso fazer o somatório e teremos a esperança.

X	P(X)	X.P(X)
0	27%=0,27	$0 \, . \, 0{,}27 = 0$
1	36%=0,36	$1 \, . \, 0{,}36 = 0{,}36$
2	18%=0,18	$2 \, . \, 0{,}18 = 0{,}36$
3	16%=0,16	$3 \, . \, 0{,}16 = 0{,}48$

4	3%=0,03	$4 \,.\, 0{,}03 = 0{,}12$
Total	1	1,32

Portanto, o valor esperado é 1,32.

Gabarito: E

Q.02 (CEBRASPE (CESPE) - Profissional de TI (ME)/Ciência de Dados/2020)

		A			
		-1	0	1	total (%)
B	-1	10	5	5	20
	0	0	60	0	60
	1	10	5	5	20
	total(%)	20	70	10	100

Considerando que a tabela precedente mostra o cruzamento de duas variáveis categorizadas A e B, que foram codificadas em três níveis numéricos de resposta: −1, 0 e 1, julgue o item que se segue.

A média da variável A é negativa.

C – CERTO

E - ERRADO

Comentários:

Pessoal, como sabemos a média é dada por:

$$E(X) = \sum_{i=1}^{n} X_i \,.\, P(X_i)$$

A banca quer saber se a média A é negativa. Podemos dispor a tabela de valores de A da seguinte forma.

A	P(1)	P(2)	P(3)	P(total)
-1	10	0	10	20%=0,2

0	5	60	5	70%=0,7
1	5	0	5	10%=0,1
Total				**100%=1**

A esperança será a seguinte:

$$E(X) = (-1)\,.\,0{,}2 + 0\,.\,0{,}7 + 1\,.\,0{,}1 = -0{,}2 + 0 + 0{,}1 = -0{,}1$$

Portanto, correta a questão.

Gabarito: Certo

Q.03 (CEBRASPE (CESPE) - Profissional de TI (ME)/Ciência de Dados/2020)

		A			
		-1	0	1	total (%)
B	-1	10	5	5	20
	0	0	60	0	60
	1	10	5	5	20
	total(%)	20	70	10	100

Considerando que a tabela precedente mostra o cruzamento de duas variáveis categorizadas A e B, que foram codificadas em três níveis numéricos de resposta: −1, 0 e 1, julgue o item que se segue.

As modas e as medianas das variáveis A e B são iguais a zero.

C – CERTO

E - ERRADO

Comentários:

Nessa questão, a banca quer saber se as modas e as medianas de A e B são zero. A primeira coisa a ser feita é colocar a tabela da outra forma para melhor visualização. Além disso, fazer a coluna da probabilidade acumulada.

Para variável A.

A	P(1)	P(2)	P(3)	P(total)	P(A) acum
-1	10	0	10	20%=0,2	0,2
0	5	60	5	70%=0,7	0,9

1	5	0	5	10%=0,1	0,1
Total				100%=1	

Analisando a tabela, podemos observar que na coluna de P(total) que a maior probabilidade é de 70% e corresponde a um valor de A de ***zero***.

Já na mediana, basta analisar a coluna da probabilidade acumulada. Como a mediana representa 50% dos valores, podemos observar que a mediana é ***zero*** (pois os 0,9 ultrapassou os 50%).

Para a variável B.

B	P(1)	P(2)	P(3)	P(total)	P(B) acum
-1	10	5	5	20%=0,2	0,2
0	0	60	0	60%=0,6	0,8
1	10	5	5	20%=0,2	1,0
Total				100%=1	

Analisando a tabela, podemos observar que na coluna de P(total) que a maior probabilidade é de 60% e corresponde a um valor de B de ***zero***.

Já na mediana, basta analisar a coluna da probabilidade acumulada. Como a mediana representa 50% dos valores, podemos observar que a mediana é ***zero*** (pois os 0,8 ultrapassou os 50%).

Desta forma, as modas e medianas são zero.

Gabarito: Certo

Q.04 (IBFC - Engenheiro (EBSERH)/Mecânico/2020)

Analise as afirmativas abaixo sobre Probabilidade e Estatística e dê valores Verdadeiro (V) ou Falso (F).

( ) São classificadas como variáveis discretas as funções para as quais é possível associar um único número real a cada evento de uma partição do espaço amostral.

( ) São classificadas como variáveis contínuas as funções para as quais é possível associar infinitos valores a um intervalo (a, b), sendo que para valores que não pertencem ao intervalo no qual se limita o experimento, a probabilidade de ocorrência é zero.

( ) A variância, ou seja, a medida estatística que concentra as probabilidades em torno da média é indicada por Var(x) ou σ^2 e dada por: $Var(x) = E(x^2) - [E(x)]^2$

( ) O desvio padrão indicado por DP (x) = σ é a raiz quadrada da variância.

Assinale a alternativa que apresenta a sequência correta de cima para baixo.

a) ***F, F, F, F.***

b) ***F, V, V, F.***

c) ***V, F, F, V.***

d) ***V, V, F, F.***

e) ***V, V, V, V.***

Comentários:

Pessoal, essa é uma questão teórica e todas as afirmações estão verdadeiras. Sendo, portanto, muito boa para revisar o assunto.

(V) São classificadas como variáveis discretas as funções para as quais é possível associar um único número real a cada evento de uma partição do espaço amostral.

Pois, as discretas podem assumir apenas certos valores, e é fruto da contagem.

(V) São classificadas como variáveis contínuas as funções para as quais é possível associar infinitos valores a um intervalo (a, b), sendo que para valores que não pertencem ao intervalo no qual se limita o experimento, a probabilidade de ocorrência é zero.

Pois, a contínua resulta de uma medida e pode assumir qualquer valor dentro de um intervalo.

(V) A variância, ou seja, a medida estatística que concentra as probabilidades em torno da média é indicada por Var(x) ou σ^2 ***e dada por:*** $Var(x) = E(x^2) - [E(x)]^2$

O item descreve exatamente a variância.

(V) O desvio padrão indicado por DP (x) = σ ***é a raiz quadrada da variância.***

O item descreve exatamente a desvio padrão.

Gabarito: E

Q.05 (IDECAN - Professor Efetivo de Ensino Básico, Técnico e Tecnológico (IF PB)/Matemática/2019)

Admita que X é uma variável aleatória discreta que assume os valores 5, 10, 15 e 20. Sua função de distribuição acumulada é:

$$\begin{cases} 0\; se\; x\; < 5 \\ 0,4\; se\; 5 \leq x < 10 \\ 0,7\; se\; 10 \leq x < 15 \\ 0,9\; se\; 15 \leq x < 20 \\ 1\; se\; x \geq 20. \end{cases}$$

O desvio padrão é

a) 5.

b) 6,5.

c) 10.

d) 25.

e) 41,7.

Comentários:

Nessa questão, temos uma função de distribuição acumulada e a banca quer saber o valor do desvio padrão. Para calcular o desvio padrão, temos que calcular E(X), E(X^2) e a variância.

Pessoal, a função acumulada dada pela banca foi a seguinte:

$$\begin{cases} 0\; se\; x\; < 5 \\ 0,4\; se\; 5 \leq x < 10 \\ 0,7\; se\; 10 \leq x < 15 \\ 0,9\; se\; 15 \leq x < 20 \\ 1\; se\; x \geq 20. \end{cases}$$

Desta forma, a função probabilidade é dada por:

$p(x) = P(X = x) = \{0\; se\; x < 5\; 0{,}4\; se\; 5 \leq x < 10\; 0{,}3\; se\; 10 \leq x < 15\; 0{,}2\; se\; 15 \leq x < 20\; 0{,}1\; se\; \geq 20$

Cálculo de E(X) e E(X^2):

$$E(X) = \sum X_i \,.\, P(X)$$

$$E(X) = 5 \,.\, P(X = 5) + 10 \,.\, P(X = 10) + 15 \,.\, P(X = 15) + 20 \,.\, P(X = 20)$$

$$E(X) = 5 \,.\, 0{,}4 + 10 \,.\, 0{,}3 + 15 \,.\, 0{,}2 + 20 \,.\, 0{,}1$$

$$E(X) = 2 + 3 + 3 + 2 = 10$$

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

$$E(X^2) = 5^2 \, . P(X = 5) + 10^2 \, . P(X = 10) + 15^2 \, . P(X = 15) + 20^2 \, . P(X = 20)$$

$$E(X^2) = 25 \, . \, 0{,}4 + 100 \, . \, 0{,}3 + 225 \, . \, 0{,}2 + 400 \, . \, 0{,}1$$

$$E(X^2) = 10 + 30 + 45 + 40 = 125$$

Sendo a variância dada por:

$$Var(X) = E(X^2) - E(X)^2$$

$$Var(X) = 125 - 10^2 = 125 - 100 = 25$$

Portanto, o desvio padrão será:

$$S = \sqrt{Var(X)}$$

$$S = \sqrt{25} = 5$$

Gabarito: A

Q.06 (VUNESP - Analista Administrativo (EBSERH HC-UFU)/Estatística/2020)

Sendo var(x) a variância de uma variável aleatória x e cov(x,y) a covariância entre duas variáveis aleatórias x e y, tem-se que

a) var(ax – by) = a var(x) – b var(y).

b) var(ax – by) = a² var(x) – b² var(y).

c) var(ax – by) = a² var(x) + b² var(y).

d) var(ax – by) = a² var(x) + b² var(y) – 2ab cov(x,y).

e) var(ax – by) = a² var(x) + b² var(y) + 2ab cov(x,y).

