INSTITUTO DE MATEMÁTICA PURA E APLICADA

JOÃO LUIS RODRIGUES FREIRE

TÍTULO: INTRODUÇÃO À INFERÊNCIA ESTATÍSTICA

Rio de Janeiro – RJ
2017

JOÃO LUIS RODRIGUES FREIRE

TÍTULO: INTRODUÇÃO À INFERÊNCIA ESTATÍSTICA

Trabalho de Conclusão de Curso apresentado ao Curso PROFMAT do Instituto de Matemática Pura e Aplicada, em cumprimento às exigências legais como requisito parcial à obtenção do título de Mestre.

Orientador: Paulo Cesar Pinto Carvalho, Doutor, Cornell Universtiy

Rio de Janeiro – RJ
2017

JOÃO LUIS RODRIGUES FREIRE

AGRADECIMENTO

Agradeço aos professores do IMPA que sempre se dispuseram a ajudar e tiveram a paciência necessária para sanar as mais prosaicas dúvidas.

Epígrafe

***“Todas as vitórias ocultam uma abdicação”.* (Simone de Beauvoir)**

RESUMO

No Brasil, o ensino da disciplina de estatística no Ensino Médio está circunscrito ao ensino da Estatística Descritiva, ou seja, parte da estatística que nos permite a representação, descrição e compreensão de dados de uma distribuição.

O ramo da Estatística chamada de Inferência Estatística, em contrapartida, permite ao estudante ter uma visão mais completa do fenômeno em estudo na medida em que podem ser feitas afirmações sobre ele a partir de uma amostra representativa desse mesmo fenômeno. Embora a Inferência Estatística tenha uma conexão natural com o estudo das Probabilidades, os dois assuntos continuam sendo estudados separadamente nas nossas escolas, o que prejudica, em muito, o bom entendimento pelos nossos estudantes dos conceitos embutidos na teoria estatística.

Este trabalho traz uma proposta de encaminhamento para o ensino de Inferência Estatística no Ensino Médio, apresentando os principais conceitos e propondo uma sequência didática para o tema.

Palavras-chave: Curva Normal. Teorema Central do Limite. Inferência Estatística.

ABSTRACT

In Brazil, the teaching of Statistics in our high schools is limited to the field of Descriptive Statistics which is the part of Statistics that allow us to represent, describe and understand the distribution of a set of data.

The field of Statistics called Statistical Inference, on the other hand, allows the student to have a broader view of the phenomenon under study due to assumptions which can be made based on a representative sample. Although Statistical Inference has a natural connection with the study of probabilities, both of the subjects are still being taught separately in our schools, which affect the understanding of the concepts of Statistics Theory by the students.

This work is a proposal of the Statistical Inference teaching in high schools, encompassing the main concepts of the Statistical Inference and a proposal for its didactic sequence.

Keywords: Bell Curve. Central Limit Theorem. Statistical Inference.

SUMÁRIO

1 INTRODUÇÃO

Os avanços no campo da probabilidade levaram ao desenvolvimento da teoria estatística permitindo que se façam generalizações científicas a partir de informações incompletas. Enquanto a estatística descritiva trata da coleta, organização e apresentação dos dados, a inferência estatística trata de generalizações da parte para o todo. Na estatística descritiva medidas como médias, medidas de dispersão etc. representam fins em si mesmos, ao passo que na inferência estatística são meios no processo de investigação.

A importância do trabalho reside no fato de que a Estatística ministrada nas escolas de Ensino Médio no Brasil além de não contemplar o estudo da inferência estatística tampouco a relaciona ao estudo da teoria das probabilidades.

Em alguns Países, a Estatística passou a fazer parte dos currículos nacionais no Ensino Fundamental nas décadas de 80 e 90. Nos Estados Unidos, de acordo com o National Council of Teachers of Mathematics (NCTM, 2016) os conteúdos principais de Estatística são apresentados durante os 12 anos (K-12) da educação básica. Assim, ao final desta escolarização, os estudantes são capazes de: formular questões para serem respondidas por meio de coleta de dados; registrar e organizar dados; escolher métodos estatísticos apropriados para análise dos dados; realizar previsões e inferências baseadas nos dados analisados; entender e aplicar conceitos básicos de probabilidade.

