Revista Pedagogia – UFMT V.5, n.2 (2018) Jul/Dez 2018

AÇÕES PEDAGÓGICAS NO ENSINO DA ESTATÍSTICA BÁSICA

Marcia Martins da Cruz dos Santos[1]
Rafael Martins Custódio Mendonça[2]
Sammia Shum de Oliveira Jesus Costa[2]
Marcos Antônio de Souza[2]
Juracy Mendes Moreira[3]

RESUMO

A estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos realizados em qualquer área do conhecimento. O objetivo proposto neste trabalho foi o de mostrar as vantagens do uso da estatística em nosso cotidiano na tentativa de despertar o interesse do aluno na aprendizagem da disciplina. As medidas de tendência central são denominadas de estatísticas por representar um conjunto de dados de forma condensada, enquanto que as medidas de dispersão são utilizadas para avaliar o grau de variabilidade ou dispersão dos valores em torna da média. Testes de hipóteses são técnicas da estatística inferencial que a partir de estudos feitos em um conjunto de dados amostrais, pode-se inferir sobre a população. Acreditamos que o conteúdo sobre estatística possa ser desenvolvido por alunos com intuito de mostrar a realidade da própria cidade, e que a participação dos alunos poderá desenvolver o interesse pela disciplina, aguçando sua curiosidade na busca de informações ate então desconhecidas.

PALAVRAS-CHAVE: Centralidade, Dispersão, Estatística e Inferência.

INTRODUÇÃO

Atualmente, o ensino da estatística se faz fundamental, visto que inúmeras informações demonstradas estatisticamente são vinculadas nos meios de comunicação, fazendo com que o cidadão necessite cada vez mais do conhecimento da estatística na hora de interpretar um gráfico, uma tabela e ate mesmo na tomada de decisão. A representação gráfica das series

1 *Pedagoga. Faculdade Montes Belos E-mail: marcia100069@hotmail.com*

2 *Prof. Faculdade Montes Belos E-mail: juramendes94@gmail.com*

3 *Autor para correspondência. Faculdade Montes Belos. E-mail: juramendes94@gmail.com*

estatísticas tem por finalidade representar os resultados obtidos, permitindo chegar-se a conclusões sobre a evolução do fenômeno ou como se relacionam os valores da serie. A escolha do tipo de gráfico ficara a critério do analista: Contudo, elementos como simplicidade, clareza e veracidade das informações devem ser considerados quando da elaboração de um gráfico. Assuntos ligados vários setores da economia, política, esporte, saúde publica entre outros, são publicados na maioria das vezes sob a forma gráfica. Assim sendo é de fundamental importância do conhecimento da estatística na leitura de uma informação representada graficamente. Geralmente, as pesquisas são realizadas através do estudo dos elementos que compõem uma amostra extraída da população que se pretende analisar. O conceito de população é intuitivo; trata-se do conjunto de indivíduos ou objetos que apresentam em comum determinadas características definidas para o estudo. É compreensível que o estudo de todos os elementos da população possibilita preciso conhecimento das variáveis que estão sendo pesquisadas; porem, nem sempre é possível obter informações de todos os elementos da população, limitações de tempo, custo e as vantagens do uso de técnicas estatísticas de inferência justificam o uso de planos amostrais.

Segundo Shulman (1986) a utilização da estatística como ferramenta em nosso cotidiano tem se tornado cada vez mais frequente em qualquer atividade profissional da vida moderna nos mais diversos seguimentos de atuação, as pessoas estão sempre expostas à estatística, e isso se deve às infinitas aplicações que o método estatístico proporciona àqueles que dele necessitam.

Estatística descritiva

É um ramo da estatística que aplica varias técnicas para descrever e interpretar um conjunto de dados, sendo que a realização de qualquer trabalho de pesquisa se fundamenta no uso de técnicas e métodos na representação dos fenômenos da realidade.

Segundo Lakatos (2007), a pesquisa configura-se como um procedimento formal, que precisa de um respaldo científico. Assim sendo, pesquisar em qualquer área do conhecimento significa responder ou aproximar-se das mais variadas questões que fazem parte de nosso cotidiano, como no caso do âmbito educacional. Assim sendo, a pesquisa é uma ferramenta muito importante ao responder as inquietações relativas ao objeto em estudo, pois é mediante o uso da pesquisa que se consegue chegar situações até então desconhecidas. Com isso, os pesquisadores estão atentos às transformações existentes na sociedade. A pesquisa é necessária em todas as situações, devido ao fato de proporcionar aproximações contínuas à realidade social (BATANERO, 1994).