Comentários:

Vejam que banca que saber a fórmula variância da subtração duas variáveis aleatórias multiplicadas por constantes.

$$Var(aX - bY) = a^2 . Var(X) + b^2 . Var(Y) - 2ab \, . cov(X,Y)$$

Portanto, letra D resposta da questão.

Gabarito: D

Q.07 (VUNESP - Analista Técnico Científico (MPE SP)/Economista/2019)

Uma variável x tem distribuição normal com média 5 e desvio padrão igual a 3. Já a variável y também tem distribuição normal, mas com média 10 e desvio padrão igual a 4. Sabe-se que x e y são independentes. A variável z=x+y tem

a) normal com média 15 e desvio padrão igual a 5.

b) normal com média 15 e desvio padrão igual a 7.

c) desconhecida com média 15 e desvio padrão igual a 7.

d) t, de Student com média 7,5 e desvio padrão igual a 5.

e) qui-quadrado com dois graus de liberdade.

Comentários:

Pessoal, as variáveis X e Y são normais e independentes. Como a variável Z é a soma de X e Y, podemos calcular a média dessa nova variável. Logo, descartamos a alternativa C.

Como X e Y é normal, Z também será normal. Logo, já descartamos a alternativa D. A letra E não temos como analisar (extrapolação). Portanto, ficamos com as alternativas A e B que pedem a média e desvio padrão de Z, isto é, X+Y.

A esperança da soma de duas variáveis é dada por:

$$E(X+Y) = E(X) + E(Y)$$

Onde,

E(X) = 5

E(Y) = 10

$$E(X+Y) = 5 + 10 = 15$$

Já para calcular o desvio padrão da soma de duas variáveis, temos primeiro que encontrar a variância e depois extrair a raiz quadrada.

$$Var(X+Y) = Var(X) + Var(Y)$$

Sendo,

S(X) = 3

Var(X) =9

S(Y) = 4

Var(Y) = 16

$$Var(X+Y) = 9 + 16 = 25$$

Portanto,

$$S(X) = \sqrt{Var(X+Y)}$$

$$S(X) = \sqrt{25} = 5$$

Portanto, teremos uma variável normal com média igual a 15 e desvio padrão igual a 5.

Gabarito: A

Q.08 (IBFC - Analista Administrativo (EBSERH)/Estatística/2020)

Com relação ao coeficiente de correlação linear (r), é incorreto afirmar que:

a) Se r for um número próximo de 1, então x e y têm forte correlação linear.

b) Se r = 0,2, então x e y têm forte correlação linear.

c) Se r for um número próximo de -1, então x e y têm forte correlação linear.

d) Se r for um número próximo de 0, então x e y têm fraca correlação linear.

e) Se r = - 0,8, então x e y têm forte correlação linear.

Comentários:

Nessa questão, temos que saber interpretar os valores do coeficiente de correlação. Na nossa aula apresentamos um quadro dessa interpretação. Como sabemos, o coeficiente de correlação varia de -1 a 1. E quanto mais próximo de -1 ou 1 mais forte será a correlação.

$\rho(X,Y) = 1$, relação linear perfeita e positiva;

$\rho(X,Y) = 0$, relação linear inexistente;

$\rho(X,Y) = -1$, relação linear perfeita e negativa;

$\rho(X,Y) > 0$, relação linear positiva;

$\rho(X,Y) < 0$, relação linear negativa.

Comparando as alternativas com esse quadro, podemos observar que a alternativa B está errada, pois uma correlação de 0,2 linear positiva, mas não fortemente. As demais alternativas estão corretas.

Gabarito: B

Q.09 (VUNESP - Analista Administrativo (EBSERH HC-UFU)/Estatística/2020)

Dados para responder à questão.

A variável x tem média 4 e desvio padrão 2, enquanto a variável y tem média 3 e desvio padrão 1. A covariância entre x e y é –1.

O coeficiente de correlação entre x e y é

a) 0,5.

b) -0,5.

c) 1.

d) -1.

e) -0,25.

Comentários:

Nessa questão, queremos o coeficiente de correlação. E como sabemos ele é dado por:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X.\sigma_y}$$

A questão forneceu as seguintes informações:

$cov(X,Y) = -1$

$\sigma_X = 2$

$\sigma_Y = 1$

Logo,

$$\rho(X,Y) = \frac{-1}{2\ .\ 1} = -\frac{1}{2} = -0{,}5$$

Gabarito: B

Q.10 (VUNESP - Curso de Formação de Oficiais do Quadro Complementar (EsFCEx)/Estatística/2020)

Dados mensais sobre a despesa com publicidade (em milhares de dólares) e a receita (em milhares de dólares) do Four Seasons Restaurant são apresentados a seguir.

Despesas (x)	1	2	3,5	4,5	5	7
Receitas (y)	5	9				29

Assinale a alternativa que contém os três valores da Receita que completam, respectivamente, a tabela apresentada para que a correlação entre essas duas variáveis seja 1.

a) 15, 19, 21.

b) 13, 19, 21.

c) 15, 19, 25.

d) 13, 17, 21.

e) 17, 21, 25.

Comentários:

Pessoal, nessa questão a banca traz uma tabela que mostra a correlação entre as variáveis X e Y. E pede para completar os espaços da variável Y.

A reta de regressão linear entre X e Y é dada por:

$Y = a\,.X + b$

Onde,

"a" é o coeficiente linear

"b" é o coeficiente angular.

Vejam que temos os dois primeiros valores de X e Y. Com isso, podemos encontrar os valores de "a" e "b".

$$5 = a\,.\,1 + b \quad (1)$$

$$9 = a\,.\,2 + b \quad (2)$$

Podemos isolar, por exemplo, b na equação (1) e obter a equação (3).

$$b = 5 - a \quad (3)$$

Agora podemos substituir a equação (3) na equação (2) e obter o valor de "a".

$$9 = 2a + 5 - a$$

$$9 - 5 = 2a - a$$

$$a = 4$$

Para encontrar "b", basta substituir o valor de "a" na equação (3).

$$b = 5 - 4$$

$$b = 1$$

De posse dos valores dos coeficientes, basta substituir na equação da reta.

$$Y = a\,.\,X + b$$

$$Y = 4.X + 1$$

Agora é só calcular os valores de Y que correspondem aos valores X (3,5; 4,5; 5).

Para X = 3,5

$$Y = 4\,.\,3{,}5 + 1 = 15$$

Para X = 4,5

$$Y = 4\,.\,4{,}5 + 1 = 19$$

Para X = 5

$$Y = 4\,.\,5 + 1 = 21$$

Gabarito: A

Allan Maux

LISTA DE QUESTÕES ESTRATÉGICAS

Q.01 (IBFC - Analista Administrativo (EBSERH)/Estatística/2020)

Numa distribuição de probabilidade discreta a variável X representa a quantidade de relógios que uma pessoa(adulta) possui:

X = quantidade de relógios	P(X)
0	27%
1	36%
2	18%
3	16%
4	3%

De acordo com a tabela, assinale a alternativa que apresenta o valor esperado para a quantidade de relógios.

a) 1,23.

b) 1,18.

c) 1,46.

d) 1,52.

e) 1,32.

