Aula 00

TCE-PI (Auditor de Controle Externo - Área Comum) Passo Estratégico de Estatística - 2024 (Pós-Edital)

Autor:

Allan Maux Santana

24 de Agosto de 2024

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

Conceitos Iniciais

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:

INTRODUÇÃO À ESTATÍSTICA / GRÁFICOS / DIAGRAMAS / TABELAS / VARIÁVEIS DISCRETAS E CONTÍNUAS	Grau de incidência
FORMAS DE APRESENTAÇÃO DE DADOS	*75,0%*
TIPOS DE VARIÁVEIS	*16,5%*
CONCEITOS INICIAIS DE ESTATÍSTICA	*8,5%*
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:

Conceitos Iniciais

Vamos, de imediato, ao conceito de Estatística:

Basicamente, a Estatística tem por finalidade a ***Tomada de Decisões,*** acerca de um processo ou população, baseadas em intepretações de dados de uma amostra da população.

Podemos assim dividir a Estatística:

CONCEITOS BÁSICOS	
Estatística Descritiva Básica (Dedutiva)	*Estatística Inferencial Avançada (Indutiva)*

Descreve e Resume os dados, através da média, moda, mediana, variância, etc – são as primeiras etapas do processo estatístico, aqui ***não*** são tomadas as decisões.	Estuda processos para a partir de dados de uma amostra, ***inferir (concluir)*** algo de uma população, objetivando a tomada de decisões, sendo a parte final do processo estatístico.

Ainda, temos a ***Estatística Probabilística*** que consiste, basicamente, na aplicação da Teoria da Probabilidade.

Seu principal objetivo é o de calcular a chance de determinado Evento acontecer.
Vamos conhecer alguns elementos do estudo da Estatística:

CONCEITOS BÁSICOS		
POPULAÇÃO	***CENSO***	***AMOSTRA***
Conjunto de todos os elementos	Contagem de toda a População	Subconjunto não vazio da População

Dados Estatísticos

Vamos logo esquematizar esse conceito e sua classificação:

Dados Estatísticos – Conceito: (Observações/Características obtidas através de uma pesquisa)		
Univariado	***Bivariado***	***Multivariado***
Apenas uma característica do objeto	Duas características...	Mais de duas características...

Os dados estatísticos, quando do momento de sua coleta, se apresentam de forma ***desorganizada***, por isso recebem o nome de ***DADOS BRUTOS***.
Já a ***organização*** desses dados brutos em ordem, crescente ou decrescente, recebe o nome de ***ROL***.
Suponha uma pesquisa com 11 concurseiros a respeito da quantidade de dias que cada um estuda por semana. Seguem as respostas:

1, 3, 5, 2, 1, 6, 3, 4, 5, 7, 7

Obviamente, a forma como os dados estão dispostos não demonstra qualquer tipo de organização, sendo chamados de ***DADOS BRUTOS***.

Ao reescrever os dados em ordem, crescente ou decrescente, passaremos a ter um ***ROL***:

1, 1, 2, 3, 3, 4, 5, 5, 6, 7, 7

A diferença entre o maior valor (7) e o menor (1) é chamada de ***AMPLITUDE (6).***

Variáveis Estatísticas

Dentre os tópicos relacionados a essa parte inicial do nosso estudo, saber classificar as variáveis estatísticas é um do mais cobrados.

Mas, Allan, o que seria uma ***variável estatística***?

Variável Estatística é justamente o objeto de nosso estudo. É aquilo que estamos pesquisando. O levantamento estatístico é justamente focado nas variáveis pesquisadas. No nosso exemplo, seria a quantidade de dias da semana que o concurseiro estuda.

VARIÁVEIS	
Quantitativas	***Qualitativas***
Está relacionada a uma informação ***Numérica*** ***(Quanto?)*** Ex.: a idade de um grupo de pessoas	Está relacionada a um ***Atributo*** ***(Qual?)*** Ex.: o grau de instrução desse grupo

Acredito não haver dificuldades quanto à classificação entre ***Variáveis Quantitativas*** e ***Qualitativas***, confere?
Se a resposta à pesquisa for um ***número***, logo a variável será ***quantitativa***, caso seja um ***atributo*** (qualidade), será ***qualitativa***.
No entanto, temos ainda mais algumas classificações que fogem um pouco dessas tão triviais, mas nada de muito difícil, vejam?
Se, no momento de uma pesquisa, você tiver que responder a essas duas perguntas:

1. Quanto filhos você tem?
2. Qual a sua altura?

Podemos classificá-las como ***Variáveis Quantitativas***, ok?
Comparando-as, percebemos que existe uma simples, porém importante diferença entre as duas. Ainda não perceberam?

1. Quanto filhos você tem? (Aqui você faz uma ***contagem***)
2. Qual a sua altura? (Nessa você faz uma ***medição***)

Quantitativas	
Discretas	***Contínua***
Contagem	Medição

Uma outra forma de entender essa diferença é na forma de se encontrar o resultado. Quando a gente conta, temos um limite de possíveis resultados dentro de um intervalor, certo? Não dá para você responder à pesquisa dizendo que 2,38 filhos, dá? Já em relação à altura, podemos dizer sim que temos 1,735m, ok?