No Brasil, o conteúdo de Estatística foi inserido de forma incipiente nos Parâmetros Curriculares Nacionais (PCN1, 1997) do ensino fundamental em 1997, e do ensino médio em 1999 (PCN2, 1999) com a denominação de tratamento da informação. Nas discussões recentes a respeito da proposta de uma Base Nacional Comum Curricular (BNCC, 2016), referente ao ensino de Matemática, a Estatística como um dos cinco eixos da Matemática tem aparecido como a principal novidade a ser incluída desde os anos iniciais do Ensino Fundamental, até as séries finais do Ensino Médio.

Portanto, o objetivo desse trabalho consiste na preparação de sequências didáticas que explorem essa conexão e, assim, possibilitar uma maior compreensão por parte de nossos professores da necessidade de se abordar os dois assuntos conjuntamente.

Em resumo, o trabalho consiste basicamente no desenvolvimento da Curva de Gauss, sua relação com a teoria das probabilidades, e sua aplicação no universo da inferência estatística.

Esse trabalho foi dividido em cinco capítulos, Introdução, Conhecimentos Básicos, Desenvolvimento, Sequência Didática e Considerações Finais. O capítulo 2, Conhecimentos Básicos, foi dividido em três seções e nele fornecemos conhecimentos essenciais de estatística necessários ao desenvolvimento do trabalho. Na seção 2.1 discorremos sobre a Função Probabilidade e os conceitos de valor esperado e variância. Na seção 2.2 abordamos os ensaios de Bernoulli e a distribuição binomial. Na seção 2.3 o foco é a Curva Normal e suas propriedades.

O capítulo 3 trata da inferência estatística propriamente dita e divide-se em quatro seções. Na seção 3.1 descrevemos o que significa fazer uma inferência estatística a partir de uma amostra selecionada. Na seção 3.2 fizemos algumas considerações sobre a natureza de uma distribuição amostral, da distribuição amostral da média e suas implicações. Apresentamos, também, o Teorema Central do Limite, ferramenta essencial no estudo da Inferência Estatística. Na seção 3.3 apresentamos algumas técnicas de seleção de uma amostra, e na seção 3.4 expomos duas técnicas utilizadas na inferência estatística, a saber, estimação e teste de hipótese.

O capítulo 4 consiste de uma sequência didática para o ensino da Curva Normal e uma sequência didática para a introdução aos alunos do ensino médio do tema “Inferência Estatística”.

O capítulo 5 apresenta as considerações finais.

2 CONHECIMENTOS BÁSICOS

2.1 Função Probabilidade

Seja X uma variável aleatória, isto é, uma variável cujo valor depende do resultado de um experimento aleatório, ou seja, aquela cujo valor não pode ser plenamente controlado ou determinado antes da observação. A função probabilidade é a função que associa a cada possível ocorrência de uma variável aleatória discreta uma probabilidade.

Exemplo 1: Seja o lançamento de um dado não viciado. Existem somente seis possíveis valores e todos os resultados são igualmente prováveis, pois não existe razão alguma para atribuirmos probabilidades diferentes em um dado. Logo:

$P(X=1)=1/6$

$P(X=2)=1/6$

$P(X=3)=1/6$

$P(X=4)=1/6$

$P(X=5)=1/6$

$P(X=6)=1/6$

Exemplo 2: No lançamento de uma moeda não viciada 3 vezes, seja X o número de caras obtidas. Então,

$P(X=0)=1/8$

$P(X=1)=3/8$

$P(X=2)=3/8$

$P(X=3)=1/8$

Em ambos os casos é fácil entender o processo de geração da variável aleatória X e, portanto calcular a probabilidade de cada valor ocorrer. Em outras circunstâncias, não podemos calcular as probabilidades porque não sabemos sobre o processo suficientemente bem.

Assim, podemos dizer que o valor da função probabilidade ou função de massa de probabilidade para uma determinada variável é a probabilidade com que a variável aleatória assumirá um valor a. Sendo f a função probabilidade então $f(a)=P(X=a)$.