Q.02 (CEBRASPE (CESPE) - Profissional de TI (ME)/Ciência de Dados/2020)

		A			
		-1	0	1	total (%)
B	-1	10	5	5	20
	0	0	60	0	60
	1	10	5	5	20
	total(%)	20	70	10	100

Considerando que a tabela precedente mostra o cruzamento de duas variáveis categorizadas A e B, que foram codificadas em três níveis numéricos de resposta: −1, 0 e 1, julgue o item que se segue.

A média da variável A é negativa.

C – CERTO

E - ERRADO

Q.03 (CEBRASPE (CESPE) - Profissional de TI (ME)/Ciência de Dados/2020)

		A			
		-1	0	1	total (%)
B	-1	10	5	5	20
	0	0	60	0	60
	1	10	5	5	20
	total(%)	20	70	10	100

Considerando que a tabela precedente mostra o cruzamento de duas variáveis categorizadas A e B, que foram codificadas em três níveis numéricos de resposta: −1, 0 e 1, julgue o item que se segue.

As modas e as medianas das variáveis A e B são iguais a zero.

C – CERTO

E - ERRADO

Q.04 (IBFC - Engenheiro (EBSERH)/Mecânico/2020)

Analise as afirmativas abaixo sobre Probabilidade e Estatística e dê valores Verdadeiro (V) ou Falso (F).

( ) São classificadas como variáveis discretas as funções para as quais é possível associar um único número real a cada evento de uma partição do espaço amostral.

( ) São classificadas como variáveis contínuas as funções para as quais é possível associar infinitos valores a um intervalo (a, b), sendo que para valores que não pertencem ao intervalo no qual se limita o experimento, a probabilidade de ocorrência é zero.

( ) A variância, ou seja, a medida estatística que concentra as probabilidades em torno da média é indicada por Var(x) ou σ^2e dada por: Var(x) = E (x^2) - [E (x)] ^{2}

( ) O desvio padrão indicado por DP (x) = σ é a raiz quadrada da variância.

Assinale a alternativa que apresenta a sequência correta de cima para baixo.

a) F, F, F, F.

b) F, V, V, F.

c) V, F, F, V.

d) V, V, F, F.

e) V, V, V, V.

Q.05 (IDECAN - Professor Efetivo de Ensino Básico, Técnico e Tecnológico (IF PB)/Matemática/2019)

Admita que X é uma variável aleatória discreta que assume os valores 5, 10, 15 e 20. Sua função de distribuição acumulada é:

$$\begin{cases} 0\ se\ x < 5 \\ 0,4\ se\ 5 \leq x < 10 \\ 0,7\ se\ 10 \leq x < 15 \\ 0,9\ se\ 15 \leq x < 20 \\ 1\ se\ x \geq 20. \end{cases}$$

O desvio padrão é

a) 5.

b) 6,5.

c) 10.

d) 25.

e) 41,7.

Q.06 (VUNESP - Analista Administrativo (EBSERH HC-UFU)/Estatística/2020)

Sendo var(x) a variância de uma variável aleatória x e cov(x,y) a covariância entre duas variáveis aleatórias x e y, tem-se que

a) var(ax – by) = a var(x) – b var(y).

b) var(ax – by) = a^2 var(x) – b^2 var(y).

c) var(ax – by) = a^2 var(x) + b^2 var(y).

d) var(ax – by) = a^2 var(x) + b^2 var(y) – 2ab cov(x,y).

e) var(ax – by) = a^2 var(x) + b^2 var(y) + 2ab cov(x,y).

Q.07 (VUNESP - Analista Técnico Científico (MPE SP)/Economista/2019)

Uma variável x tem distribuição normal com média 5 e desvio padrão igual a 3. Já a variável y também tem distribuição normal, mas com média 10 e desvio padrão igual a 4. Sabe-se que x e y são independentes. A variável z=x+y tem

a) normal com média 15 e desvio padrão igual a 5.

b) normal com média 15 e desvio padrão igual a 7.

c) desconhecida com média 15 e desvio padrão igual a 7.

d) t, de Student com média 7,5 e desvio padrão igual a 5.

e) qui-quadrado com dois graus de liberdade.

Q.08 (IBFC - Analista Administrativo (EBSERH)/Estatística/2020)

Com relação ao coeficiente de correlação linear (r), é incorreto afirmar que:

a) Se r for um número próximo de 1, então x e y têm forte correlação linear.

b) Se r = 0,2, então x e y têm forte correlação linear.

c) Se r for um número próximo de -1, então x e y têm forte correlação linear.

d) Se r for um número próximo de 0, então x e y têm fraca correlação linear.

e) Se r = - 0,8, então x e y têm forte correlação linear.

Q.09 (VUNESP - Analista Administrativo (EBSERH HC-UFU)/Estatística/2020)

Dados para responder à questão.

A variável x tem média 4 e desvio padrão 2, enquanto a variável y tem média 3 e desvio padrão 1. A covariância entre x e y é –1.

O coeficiente de correlação entre x e y é

a) 0,5.

b) -0,5.

c) 1.

d) -1.

e) -0,25.

Q.10 (VUNESP - Curso de Formação de Oficiais do Quadro Complementar (EsFCEx)/Estatística/2020)

Dados mensais sobre a despesa com publicidade (em milhares de dólares) e a receita (em milhares de dólares) do Four Seasons Restaurant são apresentados a seguir.

Despesas (x)	1	2	3,5	4,5	5	7
Receitas (y)	5	9				29

Assinale a alternativa que contém os três valores da Receita que completam, respectivamente, a tabela apresentada para que a correlação entre essas duas variáveis seja 1.

a) 15, 19, 21.

b) 13, 19, 21.

c) 15, 19, 25.

d) 13, 17, 21.

e) 17, 21, 25.

Gabarito

1	2	3	4	5	6	7	8	9	10
E	C	C	E	A	D	A	B	B	A

VARIABILIDADE / MEDIDAS DE DISPERSÃO

Sumário

O que é mais cobrado dentro do assunto

Vamos analisar agora como se comporta a incidência dos sub assuntos da nossa aula de hoje. Assim, você será melhor direcionado nos seus estudos, vejam:

MEDIDAS DE DISPERSÃO	Grau de incidência
DESVIO PADRÃO E VARIÂNCIA	*73,5%*
COEFICIENTES DE VARIAÇÃO / VARIAÇÃO RELATIVA	*20,9%*
INTERVALO INTERQUARTÍLICO	*2,8%*
AMPLITUDE	*2,8%*
TOTAL	*100,00%*

ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

A ideia desta seção é apresentar um roteiro para que você realize uma revisão completa do assunto e, ao mesmo tempo, destacar aspectos do conteúdo que merecem atenção.

Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

MEDIDAS DE VARIABILIDADE

Pessoal, as ***Medidas de Variabilidade*** ou ***Medidas de Dispersão*** têm a finalidade de medir a dispersão dos valores de um conjunto em torno de um valor médio. Essas medidas são divididas em dois grupos: ***Medidas de dispersão Absoluta*** e ***Medida de dispersão Relativa***.

MEDIDAS DE DISPERSÃO ABSOLUTA

Amplitude Total

A ***Amplitude total*** é a diferença entre os valores extremos de um conjunto de dados. Essa medida é muito pobre, pois não mostra o grau de dispersão existente no conjunto de dados, pois apenas leva em conta o valor máximo e mínimo.

$$AT = X_{máx} - X_{mín}$$

Portanto, como são desconsiderados os valores entre os extremos, essa medida pode nos levar a uma conclusão errada dos dados em análise. Além disso, é sensível ao tamanho da amostra, uma vez que, pode apresentar muita variação de uma amostra para outra, isso considerando uma mesma população.

A tabela abaixo mostra exemplos do cálculo da Amplitude Total.