Então, ainda podemos assim diferenciá-las:

Quantitativas	
Discretas	***Contínua***
Valores ***restritos*** dentro de um intervalo de possíveis resultados	Valores ***irrestritos*** dentro de um intervalo de possíveis resultados

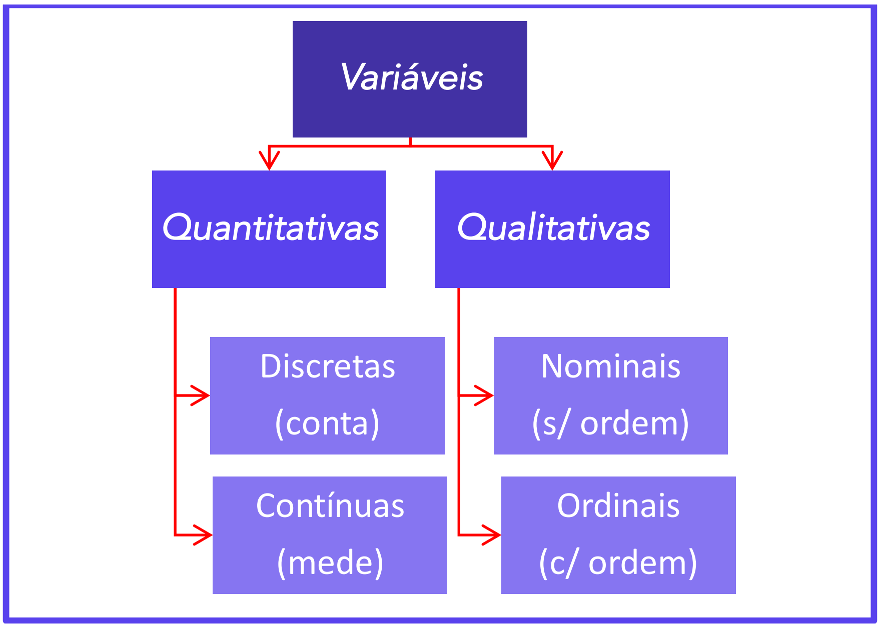
E se as perguntas fossem essas?

1. Qual o bairro que você mora?
2. Qual seu grau de instrução? (Fundamental, médio, superior, pós-graduação)

Facilmente, vemos que estamos diante de ***Variáveis Qualitativas***, certo? A primeira é chamada de ***nominais***, já a segunda de ***ordinais***, visto que a gente pode organizá-las (ordená-las) de maneira hierárquica.

Qualitativas	
Nominais	***Ordinais***
Não há hierarquia	Há hierarquia (ordem)

Resumindo:



Guardem esse esquema, ele lhe ajudará com umas algumas questões teóricas.

Distribuições de Frequências

Vamos voltar ao nosso exemplo do início da aula.
Suponha uma pesquisa com 11 concurseiros a respeito da quantidade de dias que cada um estuda por semana. Seguem as respostas:
Ao reescrever os dados em ordem, ***crescente*** ou ***decrescente***, passaremos a ter um ***ROL***:

1, 1, 2, 3, 3, 4, 5, 5, 6, 7, 7

Podemos organizar esses valores da seguinte forma:

QUANT. DIAS/SEMANA	FREQUÊNCIA
1	2
2	1
3	2
4	1
5	2
6	1
7	2
TOTAL	11

Essa ***Tabela*** é conhecida como ***Distribuição*** de ***Frequência de Dado não-agrupados***.
Frequência nada mais é do que a quantidade de vezes que cada valor de uma variável aparece num levantamento de dados. Fácil demais...

Vejam que falamos em ***dados não-agrupados***, pois os valores que estão na tabela são ***individuais***. E se a pesquisa fosse feita sobre a idade de 50 pessoas, vocês acham que essa tabela poderia ser individualizada? Ou seria mais eficiente apresentá-la através de um grupo de intervalos?

IDADES	FREQUÊNCIA ABSOLUTA (fi)	FREQUÊNCIA RELATIVA (Fi)
0 ⊢ 10	12	24%
10 ⊢ 20	5	10%
20 ⊢ 30	13	26%
30 ⊢ 40	5	10%
40 ⊢ 50	6	12%
50 ⊢ 60	9	18%
TOTAL	50	100%

Percebam que aqui os dados foram agrupados de acordo com a faixa com cada faixa etária.

> *0 ⊢ 10* : Nessa representação, chamada de ***classe***, o 0 (zero é incluído), já o 10 é excluído do intervalo.

No exemplo, temos um total de ***06 classes***.
A ***Amplitude*** de cada intervalo de classe é a diferença entre o Limite Superior e o Inferior, que nosso caso será de: (10 – 0) = ***10***

> **h = Amplitude de Classe = $L_{superior}$ - $L_{inferior}$**

Para determinarmos o ***Ponto Médio (PM)*** de cada Classe, basta calcularmos a média aritmética simples entre os limites de cada classe. Vejam:

IDADES	PONTO MÉDIO (PM)
0 ⊢ 10	***5*** = (0 + 10)/2
10 ⊢ 20	***15*** = (10 + 20)/2
20 ⊢ 30	***25*** = (20 + 30)/2

30 ⊢ 40	***35*** = (30 + 40)/2
40 ⊢ 50	***45*** = (40 + 50)/2
50 ⊢ 60	***55*** = (50 + 60)/2

Já a ***Amplitude Total*** é a diferença entre o Limite Superior da última classe e o Limite Inferior da primeira classe, sendo o tamanho de todo o conjunto a ser observado.

Representações Gráficas de Dados

Aqui vamos estudar os famosos gráficos. Precisaremos diferenciar cada tipo de gráfico, meus amigos, por nome, uso e finalidade e, também, saber interpretá-los.

GRÁFICO EM LINHAS: Séries Temporais – evolução de valores com o passar de tempo:

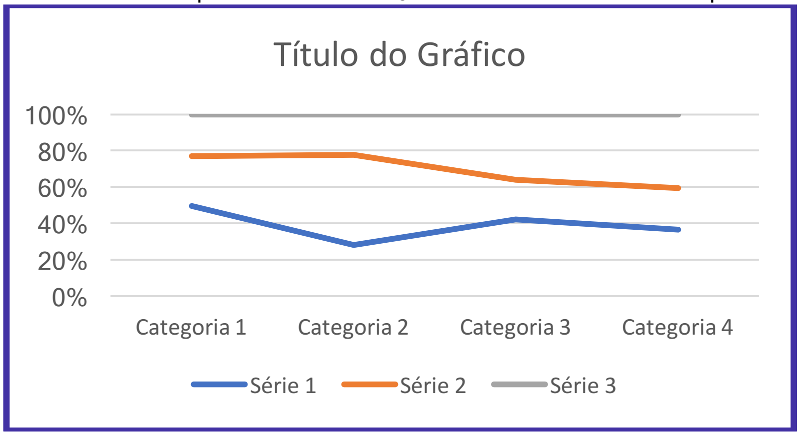


GRÁFICO DE COLUNAS: Séries Estatísticas.