Função probabilidade no lançamento de um dado não viciado:

$f(1)=1/6$

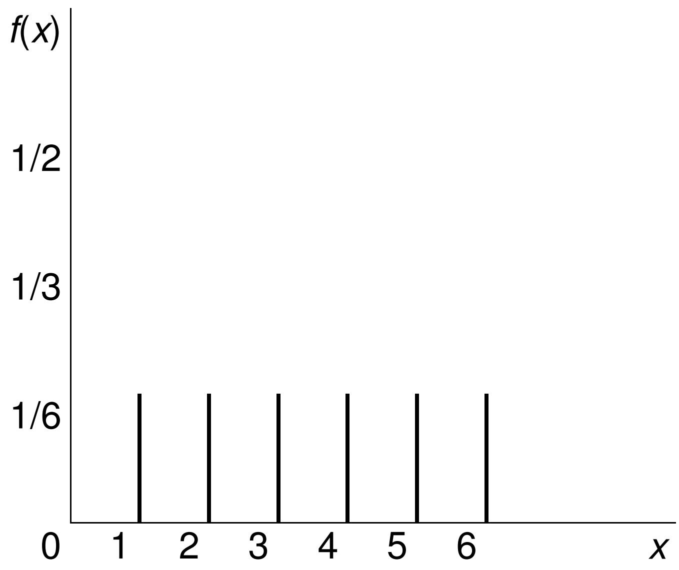
$f(2)=1/6$

$f(3)=1/6$

$f(4)=1/6$

$f(5)=1/6$

$f(6)=1/6$



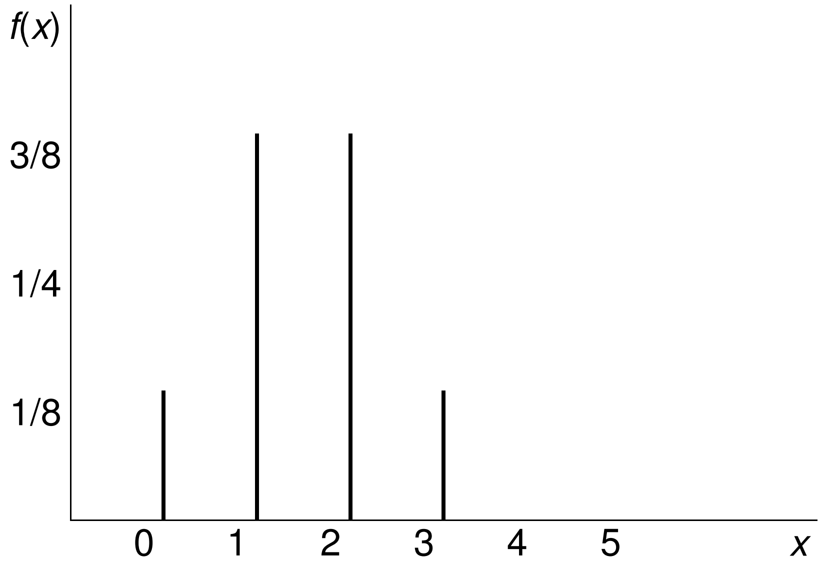
Função probabilidade do número de caras obtidas no lançamento de uma moeda 3 vezes:

$f(0)=1/8$

$f(1)=3/8$

$f(2)=3/8$

$f(3)=1/8$



Da função probabilidade seguem duas propriedades básicas:

$f(a)$ é menor ou igual a 1, para todos os valores possíveis de a

$f(a)$ é maior ou igual a 0, para todos os valores possíveis de a

Agora, suponhamos uma lista de todos os valores possíveis de uma variável aleatória e adicionemos suas probabilidades. Se o resultado dessa soma for menor que 1 ou maior que 1, algo está errado. A probabilidade deverá ser exatamente 1 pois X assumirá um de seus valores possíveis. Assim, para que uma função f seja uma função de probabilidade válida, a soma de seus valores possíveis deve ser 1. Se os valores possíveis são $x_1, x_2, x_3,...,x_n$, então:

$$f(x_1) + f(x_2) + f(x_3) + ... + f(x_n) = 1$$

Assim, se X representa o número de caras obtidas em 3 lançamentos temos:

$$1/8 + 3/8 + 3/8 + 1/8 = 1$$

Às vezes queremos saber qual é a probabilidade de que X seja menor ou igual a um determinado valor (x^*). Por exemplo, se você estiver jogando *blackjack* (conhecido também como 21) cujo objetivo é obter uma pontuação mais próxima possível de 21 sem ultrapassar essa barreira, e você já tirou 7 e 8, então você estará interessado na probabilidade de que o valor da sua próxima carta seja menor ou igual a 6.