<table>
<tr><th>Amplitude Total em dados não agrupados</th><th>Amplitude Total em dados agrupados sem intervalo de classe</th><th>Amplitude Total em dados agrupados em classe</th></tr>
<tr><td>{1, 2, 2, 3, 4, 5, 5, 6, 9, 9, 12}</td><td><table><tr><th>Idades</th><th>Frequência Absoluta (fi)</th></tr><tr><td>10</td><td>12</td></tr><tr><td>20</td><td>5</td></tr><tr><td>30</td><td>13</td></tr><tr><td>40</td><td>5</td></tr><tr><td>50</td><td>6</td></tr><tr><td>60</td><td>9</td></tr></table></td><td><table><tr><th>Idades</th><th>Frequência Absoluta (fi)</th></tr><tr><td>0 ⊢ 10</td><td>12</td></tr><tr><td>10 ⊢ 20</td><td>5</td></tr><tr><td>20 ⊢ 30</td><td>13</td></tr><tr><td>30 ⊢ 40</td><td>5</td></tr><tr><td>40 ⊢ 50</td><td>6</td></tr><tr><td>50 ⊢ 60</td><td>9</td></tr></table></td></tr>
<tr><td>Diferença entre o valor máximo e o mínimo.</td><td>Diferença entre o valor máximo e o mínimo.</td><td>Duas formas:

1) Diferença entre o limite superior da última classe e o limite inferior da primeira classe.

2) Diferença entre o ponto médio da última classe e o</td></tr>
</table>

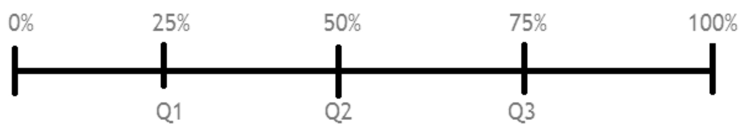
		ponto médio da primeira classe.
$AT = X_{máx} - X_{mín}$ $AT = 12 - 1 = 11$	$AT = X_{máx} - X_{mín}$ $AT = 60 - 10 = 50$	1) $AT = l_{sup} - l_{inf}$ $AT = 60 - 0 = 60$ 2) $AT = PM_{últ} - PM_{prin}$ $AT = 55 - 5 = 50$

Propriedades da Amplitude Total

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, a Amplitude Total não se altera;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, a Amplitude Total fica multiplicada ou dividida por essa constante.

Amplitude Interquartílica

A ***Amplitude Interquartílica***, assim como a Amplitude Total, é uma medida de dispersão pobre. Como sabemos, os ***quartis*** dividem os dados em *quatro partes* de mesma frequência. Desta forma, teremos sempre *três quartis* e cada um deles corresponderá a 25% do conjunto de dados. Observem a figura abaixo.



A Amplitude interquartílica é dada pela diferença entre o terceiro quartil (Q_3) e o primeiro quartil (Q_1).

Amplitude interquartílica = $Q_3 - Q_1$

Além disso, a metade dessa amplitude é chamada de Amplitude semi-interquartílica ou desvio quartílico.

Amplitude semi-interquartílica = $\frac{Q_3 - Q_1}{2}$

Propriedades da Amplitude Interquartílica

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, a Amplitude Interquartílica não se altera;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, a Amplitude Interquartílica fica multiplicada ou dividida por essa constante.

Fiquem atentos, pois as fórmulas da Amplitude Interquartílica e semi-interquartílica são semelhantes. As propriedades da Amplitude Interquartílica e Amplitude Total são iguais.

Desvios em Relação à Média Aritmética e Mediana

Pessoal, um ***desvio*** é a distância entre qualquer observação de um conjunto de dados e uma medida descritiva desse conjunto. Normalmente, calcula-se os desvios em relação à média e à mediana.

$$d_i = X - \underline{X}$$

ou

$$d_i = X - M_d$$

Onde,

$\underline{X}$ é a média.

M_d é a mediana.

Quando os desvios em relação a uma medida descritiva são pequenos, e como as observações estão concentradas em torno dessa medida, a variabilidade dos dados é pequena.

Quando os desvios em relação a uma medida descritiva são maiores, as observações estão mais dispersas, a variabilidade dos dados é grande.

Propriedades dos Desvios em Relação à Média Aritmética e Mediana

- A soma dos desvios em relação à média é nula;
- A soma dos quadrados dos desvios da sequência de números, em relação a número "k", é mínima se "k" for a média aritmética dos números;
- A soma dos desvios absolutos de uma sequência de números, em relação a um número "k", é mínima quando "k" é a mediana dos números.

Desvio Absoluto Médio

O ***Desvio Absoluto Médio*** ou ***Desvio Médio***, mede a dispersão entre os valores da distribuição e a média dos dados calculados.

Pessoal, se tivermos uma distribuição muito grande e quiséssemos calcular os desvios, teríamos que calcular desvios para cada número dessa distribuição e depois tirar conclusões desses valores encontrados. Logo, isso não seria viável.

Desta forma, uma solução seria calcular um único número que representasse toda a distribuição. Para isso, poderíamos calcular a média de todos os desvios, isto é, somar todos os desvios e dividir pela quantidade de observações (n). Sendo que, temos desvios positivos, negativos e nulos. Os positivos anulam os negativos e o resultado da soma será sempre zero.

$$\underline{d} = \frac{\sum_{i=1}^{n} \quad (X_i - \underline{X})}{n} = \frac{0}{n} = 0$$

Portanto, uma solução para esse problema seria colocar cada desvio em módulo, surgindo assim o ***Desvio Absoluto Médio*** ou ***Desvio Médio***.

$$\underline{d} = \frac{\sum_{i=1}^{n} \left| X_i - \underline{X} \right|}{n}$$

O ***Desvio Médio*** é uma medida de dispersão mais robusta do que a Amplitude Total e a Amplitude interquartílica, pois são considerados todos os valores da distribuição.

Para dados agrupados, a fórmula é a seguinte:

$$\underline{d} = \frac{\sum_{i=1}^{n} \left| X_i - \underline{X} \right| . f_i}{\sum f_i}$$

Onde:

f_i é a frequência absoluta simples.

Já para dados agrupados em classe, a fórmula é a seguinte:

$$\underline{d} = \frac{\sum_{i=1}^{n} \left| PM - \underline{X} \right| . f_i}{\sum f_i}$$

Onde:

PM é o ponto médio.

Iremos fazer um exemplo para os dados agrupados em classe, para exemplificar o cálculo do ***Desvio Médio***.

Exemplo: Calcule o desvio médio das idades de um conjunto de pessoas:

A primeira coisa a ser feita é calcular os pontos médios e em seguida a média.

idades	frequência Absoluta (fi)	PM	PM.fi
0 ⊢ 10	12	5	60
10 ⊢ 20	5	15	75
20 ⊢ 30	13	25	325
30 ⊢ 40	5	35	175
40 ⊢ 50	6	45	270
50 ⊢ 60	9	55	495
Total	**50**		**1.400**

Logo, a média será

$$\underline{X} = \frac{\sum \quad PM \,.\, f_i}{\sum \quad f_i} = \frac{1.400}{50} = 28$$

De posse da média, podemos calcular o módulo do desvio em relação à média.

idades	frequência Absoluta (fi)	PM	PM.fi	$\lvert PM - \underline{x}\rvert$
0 ⊢ 10	12	5	60	23
10 ⊢ 20	5	15	75	13
20 ⊢ 30	13	25	325	3
30 ⊢ 40	5	35	175	7
40 ⊢ 50	6	45	270	17
50 ⊢ 60	9	55	495	27
Total	**50**		**1.400**	**90**

Agora, basta multiplicar os módulos dos desvios pela frequência absoluta simples.

idades	frequência Absoluta (fi)	PM	PM.fi	$\lvert PM - \underline{x}\rvert$	$\lvert PM - \underline{x}\rvert . f_i$
0 ⊢ 10	12	5	60	23	276
10 ⊢ 20	5	15	75	13	65
20 ⊢ 30	13	25	325	3	39
30 ⊢ 40	5	35	175	7	35
40 ⊢ 50	6	45	270	17	102
50 ⊢ 60	9	55	495	27	243
Total	**50**		**1.400**		**760**

Aplicando a fórmula, teremos o seguinte:

$$\underline{d} = \frac{\sum_{i=1}^{n} \quad \lvert PM - \underline{X}\rvert \,.\, f_i}{\sum \quad f_i} = \frac{760}{50} = 15{,}2$$

Propriedades do Desvio Médio

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, o desvio médio não se altera;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, o desvio médio fica multiplicado ou dividido por essa constante.

As propriedades da Amplitude Interquartílica, Amplitude Total e Desvio Médio são iguais.

Variância e Desvio-Padrão

Quando introduzimos o assunto ***desvios médios*** observamos que não era viável calcular uma medida de dispersão que simplesmente calcula a média dos desvios, pois essa medida sempre seria zero.