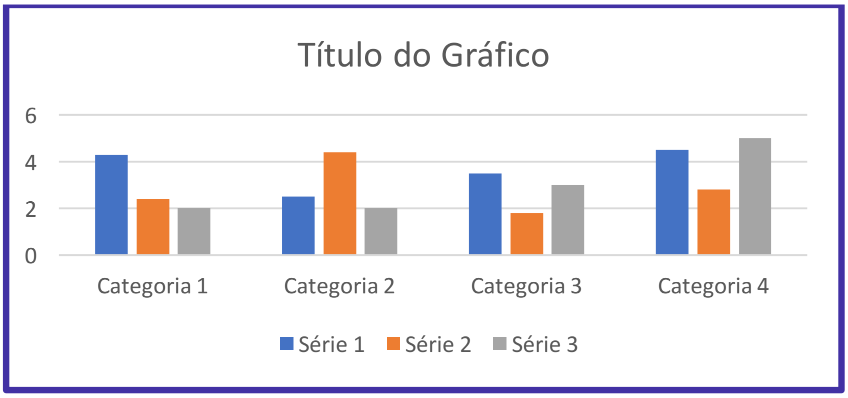
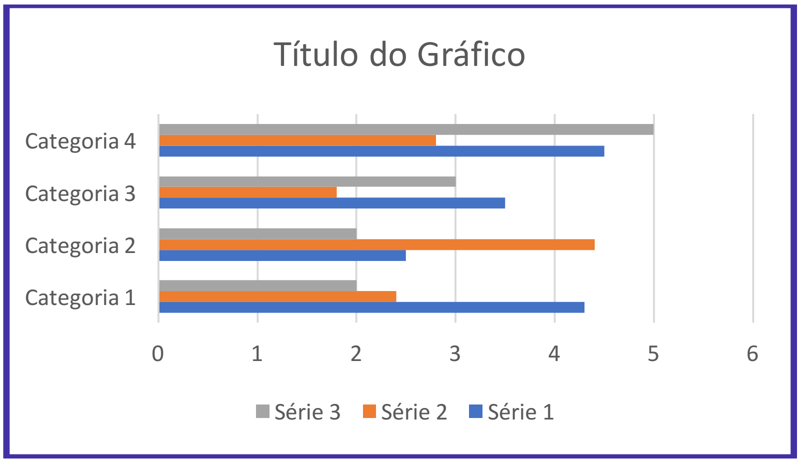
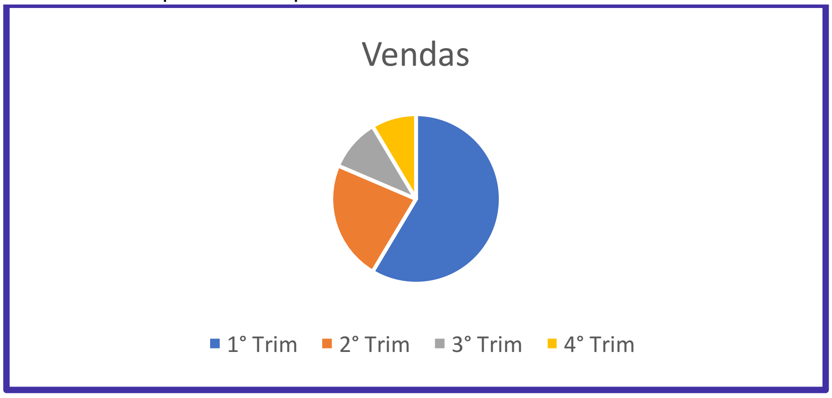


GRÁFICO DE BARRAS: Séries Estatísticas.

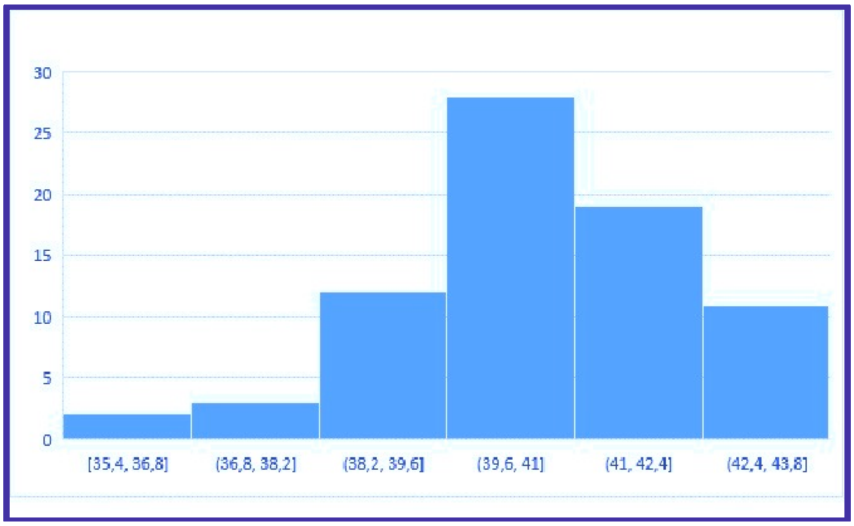
8



GRÁFICOS DE SETORES: Comparar Proporções.