Distribuição de frequência

Para Gal (2002), a tabela de distribuição de frequência constitui o tipo de tabela mais importante da estatística descritiva, sendo que um dos objetivos da estatística descritiva é a redução na quantidade de dados com os quais devemos trabalhar, e isto pode ser feito através de uma modificação na forma de apresentação destes dados, uma maneira de organizar um conjunto de dados de forma reduzida, é representá-lo é por meio de uma tabela de distribuição de frequências, sendo que: Frequência absoluta F_i: é o número de vezes que uma variável aparece no conjunto de dados: frequência relativa f_r: é a divisão da frequência absoluta, pelo tamanho da amostra n. Podendo ser expressa em porcentagem.

A seguir segue uma tabela de distribuição de frequência com suas frequências absolutas e relativas, referente ao peso de 26 crianças em uma creche.

Tabela 1. Distribuição d frequência referente ao peso de 26 crianças em uma creche

	$Peso$	$X_i = (Li + Ls)/2$	F_i	F_{ac}	f_r	f_{rac}	$f_p = f_r * 100$	f_{pac}
1	$10 \vdash 12$	$(10 + 12) = 11$	6	6	$6/26 = 0{,}23$	0,23	$0{,}23 * 100 = 23$	23
2	$12 \vdash 14$	$(12 + 14) = 13$	8	14	$8/26 = 0{,}31$	0,54	$0{,}31 * 100 = 31$	54
3	$14 \vdash 16$	$(14 + 16) = 15$	6	20	$6/26 = 0{,}23$	0,77	$0{,}23 * 100 = 23$	77
4	$16 \vdash 18$	$(16 + 18) = 17$	4	24	$4/26 = 0{,}15$	0,92	$0{,}15 * 100 = 15$	92
5	$18 \vdash 20$	$(18 + 20) = 19$	2	26	$2/26 = 0{,}08$	1,00	$0{,}20 * 100 = 8$	100
$\sum$			26		1,00		100	

Esse tipo de tabela é extremamente importante dentro da estatística descritiva, pois nos permite conhecer varias características da população ou objeto de estudo. Esse tipo de tabela nos permite conhecer, por exemplo, quantas crianças têm peso menor ou igual a 16 kg, isso pode facilmente ser identificado na coluna F_{ac} que representa a soma das frequências anterior.

Medidas de posição

São chamadas de medidas de tendência central, pois representam os fenômenos pelos seus valores médios, em torno dos quais tendem a se concentrar os dados, são medidas denominadas de estatísticas por representar um conjunto de dados de forma condensada, são conhecidas também como medidas de localização, que nos fornece uma descrição resumida sobre o comportamento de um determinado fenômeno.

Abordaremos os aspectos mais importantes de seis medidas de posição, para dados isolados. Essas medidas são: (CHOPPIN, 2004).

$$\left.\begin{matrix} Média \ldots. \\ Mediana \\ Moda \ldots.. \end{matrix}\right\} São medidas que representa um conjunto de dados pelo seu valor médio.$$

$$\left.\begin{matrix} Quartil \ldots \\ Decil \ldots\ldots \\ Percentil \end{matrix}\right\} Separatrizes: separa um conjun de dados em partes percentuais iguais.$$

Média Aritmética

É o obtido pela soma de todos os valores do conjunto de dados dividido pelo número de dados. Embora seja a mais usada das medidas de posição, é extremamente influenciada por valores extremos. O modelo de cálculo da média amostral é dado por:

$$\acute{X} = \sum_{i=1}^{n} X_i \frac{1}{n}$$

Sendo:

$\sum_{i=1}^{n} X_i$ = Somatório dos valores de X_i; i = índice que varia de 1 a n elementos da amostra;

n = tamanho da amostra em estudo.

Quando os dados estiverem agrupados numa distribuição de frequência usaremos a média aritmética dos valores $X_1, X_2, \dots, X_n$ ponderados pelas respectivas frequências absolutas: $F_1, F_2, \dots, F_n$. Um dispositivo prático para esse cálculo é a composição da seguinte tabela, que representa a renda familiar em salários mínimos de 40 famílias. A distribuição das rendas dessas famílias pode ser vista na tabela 2.