Para resolver esse problema, calculamos a média do módulo dos desvios. Sendo que, outra forma de ser resolvido esse problema é calcular, ao invés de calcular a média do módulo dos desvios, à média do quadrado dos desvios. Com isso, obteremos a ***variância***, isto é, a média aritmética do quadrado dos desvios.

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

Onde:

μ é a média populacional

X_i é a variável em análise

n é o número de elementos da população

Pessoal, como os desvios são elevados ao quadrado, a unidade da variável em análise também é elevada ao quadrado. Por exemplo, se estivermos calculando a variância das idades (anos) de um conjunto de pessoas, essa medida ficará anos ao quadrado (anos^2). Logo, essa medida não terá sentido físico. Para acabar com esse inconveniente, utilizamos o conceito de ***Desvio-Padrão***, que nada mais é que a raiz quadrada da ***variância***.

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

É importante salientar que temos variância e desvio-padrão populacional e amostral. O que tratamos até o momento é o populacional. Logo, para o amostral temos que fazer uma modificação na fórmula, ficando da seguinte forma:

Variância amostral: $S^2 = \frac{\sum_{i=1}^{n} (X-\underline{X})^2}{n-1}$

Desvio-padrão amostral: $S = \sqrt{\frac{\sum_{i=1}^{n} (X-\underline{X})^2}{n-1}}$

Outra forma de calcular a variância populacional é através da seguinte fórmula:

$$\sigma^2 = \underline{X^2} - (\underline{X})^2$$

Onde:

$\underline{X^2}$ é a média dos quadrados.

$(\underline{X})^2$ é o quadrado da média.

Para uma amostra essa fórmula fica da seguinte forma:

- $$S^2 = \frac{n}{n-1} \cdot \left[\underline{X^2} - (\underline{X})^2\right]$$

Variância e Desvio-Padrão
A variância/desvio-padrão de um conjunto será zero quando todos os elementos forem iguais; Sempre será maior ou igual a zero; No cálculo da variância/desvio-padrão haverá uma pequena diferença entre o cálculo populacional e amostral.

Para dados agrupados, as fórmulas são as seguintes:

1) Populacional

$$\sigma^2 = \frac{\sum_{i=1}^{n} \quad (X_i - \mu)^2 \,.\, f_i}{n}$$

ou

$$\sigma^2 = \frac{\sum_{i=1}^{n} \quad X_i^2 - \frac{(\sum_{i=1}^{n} \quad X_i . f_i)^2}{n}}{n}$$

2) Amostral

$$S^2 = \frac{\sum_{i=1}^{n} \quad \left(X_i - \underline{X}\right)^2 \,.\, f_i}{n-1}$$

ou

$$S^2 = \frac{\sum_{i=1}^{n} \quad (X_i^2 . f_i) - \frac{(\sum_{i=1}^{n} \quad X_i . f_i)^2}{n}}{n-1}$$

Onde:

f_i é a frequência absoluta simples.

Pessoal, para calcular o desvio-padrão, basta encontrar a raiz quadrada da variância.

Já para dados agrupados em classe, a fórmula é a seguinte:

1) Populacional

$$\sigma^2 = \frac{\sum_{i=1}^{n} \quad (PM_i - \mu)^2 \,.\, f_i}{n}$$

ou

$$\sigma^2 = \frac{\sum_{i=1}^{n} \quad PM_i^2 - \frac{(\sum_{i=1}^{n} \quad PM_i . f_i)^2}{n}}{n}$$

2) Amostral

$$S^2 = \frac{\sum_{i=1}^{n} \quad \left(PM_i - \underline{X}\right)^2 \,.\, f_i}{n-1}$$

ou

$$S^2 = \frac{\sum_{i=1}^{n} (PM_i^2 . f_i) - \frac{(\sum_{i=1}^{n} PM_i . f_i)^2}{n}}{n-1}$$

Onde:

PM é o ponto médio.

Pessoal, para calcular o desvio-padrão, basta encontrar a raiz quadrada da variância.

Iremos fazer um exemplo para os dados agrupados em classe, para exemplificar o cálculo da ***Variância*** e ***Desvio-Padrão***.

Exemplo: Calcule a variância amostral e desvio padrão amostral das idades de um conjunto de pessoas:

A primeira coisa a ser feita é calcular os pontos médios e em seguida a média.

idades	frequência Absoluta (fi)	PM	PM.fi
0 ⊢ 10	12	5	60
10 ⊢ 20	5	15	75
20 ⊢ 30	13	25	325
30 ⊢ 40	5	35	175
40 ⊢ 50	6	45	270
50 ⊢ 60	9	55	495
Total	**50**		**1.400**

Logo, a média será

$$\underline{X} = \frac{\sum PM . f_i}{\sum f_i} = \frac{1.400}{50} = 28$$

De posse da média, podemos calcular os quadrados dos desvios em relação à média.

idades	frequência Absoluta (fi)	PM	PM.fi	$(PM - \underline{x})$	$(PM - \underline{x})^2$
0 ⊢ 10	12	5	60	-23	529
10 ⊢ 20	5	15	75	-13	169
20 ⊢ 30	13	25	325	-3	9
30 ⊢ 40	5	35	175	7	49
40 ⊢ 50	6	45	270	17	289
50 ⊢ 60	9	55	495	27	729

Total	***50***		1.400		1.774

Agora, basta multiplicar os quadrados dos desvios em relação à média pela frequência absoluta simples.

idades	*frequência Absoluta (fi)*	*PM*	*PM.fi*	$(PM - \underline{x})$	$(PM - \underline{x})^2$	$(PM - \underline{x})^2 . f_i$
0 ⊢ 10	*12*	5	60	-23	529	6.348
10 ⊢ 20	*5*	15	75	-13	169	845
20 ⊢ 30	*13*	25	325	-3	9	117
30 ⊢ 40	*5*	35	175	7	49	245
40 ⊢ 50	*6*	45	270	17	289	1.734
50 ⊢ 60	*9*	55	495	27	729	6.561
Total	***50***		**1.400**			**15.850**

Aplicando a fórmula, teremos o seguinte:

$$S^2 = \frac{\sum_{i=1}^{n} \quad (PM_i - \underline{X})^2 . f_i}{n-1} = \frac{15.850}{50-1} = \frac{15.850}{49} \approx 324 \; anos^2$$

O desvio-padrão será a raiz desse valor. Logo,

$$S = \sqrt{S^2} = \sqrt{324} = 18 \; anos$$

Propriedades do Variância e Desvio-Padrão

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, variância e o desvio padrão não se alteram;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, a variância fica multiplicada ou dividida pelo quadrado dessa constante.
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, o desvio-padrão fica multiplicado ou dividido por essa constante.

MEDIDAS DE DISPERSÃO ABSOLUTA

Coeficiente de Variação e Variância Relativa

O ***Coeficiente de Variação*** é uma medida de dispersão que relaciona o desvio-padrão com a média.

$$CV = \frac{\sigma}{\mu}$$

Isso para uma população. Se for uma amostra a fórmula é a seguinte:

$$CV = \frac{S}{\underline{X}}$$

Coeficiente de Variação

É uma medida adimensional, pois o desvio-padrão e média possuem a mesma unidade;
Normalmente é expresso em porcentagem;
Quanto menor o CV, mais homogêneo será o conjunto de dados, ou seja, menor será a dispersão em torno da média.

Outra medida de dispersão relativa é a ***Variância Relativa***. Essa medida é simplesmente o quadrado do CV.

$$VR = \left(\frac{\sigma}{\mu}\right)^2 = \frac{\sigma^2}{\mu^2}$$

Isso para uma população. Se for uma amostra a fórmula é a seguinte:

$$VR = \left(\frac{S}{\underline{X}}\right)^2 = \frac{S^2}{\underline{X}^2}$$

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.

Questões CEBRASPE

Q.01 (CEBRASPE (CESPE) /ACE (TCE-RJ) / Controle Externo/TI/2021)

X	frequência absoluta
0	5
1	10
2	20
3	15
total	50

Considerando que a tabela precedente mostra a distribuição de frequências de uma variável quantitativa X, julgue o item a seguir.

A variância amostral de X é superior a 0,89.