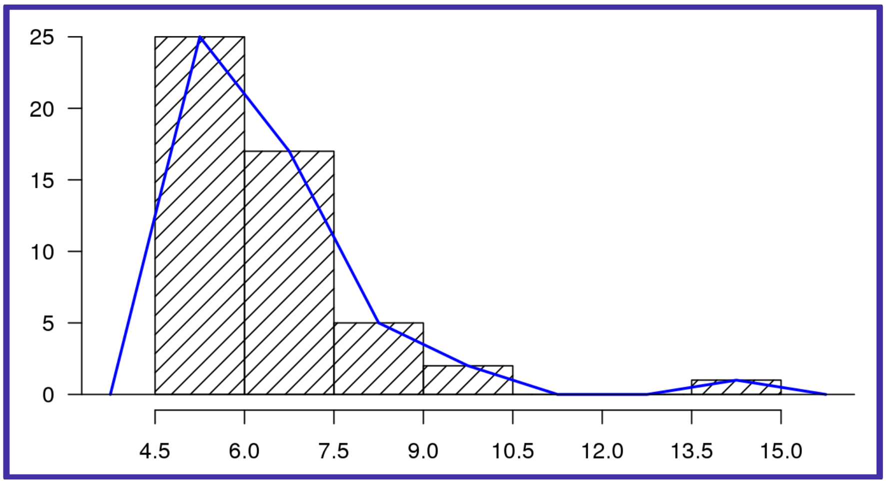
HISTOGRAMA: Utilizado em Dados Agrupados – Distribuição de Frequência em Classes.



O histograma é bem parecido com o gráfico de colunas, porém reparem que no histograma não existem espaço entre as colunas.

9

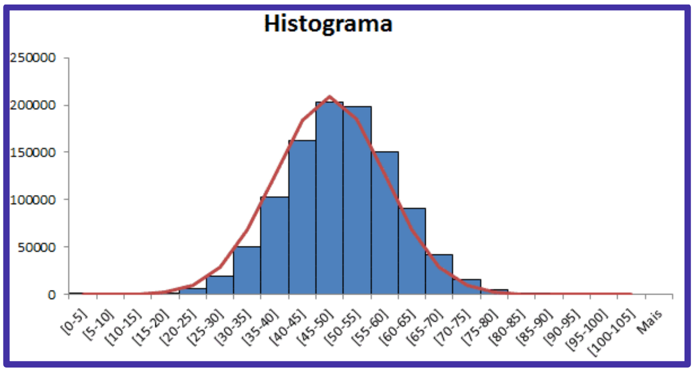
POLÍGONO DE FREQUÊNCIAS: Obtido através da união dos pontos médios das bases superiores das colunas do Histograma.



CURVA DE FREQUÊNCIAS: obtida a partir do ***Polígono de Frequências*** – Polígono de Frequência ***Polido***.

POLÍGONO DE FREQUÊNCIA	CURVA DE FREQUÊNCIA
Imagem real	*Imagem tendenciosa*

Basicamente o que é feito é a eliminação dos vértices do Polígono de Frequência. Vejam como ficaria:



A linha vermelha seria justamente nossa ***curva de frequência*** simétrica.

Porem, podemos ter uma ***assimetria*** à esquerda conforme ***gráfico 1***, ou à direita ***gráfico 2***.

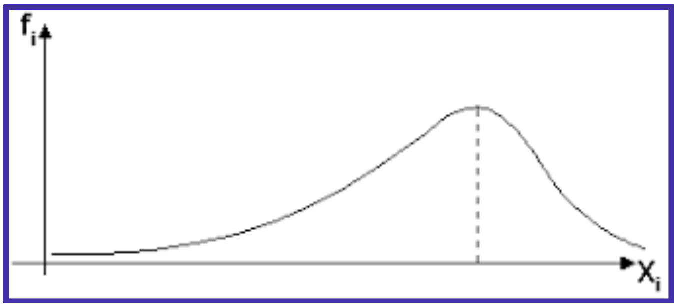


Gráfico 1

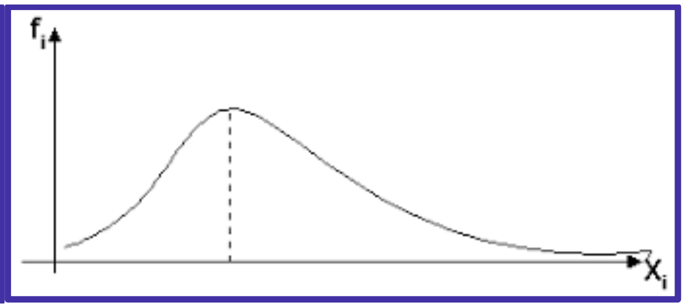
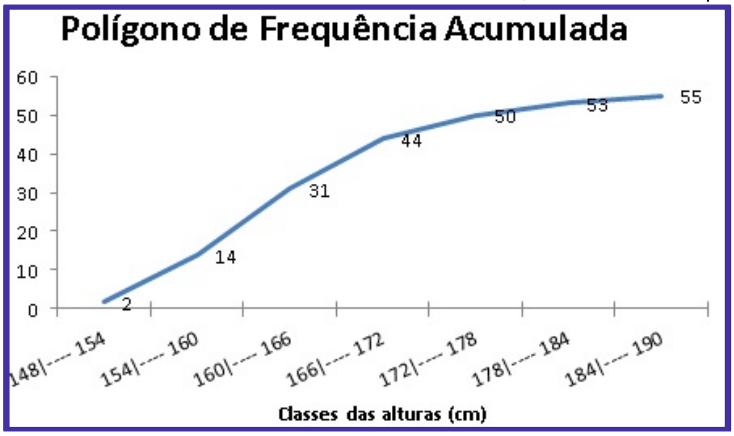


Gráfico 2

$$\text{Frequência Polida} = \frac{f_{anterior} + 2\, f_{classe\ considerada} + f_{posterior}}{4}$$

OGIVAS: Obtido através de ***Frequências Acumuladas*** – Polígono de Frequência Acumulada.