	$Renda$	$X_i = (Li + Ls)/2$	F_i	$X_i F_i$
1	2 ⊢4	$(2 + 4)/2 = 3$	5	$3 * 5 = 15$
2	4 ⊢6	$(4 + 6)/2 = 5$	10	$5 * 10 = 50$
3	6 ⊢8	$(6 + 8)/2 = 7$	14	$7 * 14 = 98$
4	8 ⊢10	$(8 + 10)/2 = 9$	8	$9 * 8 = 72$
5	10 ⊢12	$(10 + 12)/2 = 11$	3	$11 * 3 = 33$
$\sum$			26	268

O salario médio das famílias pode ser determinado com o uso da seguinte formula:

$$\acute{X} = \frac{\sum_{i=1}^{n} X_i F_i * 1}{n}$$

$$\acute{X} = \frac{268 * 1}{26}$$

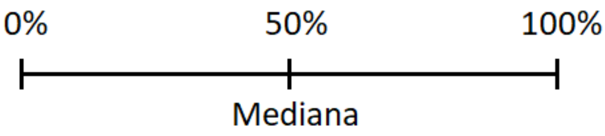
$$\acute{X} = \frac{268}{26}$$

$$\acute{X} = 6{,}7$$

Como a renda familiar foi dada em salários mínimos, pode-se dizer que a renda média desse grupo de 40 famílias é de 6,7 salários mínimos.

Mediana

Colocados em ordem crescente ou decrescente, mediana ($\tilde{X}$), é o valor central que divide a amostra ou população em duas partes iguais.



Mediana para uma variável discreta:

Uma variável será chamada discreta quando assumir qualquer valor pontual na reta real. Se n for impar, a mediano será o elemento central de ordem $\frac{n+1}{2}$. Caso n seja par, a mediana será a média entre os elementos centrais de ordem $\frac{n}{2}$ e $\frac{n}{2}+1$. Dada a distribuição a seguir, podemos determinar o valor mediano considerando uma amostra de tamanho $n = 11$

X_i	F_i	F_{ac}
1	1	1
2	3	4
3	5	9
4	2	11
$\sum$	11	

Como n é impar, a mediana será o elemento de ordem $\frac{n+1}{2}$, ou seja, $\frac{11+1}{2} = 6$ºelemento. Será, portanto o número 3, elemento facilmente identificado pela coluna F_{ac}, que acumula a frequência simples.

Mediana para uma variável continua

Uma variável será chamada continua quando assumir qualquer valor em certo intervalo na reta real. E pode ser determinado por:

$$\tilde{X} = l_{Md} + \frac{\left(\frac{n}{2} - \sum f\right) * h}{F_{Md}}$$

Sendo l_{Md} = Limite inferior da classe mediana; n = Tamanho da amostra; $\sum f$ = Soma das frequências anteriores à classe mediana; h = amplitude da classe mediana e F_{Md} = frequência da classe mediana.

Para o calculo da mediana para uma variável continua temos três passos a seguir:

1º Passo: Calcula-se $\frac{n}{2}$,como a variável é continua, não importa se n é par ou impar;

2º Passo: Pela F_{ac}, identifica a classe que contem a mediana e

3° Passo: Utiliza-se a formula imediatamente anterior.

Dada a distribuição a seguir, determine a mediana.

$Classes$	F_i	F_{ac}
35 ⊢45	5	5
45 ⊢55	12	17
55 ⊢65	**18**	35
65 ⊢75	14	49
75 ⊢85	6	55
85 ⊢95	3	58
$\sum$	58	

$$\tilde{X} = l_{Md} + \frac{\left(\frac{n}{2} - \sum f\right) * h}{F_{Md}}$$

$$\tilde{X} = 55 + \frac{\left(\frac{58}{2} - 17\right) * 10}{18}$$

$$\tilde{X} = 61{,}7$$

Logo o valor mediano será ($\tilde{X}$ = 61,7,que pertencer ao intervalo 55 ⊢65.