C - CERTO

E – ERRADO

Comentários:

Nessa questão a banca que saber se a variância amostral é superior a 0,89. Para isso, temos que saber a fórmula da variância. Aqui utilizaremos a seguinte:

$$S^2 = \frac{1}{n-1}\left[\left(\sum f.X_i^2\right) - \frac{\left(\sum f.X_i\right)^2}{n}\right]$$

Para resolver essa questão teremos que calcular duas colunas, a "f.X" e a "f.X²" e em seguida fazer o somatório delas.

X	Frequência Absoluta	f.X	f.X²

0	**5**	*0*	*0*
1	**10**	*10*	*10*
2	**20**	*40*	*80*
3	**15**	*45*	*135*
total	**50**	**95**	**225**

Aplicando os valores na fórmula teremos o seguinte:

$$S^2 = \frac{1}{50-1}\left(225 - \frac{95^2}{50}\right)$$

$$S^2 = \frac{1}{49}\left(225 - \frac{9025}{50}\right)$$

$$S^2 = \frac{1}{49}(225 - 180{,}5)$$

$$S^2 = \frac{44{,}5}{49}$$

$$S^2 = 0{,}908$$

Logo, a variância foi superior a 0,89 como diz a questão. Correta a questão.

Gabarito: Certo

Q.02 (CEBRASPE (CESPE) / Profissional de Tecnologia da Informação (ME) /2020)

Considerando que R representa uma variável quantitativa cuja média, mediana e variância são, respectivamente, iguais a 70, 80 e 100, e que $U = \frac{R}{10} - 7$***, julgue o próximo item, acerca das variáveis U e R.***

O desvio padrão da variável U é igual a 1.

C - CERTO

E - ERRADO

Comentários:

Aqui a banca traz os valores da média, mediana e variância de uma variável R e pede o valor do desvio padrão da U. Sendo U obtido da seguinte forma:

$$U = \frac{R}{10} - 7$$

Como desejamos o valor do desvio padrão e nos foi dada a variância de R. Basta calcular o desvio padrão de R e em seguida aplicar a expressão dada.

Variância de R = 100.

$S = \sqrt{S^2}$

Desvio padrão de R = 10.

O Desvio padrão é afetado pela multiplicação e pela divisão. Logo, a expressão fica da seguinte forma:

$$U = \frac{R}{10}$$

Aplicando o valor de R, ficamos com o seguinte:

$$U = \frac{10}{10}$$

$$U = 1$$

Desta forma, o desvio padrão é igual a 1 como afirma a banca.

Gabarito: Certo

Q.03 (CEBRASPE (CESPE) / Professor (Pref São Cristóvão)/Matemática/Educação Básica/2019)

A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

idade (em anos)	*9*	*10*	*11*	*12*	*13*	*14*
quantidade de estudantes	*6*	*22*	*0*	*1*	*0*	*1*

A partir dessa tabela, julgue o item.

O desvio padrão das idades é inferior a 1ano.

C - Certo.

E – Errado.

Comentários:

Pessoal, a banca forneceu uma tabela com os valores das idades e a quantidade (frequência) de um grupo de alunos da turma A e pergunta se o desvio padrão das idade é inferior a 1 ano.

A tabela apresentada da questão é a seguinte.

X	f	f · X
9	6	54
10	22	220
11	0	0
12	1	12
13	0	0
14	1	14
Total	**30**	**300**

Onde "f" é a quantidade de alunos e "X" é a idade. A primeira coisa a ser feita é calcular a média da turma A, para isso temos que construir uma coluna f · X (ver tabela acima) e em seguida fazer o somatório. Temos que a média pode ser obtida da seguinte forma:

$$\underline{X} = \frac{\sum_{i=1}^{n} \quad X_i \cdot f}{n}$$

$$\underline{X} = \frac{300}{30} = 10$$

De posse da média, podemos calcular os quadrados dos desvios em relação a média e depois multiplicar pela frequência.

X	f	f · X	$(X_i - \underline{X})$	$(X_i - \underline{X})^2$	$(X_i - \underline{X})^2 . f_i$
9	6	54	9-10=-1	1	6
10	22	220	10-10=0	0	0
11	0	0	11-10=1	1	0
12	1	12	12-10=2	4	4
13	0	0	13-10=3	9	0
14	1	14	14-10=4	16	16
Total	**30**	**300**			**26**

O desvio padrão é dado por:

$$S = \sqrt{\frac{\sum_{i=1}^{n} \quad (X_i - \underline{X})^2 \,.\, f_i}{n-1}} = \sqrt{\frac{26}{30-1}} = \sqrt{\frac{26}{29}} < 1$$

Para tirar a prova, podemos elevar ao quadrado, os dois lados.

$$\left(\sqrt{\frac{26}{29}}\right)^2 < (1)^2$$

$$\frac{26}{29} < 1$$

$$0{,}89 < 1$$

Portanto, certa a questão.

Gabarito: Certa

Q.04 (CEBRASPE / IPHAN / Analista / 2018)

Cinco municípios de um estado brasileiro possuem as seguintes quantidades de patrimônios históricos: {2, 3, 5, 3, 2}.

Admitindo que a média e o desvio-padrão desse conjunto de valores sejam iguais a 3 e 1,2, respectivamente, julgue o item seguinte.

O coeficiente de variação é superior a 0,3 e inferior a 0,5.

C – CERTO

E – ERRADO

Comentários:

Vamos lá.

- *Média = 3*
- *Desvio Padrão = 1,2*

0,3 < Coeficiente de Variação < 0,5

$$Coeficiente\ de\ Variação = \frac{Desvio\ Padrão}{Média} = \frac{1{,}2}{3} = 0{,}4$$

Gabarito: Certo

Q.05 (CEBRASPE / IPHAN / Analista / 2018)

Cinco municípios de um estado brasileiro possuem as seguintes quantidades de patrimônios históricos: {2, 3, 5, 3, 2}.

Admitindo que a média e o desvio-padrão desse conjunto de valores sejam iguais a 3 e 1,2, respectivamente, julgue o item seguinte.

Para esse conjunto de valores, a variância é igual a 3.

C – CERTO

E – ERRADO

Comentários:

Opa!!

Cuidado aí para não vacilar e perder um tempo precioso, hein!!

A variância corresponde ao quadrado do desvio padrão. Como o DP foi fornecido no enunciado da questão. Logo:

$$Variância = 1{,}2^2 = 1{,}44$$

Gabarito: Errado

Q.06 (CEBRASPE - Analista (SERPRO) /Ciência de Dados/2021)

Considerando que o número X de erros registrados em determinado tipo de código computacional siga uma distribuição binomial com média igual a 4 e variância igual a 3, julgue o item a seguir.

O coeficiente de variação da distribuição de erros X é igual a 3.

C – CERTO

E – ERRADO

Comentários:

Os dados fornecidos na questão são os seguintes:

Média = 4

Variância =3

E a banca deseja saber se o coeficiente de variação é 3.

$$CV = \frac{desvio\ padrão}{média}$$

O desvio padrão é $\sqrt{3} \approx 1{,}73$. Logo,

$$CV = \frac{1{,}73}{4} = 0{,}43$$

Gabarito: Errado

Questões VUNESP

Q.07 (VUNESP - Analista Administrativo (EBSERH HC-UFU) / Estatística/2020)

Um aluno tirou as seguintes notas ao longo do semestre: 4, 8, 6, 1 e 6. A média, a mediana e o desvio padrão foram, respectivamente:

a) 5; 6 e 5,6.

b) 5; 5 e 5,6.

c) 5; 6 e 2,4.

d) 5; 6 e 31,6.

e) 6; 6 e 5,6.

Comentários:

Pessoal, os dados fornecidos na questão foram os seguintes:

4, 8, 6, 1 e 6

A primeira coisa a ser feita é ordenar em um ROL.

{1, 4, 6, 6, 8}

Temos 5 elementos e facilmente podemos observar que a mediana é 6.

{1, 4, 6, 6, 8}

Com isso, podemos eliminar a alternativa B.

Agora iremos calcular a média, para isso, basta somar todos os valores e dividir por 5.

$$\underline{X} = \frac{1+4+6+6+8}{5} = \frac{25}{5} = 5$$

Com isso, podemos eliminar a alternativa E e ficamos com as letras A, C e D.

Para calcular o desvio padrão, temos que encontrar a média dos quadrados dos desvios em relação à média.