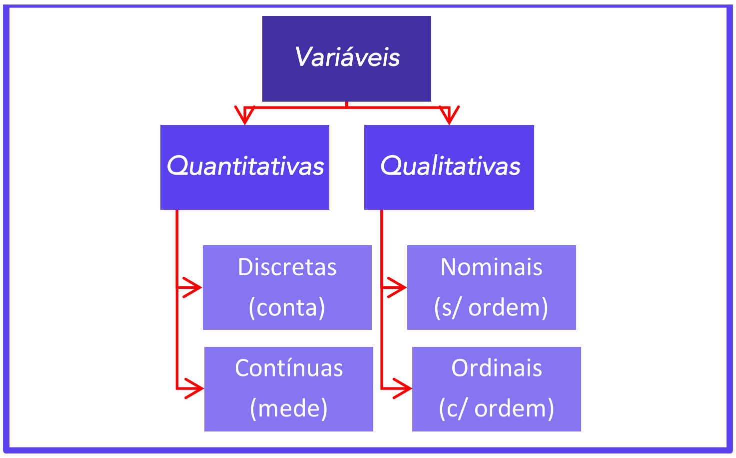
Aposta Estratégica

A ideia desta seção é apresentar os pontos do conteúdo que mais possuem chances de serem cobrados em prova, considerando o histórico de questões da banca em provas de nível semelhante à nossa, bem como as inovações no conteúdo, na legislação e nos entendimentos doutrinários e jurisprudenciais1.

Pessoal, nossa Aposta Estratégica, obviamente, não poderia deixar de ser outra:

[1] Vale deixar claro que nem sempre será possível realizar uma aposta estratégica para um determinado assunto, considerando que às vezes não é viável identificar os pontos mais prováveis de serem cobrados a partir de critérios objetivos ou minimamente razoáveis.

Classificação das Variáveis



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 / COMPESA/ 2018)

A COMPESA, em uma pesquisa de satisfação dos usuários, preparou um formulário para traçar os perfis de seus clientes e o grau de satisfação com os serviços da empresa,

Em um formulário, ela solicitou os dados a seguir.

I. Idade.

II. Grau de escolaridade.

III. Faixa de renda familiar.
IV. Nota dada ao serviço.

Assinale a opção que contempla apenas variáveis categóricas.

a) ***I e II.***
b) ***II e III.***
c) ***III e IV.***
d) ***I, II e III.***
e) ***I, II e IV.***

Comentários:

I – Idade: Variável Quantitativa
II - Grau de Escolaridade: ***Variável Qualitativa – Categórica***
III – Faixa de Renda Familiar (classe baixa, classe média, classe alta): ***Variável Qualitativa***
IV – Nota dada ao Serviço: Variável Quantitativa

Gabarito: B

Q. 02 (FGV/Auditor (ALBA) / Auditoria/2014)

Observe a tabela de frequências a seguir, que se refere aos saldos em conta, num determinado dia, de duzentas contas-correntes:

Saldos em conta (R$)	Frequência
Até 100,00	8
De mais de 100,00 a 300,00	28
De mais de 300,00 a 500,00	46
De mais de 500,00 a 700,00	54
De mais de 700,00 a 900,00	44
De mais de 900,00 a 1.100,00	13
De mais de 1.100,00 a 1.300,00	6
Acima de 1.300,00	1

A frequência relativa acumulada de saldos em R$ 900,00 é igual a

a) 22%.

b) 36%.

c) 54%.

d) 90%.

e) 97%.

Comentários:

Nessa questão temos apenas que encontrar a frequência relativa acumulada de saldos em R$ 900,00. Além disso, é dito que temos 200 contas-corretes. Podemos fazer isso de duas formas.

Primeira forma: encontrar a frequência relativa de cada classe e depois fazer a acumulação até R$ 900,00. Para tanto, basta dividir cada frequência absoluta por 200 e depois multiplicar por 100%.

Saldos em conta (R$)	Frequência	Frequência Relativa	Freq. Relativa Acum.
Até 100,00	8	$\frac{8}{200}\,.\,100\% = 4\%$	4%
De mais de 100,00 a 300,00	28	$\frac{28}{200}\,.\,100\% = 14\%$	18%
De mais de 300,00 a 500,00	46	$\frac{46}{200}\,.\,100\% = 23\%$	*41%*
De mais de 500,00 a 700,00	54	$\frac{54}{200}\,.\,100\% = 27\%$	*68%*
De mais de 700,00 a 900,00	44	$\mathbf{\frac{44}{200}\,.\,100\% = 22\%}$	90%
De mais de 900,00 a 1.100,00	13	$\frac{13}{200}\,.\,100\% = 6{,}5\%$	*96,5%*
De mais de 1.100,00 a 1.300,00	6	$\frac{6}{200}\,.\,100\% = 3\%$	99,5%
Acima de 1.300,00	1	$\frac{1}{200}\,.\,100\% = 0{,}5\%$	100%

Vejam que até a acumulação de R$ 900,00 temos 90%. Portanto, letra "D" a resposta.

Segunda forma: encontrar a frequência absoluta acumulada.

Saldos em conta (R$)	Frequência	Freq. Abs. Acum
Até 100,00	8	*8*
De mais de 100,00 a 300,00	28	36
De mais de 300,00 a 500,00	46	82
De mais de 500,00 a 700,00	54	136
De mais de 700,00 a 900,00	44	180
De mais de 900,00 a 1.100,00	13	193
De mais de 1.100,00 a 1.300,00	6	199
Acima de 1.300,00	1	200

Vejam que até a acumulação de R$ 900,00 temos um acumulado de 180.

Para encontrar a frequência acumulada correspondente, basta dividir por 200 e depois multiplicar por 100%.

$$\frac{180}{200}.100\% = 90\%$$

Portanto, letra "D" a resposta.