Moda

Dentre as principais medidas de posição, destaca-se a moda. É o valor que mais vezes ocorre numa distribuição. Para distribuição simples a identificação da moda pela simples observação do elemento que apresenta maior frequência. Para dados agrupados em classe utilizaremos a formula de Czuber.

$$Mo = l + \frac{\Delta 1}{\Delta 1 + \Delta 2} h$$

Sendo que: l = limite inferior da modal; $\Delta 1$ = diferença entre a frequência da classe modal e a anterior; $\Delta 2$ = diferença entre a frequência da classe modal e a posterior e

h = amplitude de classe. Dada a distribuição a seguir, determine a mediana.

$Classes$	
$0 \vdash 1$	3
$1 \vdash 2$	10
$2 \vdash 3$	**17**
$3 \vdash 4$	8
$4 \vdash 5$	5
$\sum$	43

$$Mo = l + \frac{\Delta 1}{\Delta 1 + \Delta 2} h$$

$$Mo = 2 + \frac{17 - 10}{(17 - 10) + (17 - 8)} 1$$

$$Mo = 2 + \frac{7}{7 + 9} 1$$

$$Mo = 2{,}44$$

Logo o valor modal será ($Mo = 2{,}44$,que deve pertencer ao intervalo $2 \vdash 3$.

Medidas de dispersão

São medidas estatísticas utilizadas para avaliar o grau de variabilidade ou dispersão dos valores em torna da média. Serve para medir representatividade da média.

Em trabalho envolvendo as medidas de posição, devemos tomar cuidado com o uso da média e sua interpretação, pois como é influenciada por valores extremos e poderá esconder aspectos métricos sobre o conjunto de dados. Por isso a média necessita de outras medidas estatísticas que auxiliem em seu uso e interpretação.

Sejam as series:

a) 20, 20, 20

b) 15, 10, 20, 25, 30

Tem-se: $\acute{X}_a = 20 e \acute{X}_b = 20$

Apesar das series terem medias iguais na série (a) não apresenta dispersão, enquanto que a série (b) apresenta dispersão. A assim a série (a) é mais representativa que a série (b). Segundo Gauthier (2008), se considerarmos apenas o número médio de cada serie, diríamos que as duas séries são iguais, porem, pela contagem individual em cada série verifica-se que há diferença de dispersão, e por isso precisamos medir o desvio de cada serie. Desde que se deseja medir a dispersão dos dados em relação à media, parece interessante a análise dos desvios em torna da média.

$$d_i = \left(X_i - \acute{X}\right)$$

Dada a série 3, 5, 7. Mostrar que a soma dos desvios é igual a zero. Sendo que a média vale 5.

$$\sum_{i=1}^{n}\left(X_i - \acute{X}\right) = (3-5)+(5-5)+(7-5) = (-2)+(0)+(2) = 0$$

Mas a soma dos desvios será sempre igual a zero. Uma maneira de se trabalhar com os desvios sem que a soma de zero é considerar o modulo dos desvios. Dessa forma define-se desvio médio como:

$$D_M = \frac{\sum d_i \vee F_i}{n} = \frac{\sum\left(X_i - \acute{X}\right)F_i}{n}$$

As principais medidas de dispersão são:

$$\left.\begin{matrix} Variância \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ DesvioPadrão \ldots\ldots\ldots\ldots.. \\ CoeficientedeVariação \end{matrix}\right\} Avaliamograudedisperssãodosvaloresemtornodamédia$$

Variância

Considera-se o quadrado de cada desvio $\left(X_i - \acute{X}\right)^2$, evitando que a soma dos desvios fosse zero.

$$\sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{n} X_i^2 F_i - \frac{\left(\sum X_i F_i\right)^2}{N}\right], \Big\} VariânciaPopulacional$$

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 F_i - \frac{(\sum X_i F_i)^2}{N}\right], \Big\} VariânciaAmostral$$

Onde: σ^2: Variância Populacional; S^2: Variância Amostral e $n-1$: Graus de liberdade, que é uma correção para o valor do cálculo na amostral.

X_i	F_i	$X_i F_i$	$X_i^2 F_i$
5	2	$5 * 2 = 10$	$5 * 10 = 50$
7	3	$7 * 3 = 21$	$7 * 21 = 147$
8	5	$8 * 5 = 40$	$8 * 40 = 320$
9	4	$9 * 4 = 36$	$9 * 36 = 324$
11	2	$11 * 2 = 22$	$11 * 22 = 242$
$\sum$	16	129	1083

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 F_i - \frac{(\sum X_i F_i)^2}{N}\right],$$

$$S^2 = \frac{1}{16-1}\left[1083 - \frac{129^2}{16}\right],$$

$$S^2 = 2{,}86,$$

Observando a formula para cálculo da variância $\left(X_i - \acute{X}\right)^2$, nota-se que é uma soma de quadrados. Dessa forma, se a unidade da variável for, por exemplo, metro (m) teremos como resultado metro ao quadrado (m^2). Para se ter a unidade original, necessita-se definir outra medida de dispersão, que é a raiz quadrada da variância. De todas as medidas de dispersão esta é a mais utilizada, pois ela exprime o resultado na mesma medida da variável em estudo.