X	$(X_i - \underline{X})$	$(X_i - \underline{X})^2$
1	1-5=-4	16
4	4-5=-1	1
6	6-5=1	1
6	6-5=1	1
8	8-5=3	9
Total		**28**

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \underline{X})^2}{n}} = \sqrt{\frac{28}{5}} = \sqrt{5{,}6}$$

Analisando as alternativas A, C e D. Podemos perceber que a resposta só pode ser a letra C.

$$\sigma \approx 2{,}4$$

Gabarito: C

Q.08 (VUNESP - Analista de Gestão (FITO) /Contabilidade/2020)

Na análise de um conjunto de valores, é muito importante calcular as medidas de tendência central e de dispersão, para se ter uma ideia das características dessa distribuição de dados. Em relação a essas medidas, é correto afirmar que

a) a moda, a média aritmética e a mediana nunca coincidem em valor.

b) se somarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a média aritmética e a variância ficam inalteradas.

c) se multiplicarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a nova variância fica multiplicada pelo quadrado dessa constante.

d) a soma dos desvios de cada valor do conjunto em relação à média apresenta normalmente um valor maior que zero.

e) se dividirmos cada elemento desse conjunto de valores por uma constante arbitrária maior que zero, o novo desvio padrão fica dividido pelo valor da constante ao quadrado.

Comentários:

Pessoal, essa é uma questão de propriedade das medidas de tendência central e de dispersão.

Vamos analisar cada alternativa:

Letra a) a moda, a média aritmética e a mediana ~~nunca~~ coincidem em valor.

Errada, pois eles podem sim coincidir, basta lembrar, por exemplo, de uma distribuição simétrica em que a média e mediana são iguais.

Letra b) se somarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a média aritmética e a variância ficam inalteradas.

Errada em relação à média. A média aritmética ficará somada a essa variável, mas a variância ficará inalterada.

Letra c) se multiplicarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a nova variância fica multiplicada pelo quadrado dessa constante.

Correta, pois a variância fica multiplicada pelo quadrado da constante.

Letra d) a soma dos desvios de cada valor do conjunto em relação à média apresenta normalmente um valor ~~maior que~~ zero.

Errada, pois sabemos que a soma dos desvios em relação à média é sempre zero.

Letra e) se dividirmos cada elemento desse conjunto de valores por uma constante arbitrária maior que zero, o novo desvio padrão fica dividido pelo valor da constante ~~ao quadrado~~.

Errada, pois o desvio padrão ficará dividido pela constante. Quem fica dividido pela constante ao quadrado é a variância.

Gabarito: C

Questões Bancas Diversas

Q.09 (IBFC - Supervisor de Pesquisas (IBGE) /Suporte Gerencial/2021)

Em dois grupos formados pela mesma quantidade de pessoas constatou-se que a média de idade do primeiro grupo é igual a 25 com variância de 16, e a média de idade do segundo grupo é igual a 40, com variância de 36. Nessas condições, é correto afirmar que o coeficiente de variação:

a) do primeiro grupo é maior que o do segundo grupo.

b) do primeiro grupo é menor que o do segundo grupo.

c) é igual para ambos os grupos.

d) é igual a 0,64 para o primeiro grupo e igual a 0,90 para o segundo grupo.

e) é igual a 1,6 para o primeiro grupo e igual a 1,5 para o segundo grupo.

Comentários:

Pessoal, essa é uma questão para comparar o coeficiente de variação de dois grupos.

$$CV = \frac{desvio\ padrão}{média}$$

Grupo 1:

Média = 25

Variância = 16

Desvio padrão = $\sqrt{variância} = \sqrt{16} = 4$

$$CV_{Grupo\ 1} = \frac{4}{25} = 0{,}16$$

Grupo 2:

Média = 40

Variância = 36

Desvio padrão = $\sqrt{variância} = \sqrt{36} = 6$

$$CV_{Grupo\ 2} = \frac{6}{40} = 0{,}15$$

Analisando as alternativas podemos observar que a resposta é a **letra A**, pois o CV foi maior para o primeiro grupo.

Gabarito: A

Q.10 (IBADE - Analista Público de Gestão (Pref Vila Velha)/Economista/2020)

Exemplos comuns de medidas de dispersão estatística são:

I – média;

II – mediana;

III – variância;

IV - desvio padrão;

V - amplitude interquartil.

Está(ão) correta(s):

a) somente V.

b) somente V e IV.

c) somente V, IV e III.

d) somente V, IV, III e II.

e) I, II, III, IV e V.

Comentários:

Pessoal, a média e a mediana são medidas de tendência central. Já a variância, desvio padrão e amplitude interquartil são medidas de dispersão absoluta, como vimos em aula. Logo, a letra C é a resposta.

Gabarito: C

Questões FGV

Q.11 (FGV/ Analista de Pesquisa e Informações (FunSaúde CE)/Estatística/2021)

Uma variável aleatória discreta X tem os seguintes valores possíveis e probabilidades associadas:

x	-1	1	3
p(x)	0,4	0,2	0,4

A variância de X é igual a

a) 2,0.

b) 2,4.

c) 2,8.

d) 3,2.

e) 3,6.

Comentários:

Considere a seguinte tabela:

X	P(X)
-1	0,4
1	0,2
3	0,4
Total	1,0

A primeira a ser feita é construir a coluna de X.P(X) e em seguida calcular a média.

X	P(X)	X.P(X)
-1	0,4	-0,4
1	0,2	0,2
3	0,4	1,2

Total	1,0	1,0

$$\bar{X} = \sum_{i=1}^{n} X_i \, . \, P(X_i) = 1{,}0$$

Para calcular a variância utilizaremos a seguinte fórmula.

$$Var(X) = \sum_{i=1}^{n} {X_i}^2 \, . \, P(X_i) - \bar{X}^2$$

X	P(X)	X.P(X)	X^2	X^2. P(X)
-1	0,4	-0,4	1^2=1	0,4
1	0,2	0,2	1^2=1	0,2
3	0,4	1,2	3^2=9	3,6
Total	1,0	1,0		4,2

Portanto, a variância será

$$Var(X) = 4{,}2 - 1{,}0^2$$

$$Var(X) = 4{,}2 - 1{,}0$$

$$Var(X) = \mathbf{3,2}$$

Gabarito: D

Q.12 (FGV/ Estatístico (SUSAM) / 2014)

Uma variável aleatória X tem média 4 e desvio padrão igual a 2. Se Y = 3X – 2 então a média e o desvio padrão de Y são, respectivamente,

a) 12 e 16.

b) 12 e 4.

c) 10 e 18.

d) 10 e 6.

e) 10 e 36.

Comentários:

Aqui temos uma questão de transformação de variável. Para resolver essa questão é bom ter o conhecimento das propriedades da média e desvio padrão.

As propriedades para a **média**:

- A soma ou subtração de uma variável aleatória "k" a todos os valores de um conjunto de dados, a média ficará somada ou subtraída por essa constante;
- A multiplicação ou divisão uma variável aleatória "k" a todos os valores de um conjunto de dados, a média fica multiplicado ou dividida por essa constante.

As propriedades para o **desvio padrão**:

- A soma ou subtração de uma constante "k" a todos os valores de um conjunto de dados, o desvio padrão não se alteram;
- A multiplicação ou divisão de uma constante "k" a todos os valores de um conjunto de dados, o desvio padrão fica multiplicado ou dividida por essa constante.

A variável X tem as seguintes medidas:

Média = 4

Desvio padrão = 2

A variável Y é a seguinte:

$$\boldsymbol{Y = 3X - 2}$$

Para encontrar a média e o desvio padrão de Y, basta utilizar as propriedades.

A média é afetada pelas 4 operações matemáticas.

$$Média\ de\ Y\ =\ 3\,.\,4 - 2\ =\ 12 - 2\ =\ \mathbf{10}$$

O desvio padrão é afetado apenas pela multiplicação e divisão.

$$Desvio\ padrão\ de\ Y\ =\ 3\,.\,2\cancel{-2}\ =\ 6 =\ \mathbf{6}$$

Gabarito: D

Q.13 (FGV / Analista Judiciário / 2022)

Os dados a seguir são uma amostra de idades:

26 28 30 32 32 34 36 38

O desvio padrão dessas idades é igual a

a) **2.**
b) **3.**
c) **4.**
d) **5.**
e) **6.**

Comentários:

Percebam que a distribuição 26 28 30 32 32 34 36 38 é simétrica, portanto a Média é o termo central (ou os termos centrais), logo a nossa média é 32.