Gabarito: D

Questões CEBRASPE

Q.03 (CESPE - CEBRASPE / DEPEN / 2015)

O diretor de um sistema penitenciário, com o propósito de estimar o percentual de detentos que possuem filhos, entregou a um analista um cadastro com os nomes de 500 detentos da instituição para que esse profissional realizasse entrevistas com os indivíduos selecionados.

A partir dessa situação hipotética e dos múltiplos aspectos a ela relacionados, julgue o item seguinte, referente a técnicas de amostragem.

A diferença entre um censo e uma amostra consiste no fato de esta última exigir a realização de um número maior de entrevistas.

C - Certo.

E - Errado.

Comentários:

CONCEITOS BÁSICOS	
CENSO	AMOSTRA
Contagem de toda a População	Subconjunto não vazio da População

Percebam que o examinador inverteu os conceitos, logo, o correto seria:

A diferença entre uma amostra e um censo consiste no fato de esta última exigir a realização de um número maior de entrevistas.

Gabarito: Errado

Q.04 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.

15

O tipo de sangue, a cor dos olhos e o número de filhos de famílias de baixa renda são todas variáveis qualitativas.
C - Certo.
E - Errado.
Comentários:
Tipo de Sangue e Cor dos Olhos são ***variáveis qualitativas***.
Número de filhos é uma ***variável quantitativa discreta***.

Gabarito: Errado

Q.05 (CESPE - CEBRASPE / Banco da Amazônia S.A. / 2006)
Na estatística, as medidas descrevem fenômenos em termos que podem ser analisados de forma objetiva. Acerca desse assunto, julgue o seguinte item.
O número de filhos é uma variável categórica ordinal, pois pode ser quantificada.
C - Certo.
E - Errado.
Comentários:
O número de filhos é uma variável ***quantitativa discreta***.

Gabarito: Errado

Q.06 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
Os valores das variáveis quantitativas discretas não podem ser contados, mas apenas mensurados.
C - Certo.
E - Errado.
Comentários:
Os valores das variáveis quantitativas discretas podem ser contados.

Gabarito: Errado

Q.07 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)
A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
O peso de crianças é considerado uma variável quantitativa contínua.
C - Certo.
E -Errado.
Comentários:

Dentro de um intervalo de dados, o peso de uma criança pode assumir qualquer valor. Além disso, vimos que as variáveis que são medidas são as ***quantitativas contínuas***. Logo, nosso enunciado está corretíssimo.

Gabarito: Correto

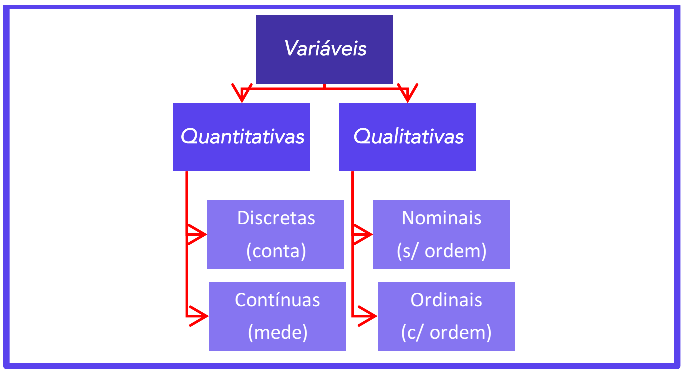
Q.08 (CESPE - CEBRASPE / Polícia Federal / 2005)

Um projeto de serviços de assistência social foi desenvolvido para ser implementado em todas as delegacias e plantões policiais de um estado brasileiro. Porém, antes da sua aplicação em todo o estado, ele foi implementado em 10 municípios, em caráter experimental, por 12 meses. Esses municípios foram escolhidos aleatoriamente entre os 250 municípios do estado. Nesse período experimental, foram registradas 48.000 ocorrências nos 10 municípios selecionados. Em 25% dessas ocorrências, as pessoas envolvidas foram encaminhadas aos assistentes sociais. A partir dessas ocorrências, os 100 assistentes sociais envolvidos nesse projeto atenderam, em média, 500 pessoas por mês. Os resultados obtidos foram positivos, observando-se uma queda na reincidência de denúncias e ocorrências registradas nesses municípios após a implementação do projeto.

A partir dos dados apresentados no texto acima, julgue o item subsequente.

O número de ocorrências registradas durante o período experimental de 12 meses nos 10 municípios selecionados (48.000) é a realização de uma variável aleatória contínua.
C - Certo.
E - Errado.
Comentários:

Vejam que estamos querendo classificar o número de ocorrência registradas. Vamos dar uma breve lembrada:



A quantidade de ocorrências é medida ou contada?
Concordam comigo que ela é contada? Logo, temos uma ***Variável Discreta***.

Gabarito: Errado

Q.09 (CESPE - CEBRASPE / Agência Nacional de Saúde Complementar / 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.

O peso de crianças é considerado uma variável quantitativa contínua.

C - Certo.

E - Errado.

Comentários:

Aquilo que pode ser medido é considerada uma Variável Quantitativa Contínua.

Gabarito: Correto

Q.10 (CESPE – CEBRASPE / Instituto Jones dos Santos Neves - ES / 2010)

quantidade *X* de processos analisados por servidor	frequência
1	6
2	9
3	10
4	3
5	2
total de servidores	30

Considerando que a tabela acima mostra a distribuição de frequências da quantidade X de processos que cada servidor de certo órgão público analisou em determinada semana, julgue o item a seguir.

A variável X é quantitativa discreta.

C - Certo.

E - Errado.

Comentários:

Perfeitamente correta nossa assertiva. Visto que "x" representa a quantidade de processos, logo podemos contá-los.

Gabarito: Correto

Q.11 (CESPE – CEBRASPE / BACEN / 2013)

2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários. Com base nesses dados, julgue o próximo item.