Desvio padrão populacional	Desvio padrão amostral
$\sigma = \sqrt{\sigma^2}$	$S = \sqrt{S^2}$

Para o calculo do desvio padrão amostral, podemos extrair a raiz quadrada da variância, assim:

$$S = \sqrt{S^2}$$

$$S = \sqrt{(2{,}86)^2}$$

$$S = 1{,}69$$

Esse valor representa a medida de dispersão dos dados em torno de média amostral.

Coeficiente de variação

É uma medida que indica a proporção do desvio padrão em relação à média, expresso em percentagem, pode ser usada para comparar a dispersão de dois conjuntos de dados, sem que eles estejam necessariamente na mesma unidade de medida. Serve para medir a representatividade de um experimento. E é dado por:

Coeficiente de variação populacional	Coeficiente de variação amostral
$CV = \frac{\sigma}{\acute{X}} * 100$	$CV = \frac{S}{\acute{X}} * 100$

Sendo: σ: Desvio padrão populacional e S: Desvio padrão amostral.

O coeficiente de variação amostral referente ao exemplo anterior pode ser dado por:

$\acute{X} = \frac{\sum_{i=1}^{n} X_i F_i * 1}{n}$ $\acute{X} = \frac{129 * 1}{16}$ $\acute{X} = \frac{129}{16}$ $\acute{X} = 8{,}06$	$CV = \frac{S}{\acute{X}} * 100$ $CV = \frac{1{,}69}{8{,}06} * 100$ $CV = \frac{1{,}69}{8{,}06} * 100$ $CV = 21\%$

Este coeficiente $(CV = 21\%)$,é uma medida padronizada de dispersão de uma distribuição de probabilidade ou de uma distribuição de frequências. É definido como a razão do desvio padrão {}S pela média $\acute{X}$. Sendo usado em varias áreas do conhecimento para expressar a precisão de um experimento. O coeficiente de variação de 21% para fins estatísticos é considerado alto, indicando baixa precisão do experimento.

Teste de hipóteses

Trata-se de uma técnica de inferência estatística, ou seja, a partir de um teste de hipóteses, realizado com dados amostrais, pode-se inferir sobre a população. Podemos definir como hipóteses questões relacionadas a característica em estudo, sendo que o papel fundamental do teste de hipótese na pesquisa científica é sugerir explicações para os fatos. Num teste hipóteses são formuladas duas hipóteses; denominadas de hipótese nula (H_0) e hipótese alternativa (H_1), Hipótese nula é aquela que ser testada, e expressa por uma igualdade, enquanto que hipótese alternativa é dada por uma desigualdade e será considerada como aceitável, caso a hipótese nula seja rejeitada.

Tipos de erro

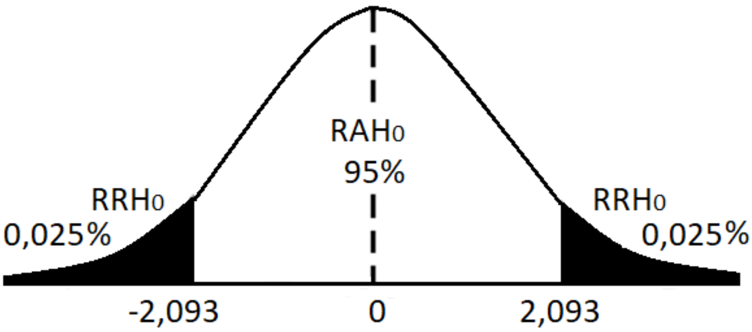
Há dois tipos possíveis de erro ao testar uma hipótese estatística. Pode rejeitar uma hipótese quando ela é, de fato, verdadeira, ou aceitar uma hipótese quando ela é, de fato, falsa. A rejeição de uma hipótese verdadeira é chamada de erro tipo I. A aceitação de uma hipótese falsa constitui o erro tipo II. As probabilidades desses dois tipos de erros são designadas, respectivamente, por α e β. A probabilidade α do erro tipo I é denominada nível de significância do teste. Os possíveis erros são apresentados na tabela 3.