Ou simplesmente:

$$\bar{X} = \frac{26+28+30+32+32+34+36+38}{8} = 32$$

Fiquem atentos ao fato de termos uma ***AMOSTRA***, portanto, para o cálculo do ***desvio padrão***, precisaremos ***dividir por (n – 1)***.

$$S^2 = \frac{(26-32)^2+(28-32)^2+(30-32)^2+(32-32)^2+(32-32)^2+(34-32)^2+(36-32)^2+(38-32)^2}{8-1}$$

$$S^2 = 16$$

$$s = 4$$

Gabarito: C

Q.14 (FGV / Pref. Paulínia / 2021)

Em uma turma de 10 alunos, as notas dos alunos em uma avaliação foram:

6	*7*	*7*	*8*	*8*	*8*	*8*	*9*	*9*	*10*

O desvio padrão dessa lista de notas é, aproximadamente,

a) 0,8.
b) 0,9.
c) 1,1.
d) 1,3.
e) 1,5.

Comentários:

Sabemos que o desvio padrão é a raiz quadrada da variância, ok?

A Média Aritmética das notas é igual 8.

$$S^2 = \frac{(6-8)^2 + (7-8)^2 + (7-8)^2 + (8-8)^2 + (8-8)^2 + (8-8)^2 + (8-8)^2 + (9-8)^2 + (9-8)^2 + (10-8)^2}{10}$$

$$S^2 = 1{,}2$$

$$S = 1{,}1$$

Gabarito: C

LISTA DE QUESTÕES ESTRATÉGICAS

Questões CEBRASPE

Q.01 (CEBRASPE (CESPE) /ACE (TCE-RJ) / Controle Externo/TI/2021)

X	frequência absoluta
0	5
1	10
2	20
3	15
total	50

Considerando que a tabela precedente mostra a distribuição de frequências de uma variável quantitativa X, julgue o item a seguir.

A variância amostral de X é superior a 0,89.

C - CERTO

E – ERRADO

Q.02 (CEBRASPE (CESPE) / Profissional de Tecnologia da Informação (ME) /2020)

Considerando que R representa uma variável quantitativa cuja média, mediana e variância são, respectivamente, iguais a 70, 80 e 100, e que $U = \frac{R}{10} - 7$*, julgue o próximo item, acerca das variáveis U e R.*

O desvio padrão da variável U é igual a 1.

C - CERTO

E - ERRADO

Q.03 (CEBRASPE (CESPE) / Professor (Pref São Cristóvão)/Matemática/Educação Básica/2019)

A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

idade (em anos)	*9*	*10*	*11*	*12*	*13*	*14*
quantidade de estudantes	*6*	*22*	*0*	*1*	*0*	*1*

A partir dessa tabela, julgue o item.

O desvio padrão das idades é inferior a 1ano.

C - Certo.

E – Errado.

Q.04 (CEBRASPE / IPHAN / Analista / 2018)

Cinco municípios de um estado brasileiro possuem as seguintes quantidades de patrimônios históricos: {2, 3, 5, 3, 2}.

Admitindo que a média e o desvio-padrão desse conjunto de valores sejam iguais a 3 e 1,2, respectivamente, julgue o item seguinte.

O coeficiente de variação é superior a 0,3 e inferior a 0,5.

C – CERTO

E – ERRADO

Q.05 (CEBRASPE / IPHAN / Analista / 2018)

Cinco municípios de um estado brasileiro possuem as seguintes quantidades de patrimônios históricos: {2, 3, 5, 3, 2}.

Admitindo que a média e o desvio-padrão desse conjunto de valores sejam iguais a 3 e 1,2, respectivamente, julgue o item seguinte.

Para esse conjunto de valores, a variância é igual a 3.

C – CERTO

E – ERRADO

Q.06 (CEBRASPE - Analista (SERPRO) /Ciência de Dados/2021)

Considerando que o número X de erros registrados em determinado tipo de código computacional siga uma distribuição binomial com média igual a 4 e variância igual a 3, julgue o item a seguir.

O coeficiente de variação da distribuição de erros X é igual a 3.

C – CERTO

E – ERRADO

Questões VUNESP

Q.07 (VUNESP - Analista Administrativo (EBSERH HC-UFU) / Estatística/2020)

Um aluno tirou as seguintes notas ao longo do semestre: 4, 8, 6, 1 e 6. A média, a mediana e o desvio padrão foram, respectivamente:

a) 5; 6 e 5,6.

b) 5; 5 e 5,6.

c) 5; 6 e 2,4.

d) 5; 6 e 31,6.

e) 6; 6 e 5,6.

Q.08 (VUNESP - Analista de Gestão (FITO) /Contabilidade/2020)

Na análise de um conjunto de valores, é muito importante calcular as medidas de tendência central e de dispersão, para se ter uma ideia das características dessa distribuição de dados. Em relação a essas medidas, é correto afirmar que

a) a moda, a média aritmética e a mediana nunca coincidem em valor.

b) se somarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a média aritmética e a variância ficam inalteradas.

c) se multiplicarmos uma constante arbitrária maior que zero a cada elemento desse conjunto de valores, a nova variância fica multiplicada pelo quadrado dessa constante.

d) a soma dos desvios de cada valor do conjunto em relação à média apresenta normalmente um valor maior que zero.

e) se dividirmos cada elemento desse conjunto de valores por uma constante arbitrária maior que zero, o novo desvio padrão fica dividido pelo valor da constante ao quadrado.

Questões Bancas Diversas

Q.09 (IBFC - Supervisor de Pesquisas (IBGE) /Suporte Gerencial/2021)

Em dois grupos formados pela mesma quantidade de pessoas constatou-se que a média de idade do primeiro grupo é igual a 25 com variância de 16, e a média de idade do segundo grupo é igual a 40, com variância de 36. Nessas condições, é correto afirmar que o coeficiente de variação:

a) do primeiro grupo é maior que o do segundo grupo.

b) do primeiro grupo é menor que o do segundo grupo.

c) é igual para ambos os grupos.

d) é igual a 0,64 para o primeiro grupo e igual a 0,90 para o segundo grupo.

e) é igual a 1,6 para o primeiro grupo e igual a 1,5 para o segundo grupo.

Q.10 (IBADE - Analista Público de Gestão (Pref Vila Velha)/Economista/2020)

Exemplos comuns de medidas de dispersão estatística são:

I – média;

II – mediana;

III – variância;

IV - desvio padrão;

V - amplitude interquartil.

Está(ão) correta(s):

a) somente V.

b) somente V e IV.

c) somente V, IV e III.

d) somente V, IV, III e II.

e) I, II, III, IV e V.

Questões FGV

Q.11 (FGV/ Analista de Pesquisa e Informações (FunSaúde CE)/Estatística/2021)

Uma variável aleatória discreta X tem os seguintes valores possíveis e probabilidades associadas:

x	-1	1	3
p(x)	0,4	0,2	0,4

A variância de X é igual a

a) 2,0.

b) 2,4.

c) 2,8.

d) 3,2.

e) 3,6.

Q.12 (FGV/ Estatístico (SUSAM) / 2014)

Uma variável aleatória X tem média 4 e desvio padrão igual a 2. Se Y = 3X – 2 então a média e o desvio padrão de Y são, respectivamente,

a) 12 e 16.

b) 12 e 4.

c) 10 e 18.

d) 10 e 6.

e) 10 e 36.

Q.13 (FGV / Analista Judiciário / 2022)

Os dados a seguir são uma amostra de idades:

26 28 30 32 32 34 36 38

O desvio padrão dessas idades é igual a

a) 2.
b) 3.
c) 4.
d) 5.
e) 6.

Q.14 (FGV / Pref. Paulínia / 2021)

Em uma turma de 10 alunos, as notas dos alunos em uma avaliação foram:

6	7	7	8	8	8	8	9	9	10

O desvio padrão dessa lista de notas é, aproximadamente,

a) 0,8.
b) 0,9.
c) 1,1.
d) 1,3.
e) 1,5.

Gabarito

1	2	3	4	5	6	7	8	9	10
CC	CC	CC	CC	ERR	ERR	C	C	A	C
11	12	13	14	15	16	17	18	19	20
D	D	C	C						

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.