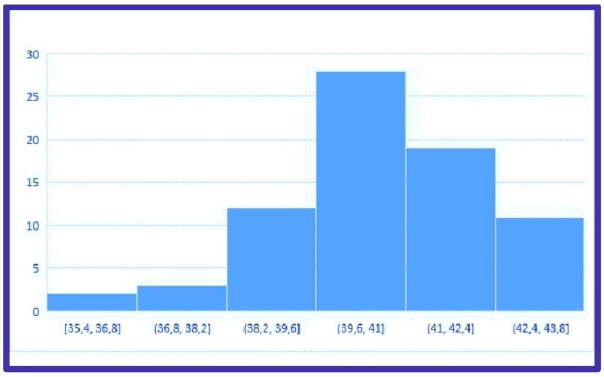
É inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos; dessa forma, apenas por meio de um gráfico de barras pode ser realizada a representação gráfica.

C - Certo.

E - Errado.

Comentários:

HISTOGRAMA: Utilizado em Dados Agrupados – Distribuição de Frequência em Classes.



Podemos muito bem agrupar os dados da questão, ok? O histograma é utilizado para indicar a frequência que certo dado aparece.

Gabarito: Errado

Q.12 (CESPE – CEBRASPE / BACEN / 2013)

número diário de denúncias registradas (X)	frequência relativa
0	0,3
1	0,1
2	0,2
3	0,1
4	0,3
total	**1,0**

A tabela precedente apresenta a distribuição de frequências relativas da variável X, que representa o número diário de denúncias registradas na ouvidoria de determinada instituição pública. A partir das informações dessa tabela, julgue o item seguinte.

A variável X é do tipo qualitativo nominal.

C - Certo.

E - Errado.

Comentários:

A variável "X" representa o número (quantidade) de denúncias registradas, logo temos uma ***Variável Quantitativa Discreta***.

Gabarito: Errado

Questões FCC

Q.13 (FCC / Defensoria Pública do Estado de SP/ 2009)

Sobre estatística aplicada, é correto o que se afirma em:

a) ***Parâmetros são medidas características de grupos, determinadas por meio de uma amostra aleatória.***
b) ***A estatística descritiva é a técnica pela são coletados dados de uma amostra, a partir do que são tomadas decisões sobre uma determinada população.***
c) ***A caracterização de uma população se dá por meio da observação de todos os seus componentes que a integram.***
d) ***A estatística inferencial compreende um conjunto de técnicas destinadas à síntese de dados numéricos.***
e) ***Censo é o processo utilizado para se medir as características de todos os membros de uma dada população.***

Comentários

Alternativa A: Errada

Parâmetros são valores ***fixos***. Portanto, não variam como as variáveis aleatórias. Ainda mais, eles são determinados pela população.

Alternativa B: Errada

CONCEITOS BÁSICOS	
Estatística Descritiva	*Estatística Inferencial*
Descreve os dados, através da média, moda, mediana, variância, etc	Estuda processos para a partir de dados de uma amostra, inferir (concluir) algo de uma população

Alternativa C: Errada

Uma população é caracterizada por meio de ***medidas*** de ***posição*** e de ***dispersão***.

Alternativa D: Errada

Conforme tabela da Alternativa "B".

Alternativa E: Certa

A Alternativa "E" traduz perfeitamente o conceito de ***Censo***.
Gabarito: E

Prof. Allan Maux

LISTA DE QUESTÕES ESTRATÉGICAS

Questões FGV

Q.01 (FGV / COMPESA/ 2018)

A COMPESA, em uma pesquisa de satisfação dos usuários, preparou um formulário para traçar os perfis de seus clientes e o grau de satisfação com os serviços da empresa,
Em um formulário, ela solicitou os dados a seguir.
I. Idade.
II. Grau de escolaridade.
III. Faixa de renda familiar.
IV. Nota dada ao serviço.

Assinale a opção que contempla apenas variáveis categóricas.

a) *I e II.*
b) *II e III.*
c) *III e IV.*
d) *I, II e III.*
e) *I, II e IV.*

Q. 02 (FGV/Auditor (ALBA)/Auditoria/2014)

Observe a tabela de frequências a seguir, que se refere aos saldos em conta, num determinado dia, de duzentas contas-correntes:

Saldos em conta (R$)	Frequência
Até 100,00	8
De mais de 100,00 a 300,00	28
De mais de 300,00 a 500,00	46
De mais de 500,00 a 700,00	54
De mais de 700,00 a 900,00	44
De mais de 900,00 a 1.100,00	13
De mais de 1.100,00 a 1.300,00	6

Acima de 1.300,00	1

A frequência relativa acumulada de saldos em R$ 900,00 é igual a

a) *22%.*
b) *36%.*
c) *54%.*
d) *90%.*
e) *97%.*

Questões CEBRASPE

Q.03 (CESPE - CEBRASPE / DEPEN / 2015)
O diretor de um sistema penitenciário, com o propósito de estimar o percentual de detentos que possuem filhos, entregou a um analista um cadastro com os nomes de 500 detentos da instituição para que esse profissional realizasse entrevistas com os indivíduos selecionados.

A partir dessa situação hipotética e dos múltiplos aspectos a ela relacionados, julgue o item seguinte, referente a técnicas de amostragem.

A diferença entre um censo e uma amostra consiste no fato de esta última exigir a realização de um número maior de entrevistas.
C - Certo.
E - Errado.
Q.04 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)
A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
O tipo de sangue, a cor dos olhos e o número de filhos de famílias de baixa renda são todas variáveis qualitativas.
C - Certo.
E - Errado.
Q.05 (CESPE - CEBRASPE / Banco da Amazônia S.A. / 2006)
Na estatística, as medidas descrevem fenômenos em termos que podem ser analisados de forma objetiva. Acerca desse assunto, julgue o seguinte item.
O número de filhos é uma variável categórica ordinal, pois pode ser quantificada.
C - Certo.
E - Errado.