Tabela 3. Tipos possíveis de erros ao testar uma hipótese estatística

		Realidade	
		H_0 Verdadeira	H_0 Falsa
Decisão	Aceitar H_0	Decisão correta $(1-\alpha)$	Erro tipo II (β)
	Rejeitar H_0	Erro tipo I (α)	Decisão correta $(1-\beta)$

Observa que o erro tipo I só poderá ser cometido se rejeitar H_0,sendo H_0 verdadeira, e o erro tipo II, quando se aceita H_0 sendo H_0 falsa. O tomador de decisão deseja, obviamente, reduzir ao máximo as probabilidades dos dois erros. Infelizmente, esta é uma tarefa difícil, porque, para uma amostra de determinado tamanho, a probabilidade de se incorrer em um erro do tipo II aumenta à medida que diminui a probabilidade do erro tipo I. E vice-versa. A redução simultânea dos erros poderá ser alcançada pelo aumento do tamanho da amostra. (BUSSAB 2003). Designa-se por H a hipótese nula que é a hipótese a ser testada, e por H a hipótese alternativa, que será aceita H_0 seja rejeitada. A hipótese nula será expressa por uma igualdade, enquanto que a hipótese alternativa por uma desigualdade. Para compreender o relacionamento dos erros e suas dimensões vamos idealizar um exemplo:

Um colégio atesta para calouros admitidos uma nota média 115 de um teste vocacional. Para testar a hipótese de que a média de uma nova turma será a mesma, tirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média de 118 e desvio padrão de 20. Admitir $\alpha = 0{,}05$, para efetuar o teste. Deseja-se testar $H_0\colon \mu = 115$contra $H_1\colon \mu \neq 115$. A área hachurada da figura abaixo corresponde à probabilidade rejeitar $H_0\colon \mu = 115.$ (RRH_0) Região de Rejeição de H_0. Sendo H_0 verdadeira, ou seja, a área representa α (probabilidade de cometer o erro tipo I). O valor 2,093 refere-se ao valor da tabela da distribuição t de Student, para 19 gruas de liberdade. Para o cálculo da estatística do teste usaremos a seguinte formula:



$$t_{cal} = \frac{\acute{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t_{cal} = \frac{118 - 115}{\frac{20}{\sqrt{20}}}$$

$$t_{cal} = 0{,}67$$

Como $-2{,}093 \leq t_{cal} \leq 2{,}093$, não se pode rejeitar $H_0\colon \mu = 115$ com esse nível de significância.

Intervalos de confiança

As medidas de posição média, mediana e moda são chamadas de estimativas pontuais, pois representam um único valor que estima características de uma amostra ou população em estudo. Existem ainda as estimativas por intervalos que são expressas por um intervalo fechado, em que acredita que contenha o verdadeiro valor do parâmetro. Por exemplo, num estudo para medir a altura media dos alunos de uma sala de aula, conclui-se que a altura media seja 1,60m, variado de 1,50m a 1,70m, temos neste caso um intervalo de confiança [1,50; 1,70].

Uma das finalidades pratica para o uso de intervalos é a determinação da dispersão ou variabilidade das estimativas, um intervalo muito grande indica que a estimativa não é tão precisa quanto outra com intervalo menor, ou seja, quanto maior a amplitude do intervalo menor a confiabilidade da estimativa. Intervalos de confiança podem ser construídos com diferentes coeficientes, sendo em geral mais utilizados 90%, 95% e 99%. A cada coeficiente corresponde um valor crítico da distribuição. A tabela 4 mostra valores dos coeficientes para o calculo para os principais níveis de confiança.

Tabela 4. Valores críticos para calculo de intervalos de confiança.

Confiança desejada	α	Valor critico $z_{\frac{\alpha}{2}}$
90%	0,10	1,645
95%	0,05	1,960
99%	0,01	2,575

Um intervalo de confiança para a média populacional quando a variância for conhecida, pode ser determinada pela seguinte formula:

$$P\left(\acute{x} - \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}} \leq \mu \leq \acute{x} + \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Sendo: P =Probabilidade; $z_{\frac{\alpha}{2}}$ =Valor critico da distribuição; $\sigma =$ Desvio padrão; n =Tamanho da amostra e $1 - \alpha$ =Nível de significância.