Para acharmos a probabilidade de que uma variável aleatória X assuma um valor menor ou igual a 3 no lançamento de um dado não viciado ou honesto, precisamos adicionar $P(X=1)+P(X=2)+P(X=3)=1/2$. Em geral, se $x_1, x_2, x_3, \ldots, x^*$ são valores possíveis de Y menores ou iguais a x^* então:

$$P(Y \leq x^*)=f(x_1)+f(x_2)+f(x_3)+\ldots.+f(x^*)$$

A função que diz que a probabilidade de Y deva ser menor ou igual a um determinado valor é usualmente conhecida como função de distribuição acumulada, e a representaremos por F:

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

Calculando a função de distribuição acumulada para o número de ocorrências de cara no lançamento de uma moeda por 3 vezes temos:

$$F(x) = \begin{cases} 0 \; ; & x<0 \\ 1/8 \; ; & 0\leq x<1 \\ 1/2 \; ; & 1\leq x<2 \\ 7/8 \; ; & 2\leq x<3 \\ 1 \; ; & 3\leq x \end{cases}$$

Valor Esperado

Considere a variável aleatória X sendo o número de vezes que cara aparece no lançamento de três moedas não viciadas simultaneamente. Façamos agora o experimento X oito milhões de vezes. Quantas vezes teremos 0 cara, 1 cara, 2 caras ou 3 caras? É razoável supor que provavelmente o valor X=0 ocorrerá aproximadamente um milhão de vezes, o valor X=1 aproximadamente 3 milhões de vezes, o valor X=2 aproximadamente 3 milhões de vezes, e o valor X=3 um milhão de vezes. Na realidade, muito provavelmente não teremos esses números exatamente.

Representamos o valor esperado de X por $E(X)$ *e* sua fórmula geral é

$E(X) = f(x_1)x_1 + f(x_2)x_2 + f(x_3)x_3 + \ldots + f(x_n)x_n = \sum f(x_n)x_n$

Exemplo: número de caras obtidas no lançamento de três moedas não viciadas

$E(X) = 0\text{x}0{,}125 + 1\text{x}0{,}375 + 2\text{x}0{,}375 + 3\text{x}0{,}125 = 1{,}5$

Variância

Embora o valor esperado de uma variável aleatória nos diga muito sobre o seu comportamento, várias outras informações não são possíveis de serem observadas.

Nesse sentido, podemos utilizar a variância para medir a dispersão de uma lista de números assim como a dispersão de uma variável aleatória. Para calcularmos a variância, primeiramente devemos calcular o desvio quadrático de cada valor possível em relação ao valor esperado. Se cada valor possível possuir igual probabilidade de ocorrência, então a variância é simplesmente a média dos desvios quadráticos.

No caso das probabilidades diferirem é necessário multiplicar cada desvio quadrático pela probabilidade desse valor ocorrer, e somarmos os seus produtos.

Representando a variância de X por *VAR(X),* sua fórmula geral é dada por:

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

EXEMPLO:

Seja X o número de caras obtidas no lançamento de 3 moedas, então sua variância é:

Valor Possível (a_i)	Probabilidade (p)	Desvio Quadrático $(a_i - E(X))^2$	$p(a_i - E(X))^2$
0	0,125	$(0 - 1,5)^2$	9/32
1	0,375	$(1 - 1,5)^2$	3/32
2	0,375	$(2 - 1,5)^2$	3/32
3	0,125	$(3 - 1,5)^2$	9/32
		$Var\ (X) =$	0,75

Alternativamente, podemos calcular a variância da seguinte maneira:

Valor Possível (a_i)	Probabilidade (p)	$(a_\mathrm{i})^2$	$p(a_\mathrm{i})^2$
0	0,125	$(0)^2$	0
1	0,375	$(1)^2$	0,375
2	0,375	$(2)^2$	1,5
3	0,125	$(3)^2$	1,125
		$E(X^2) =$	3

Conforme vimos anteriormente, o valor esperado do número de caras no lançamento de 3 moedas não viciadas é de 1,5. Logo:

$$VAR(X) = E(X^2) - [E(X)]^2 = 3 - 2,25 = 0,75$$

2.2 Ensaios de Bernoulli

Imagine um experimento repetido inúmeras vezes. Seja p a probabilidade de o experimento ser realizado com sucesso. Um experimento como esse em que existem somente duas possibilidades, sucesso ou fracasso, é chamado de tentativas de Bernoulli. É um caso especial de distribuição binomial.