Q.06 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
Os valores das variáveis quantitativas discretas não podem ser contados, mas apenas mensurados.
C - Certo.
E - Errado.

Q.07 (CESPE - CEBRASPE / Agência Nacional de Saúde Suplementar/ 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
O peso de crianças é considerado uma variável quantitativa contínua.
C - Certo.
E -Errado.

Q.08 (CESPE - CEBRASPE / Polícia Federal / 2005)

Um projeto de serviços de assistência social foi desenvolvido para ser implementado em todas as delegacias e plantões policiais de um estado brasileiro. Porém, antes da sua aplicação em todo o estado, ele foi implementado em 10 municípios, em caráter experimental, por 12 meses. Esses municípios foram escolhidos aleatoriamente entre os 250 municípios do estado. Nesse período experimental, foram registradas 48.000 ocorrências nos 10 municípios selecionados. Em 25% dessas ocorrências, as pessoas envolvidas foram encaminhadas aos assistentes sociais. A partir dessas ocorrências, os 100 assistentes sociais envolvidos nesse projeto atenderam, em média, 500 pessoas por mês. Os resultados obtidos foram positivos, observando-se uma queda na reincidência de denúncias e ocorrências registradas nesses municípios após a implementação do projeto.

A partir dos dados apresentados no texto acima, julgue o item subsequente.

O número de ocorrências registradas durante o período experimental de 12 meses nos 10 municípios selecionados (48.000) é a realização de uma variável aleatória contínua.
C - Certo.
E - Errado.

Q.09 (CESPE - CEBRASPE / Agência Nacional de Saúde Complementar / 2005)

A identificação do tipo de variável é um requisito importante para a escolha do teste estatístico mais adequado. Acerca das variáveis, julgue o seguinte item.
O peso de crianças é considerado uma variável quantitativa contínua.
C - Certo.
E - Errado.

Q.10 (CESPE – CEBRASPE / Instituto Jones dos Santos Neves - ES / 2010)

quantidade X de processos analisados por servidor	frequência
1	6
2	9
3	10
4	3
5	2
total de servidores	30

Considerando que a tabela acima mostra a distribuição de frequências da quantidade X de processos que cada servidor de certo órgão público analisou em determinada semana, julgue o item a seguir.
A variável X é quantitativa discreta.
C - Certo.
E - Errado.

Q.11 (CESPE – CEBRASPE / BACEN / 2013)

2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários. Com base nesses dados, julgue o próximo item.
É inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos; dessa forma, apenas por meio de um gráfico de barras pode ser realizada a representação gráfica.
C - Certo.
E - Errado.

Q.12 (CESPE – CEBRASPE / BACEN / 2013)

número diário de denúncias registradas (X)	frequência relativa
0	0,3
1	0,1
2	0,2
3	0,1
4	0,3
total	**1,0**

A tabela precedente apresenta a distribuição de frequências relativas da variável X, que representa o número diário de denúncias registradas na ouvidoria de determinada instituição pública. A partir das informações dessa tabela, julgue o item seguinte.

A variável X é do tipo qualitativo nominal.
C - Certo.
E - Errado.

Questões FCC

Q.13 (FCC / Defensoria Pública do Estado de SP/ 2009)

Sobre estatística aplicada, é correto o que se afirma em:

a) *Parâmetros são medidas características de grupos, determinadas por meio de uma amostra aleatória.*
b) *A estatística descritiva é a técnica pela são coletados dados de uma amostra, a partir do que são tomadas decisões sobre uma determinada população.*
c) *A caracterização de uma população se dá por meio da observação de todos os seus componentes que a integram.*
d) *A estatística inferencial compreende um conjunto de técnicas destinadas à síntese de dados numéricos.*
e) *Censo é o processo utilizado para se medir as características de todos os membros de uma dada população.*

Gabarito

1	2	3	4	5	6	7	8	9	10
B	D	ERR	ERR	ERR	ERR	CER	ERR	CER	CER
11	12	13	14	15	16	17	18	19	20
ERR	ERR	E	*	*	*	*	*	*	*

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**

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

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**

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

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**

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

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:

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

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.

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

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

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

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
Total	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:

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

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

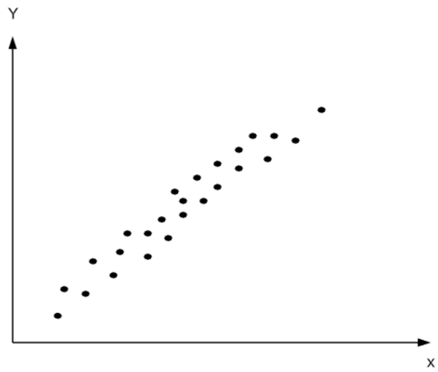
σ_Y é o desvio padrão de Y

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

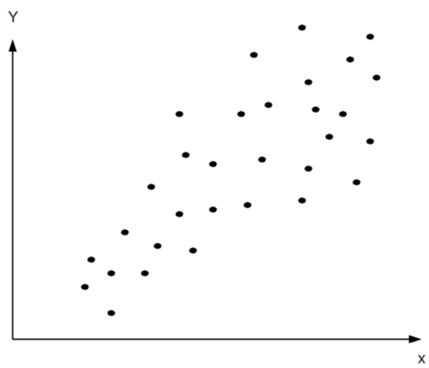
- Se $\rho(X,Y) = -1$, significa que as variáveis possuem correlação negativa perfeita. Se X aumentar em uma unidade, Y diminui na mesma magnitude;
- Se $\rho(X,Y) = 0$, significa que as variáveis não são linearmente correlacionadas. Uma variação em X não está associada a uma variação em Y de forma linear;
- 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	$4 \cdot 0{,}03 = 0{,}12$
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^2e 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):

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

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

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

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

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

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

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

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

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) = 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).

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

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.