Com exemplo de uma aplicação pratica para o uso de intervalo de confiança, podemos citar: “A duração da vida de uma peça de equipamento é tal que $\sigma = 5$ horas. Foram amostradas 100 dessas peças obtendo uma média de 500 horas. Deseja-se construir um intervalo de confiança para a verdadeira duração média da peça com 90%, 95% e 99%.”

Intervalo de confiança para 90%

$$P\left(\acute{x} - \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}} \leq \mu \leq \acute{x} + \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(500 - \frac{1{,}645 * 5}{\sqrt{100}} \leq \mu \leq 500 + \frac{1{,}645 * 5}{\sqrt{100}}\right) = 90\%$$

$$P(499{,}18 \leq \mu \leq 500{,}82) = 90\%$$

$$[499{,}18; 500{,}82]$$

O intervalo $[499{,}18; 500{,}82]$ contém a duração média da peça com 90% de confiança, isso significa que se forem construídos intervalos dessa mesma maneira para um grande número de amostras, em 90% dos casos tais intervalos incluíram a média μ.

Intervalo de confiança para 95%

$$P\left(\acute{x} - \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}} \leq \mu \leq \acute{x} + \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(500 - \frac{1{,}96 * 5}{\sqrt{100}} \leq \mu \leq 500 + \frac{1{,}96 * 5}{\sqrt{100}}\right) = 95\%$$

$$P(499{,}02 \leq \mu \leq 500{,}98) = 95\%$$

$$[499{,}02; 500{,}98]$$

O intervalo $[499{,}02; 500{,}98]$ contém a duração média da peça com 95% de confiança, Intervalo de confiança para 95%

Intervalo de confiança para 99%

$$P\left(\acute{x} - \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}} \leq \mu \leq \acute{x} + \frac{z_{\frac{\alpha}{2}} * \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(500 - \frac{2{,}575 * 5}{\sqrt{100}} \leq \mu \leq 500 + \frac{2{,}575 * 5}{\sqrt{100}}\right) = 99\%$$

$$P(499{,}02 \leq \mu \leq 500{,}98) = 99\%$$

$$[498{,}71; 498{,}71]$$

O intervalo $[498{,}71; 500{,}98]$ contém a duração média da peça com 99% de confiança. Observa-se que quanto menor for o erro, maior será o intervalo.

CONSIDERAÇÕES FINAIS

Acreditamos que o conteúdo sobre estatística possa ser desenvolvido por alunos com intuito de mostrar a realidade da escola, e que a participação dos alunos poderá desenvolver o interesse pela disciplina, aguçando sua curiosidade em buscar novas informações até então desconhecidas. Procuramos mostrar neste trabalho a necessidade da interpretação de informações estatísticas que são vinculadas nos meios de comunicação, e também a importância da estatística na tomada de decisão.

REFERÊNCIAS

BATANERO, C.; GODINO, J. D.; GREEN, D.; HOLMES P. **Errors and difficulties in understanding elementary statistical concepts**. Internation Journal of Mathematics Education in Science and Technology, v. 25, n.4, p. 527-547, 1994.

BUSSAB, W. O. e MORETTIN, P. A**. Estatística Básica**. São Paulo: Editora Saraiva, 2003.

CHOPPIN, Allain. História dos livros e das edições didáticas: sobre o estado da arte. **Revista Brasileira de Educação**, São Paulo, v. 30, n. 3, p. 549-566, set./dez. 2004.

GAL, Iddo. **Adult´s statistical literacy: meanings, components, responsibilities – appeares.** Internacional Statistical Review, Espanha, v. 70, n. 1, p. 1-33, 2002.

GAUTHIER, Clermont **O ensino da estatística e da probabilidade na educação básica e a formação dos professores**. Campinas, v. 28, n. 74, p. 57-73, abr. 2008.

LAKATOS, Eva Maria; MARCONI, Marina Andrade. **Fundamentos de metodologia científica**. 6. ed.5.reimpr. São Paulo: Atlas 2007**.**

SHULMAN, Lee S. **Those who understand: knowledge growth in teaching. Educational Researcher,** Washington (EUA), v. 15, n. 2, p. 4-14, 1986.