Seja Z uma variável aleatória na qual Z só pode assumir dois valores possíveis: 0 (fracasso) e 1 (sucesso), então:

$f(0)=P(Z=0)=1-p$
$f(1)=P(Z=1)=p$
$f(a)=0$, para todos os outros valores de a

O valor esperado pode ser calculado da seguinte forma:

$E(Z) = 0 \text{ x } P(Z=0) + 1 \text{ x } P(Z=1) = [0 \text{ x } (1-p)] + [1 \text{ x } p] = p$

A variância pode ser calculada por:

$E(Z^2) = 0^2 \text{ x } P(Z=0) + 1^2 \text{ x } P(Z=1) = [0^2\ (1-p)] + [1^2\ p] = p$

$Var(Z) = E(Z^2) - (E(Z))^2 = p - p2 = p(1-p)$

Note que a variância é maximizada quando $p=1/2$, ou seja, o grau de incerteza é o maior possível.

Distribuição Binomial

Seja um experimento em que existam somente 2 resultados possíveis: sucesso ou fracasso. Chamemos de p a probabilidade de sucesso de cada tentativa, e de $1-p$ a probabilidade de fracasso. Se o experimento for conduzido 10 vezes, quantas tentativas independentes resultarão em sucesso? Se a probabilidade de sucesso p for igual a 0,8, intuitivamente esperaremos que em média ocorreram 8 sucessos.

Se o experimento for conduzido 2 vezes, qual a probabilidade de ambas tentativas resultarem em sucesso? Seja A o evento sucesso na primeira tentativa e B o evento sucesso na segunda tentativa, então $P(A)=p$ e $P(B)=p$. O evento sucessos em ambas tentativas pode ser escrito como A∩B. Suponha que cada tentativa seja independente uma da outra. Isso significa que a chance de se obter sucesso em qualquer tentativa não é afetada pelos resultados de outras tentativas. Se as tentativas são independentes então podemos multiplicar as duas probabilidades:

$P(A \cap B) = P(A)P(B) = p^2$

Assim, a probabilidade de se obter sucessos em ambas tentativas é p^2. Podemos utilizar o mesmo raciocínio para mostrar que a probabilidade de se obter sucesso em todas 10 tentativas é de p^{10}. Por exemplo, se p=0,8, então a probabilidade de 10 sucessos é de $(0,8)^{10}$. Embora a chance de sucesso em qualquer tentativa seja boa, a chance de 10 sucessos ocorrerem em 10 tentativas é bastante pequena.

Da mesma maneira, a probabilidade de 10 tentativas resultarem em 10 fracassos é de $(1\text{-}p)^{10}$. Por exemplo, se p=0,8, a chance de 10 fracassos em 10 tentativas é de $(0,2)^{10}$.

Então, qual a probabilidade de exatamente 1 das 10 tentativas seja um sucesso?
Para achar a probabilidade de que exatamente 1 tentativa seja um sucesso, devemos somar a probabilidade de que somente a primeira tentativa seja um sucesso mais a probabilidade de que somente a segunda tentativa seja um sucesso e assim sucessivamente. Como temos 10 tentativas, a resposta ao segundo questionamento é dada por:

$P(\text{um sucesso}) = 10 \times 0{,}8 \times 0{,}2^9$

No caso de desejarmos calcular a probabilidade de somente as duas primeiras tentativas serem um sucesso o cálculo será dado por:

$0{,}8^2 \times 0{,}2^8$

Essa é a mesma probabilidade esperada para quaisquer 2 sucessos e 8 fracassos. O número de maneiras de escolher 2 posições entre as 10 possíveis para ocorrer sucesso é $C_{10,2.}$

Portanto, a probabilidade de obtermos exatamente 2 sucessos em 10 realizações da experiência aleatória é:

$C_{10,2} \times 0{,}8^2 \times 0{,}2^8$

Seguindo o mesmo raciocínio chegamos a fórmula geral para a distribuição binomial:

$P(X = i) = C_{n,i} \times p^i \times (1\text{-}p)^{n-i}$

2.3 Distribuição de probabilidade de uma variável aleatória discreta

A distribuição de probabilidade de uma variável aleatória discreta é uma lista mutuamente excludente de todos os possíveis valores assumidos por aquela variável aleatória, de modo que uma determinada probabilidade de ocorrência esteja associada a cada um desses valores.

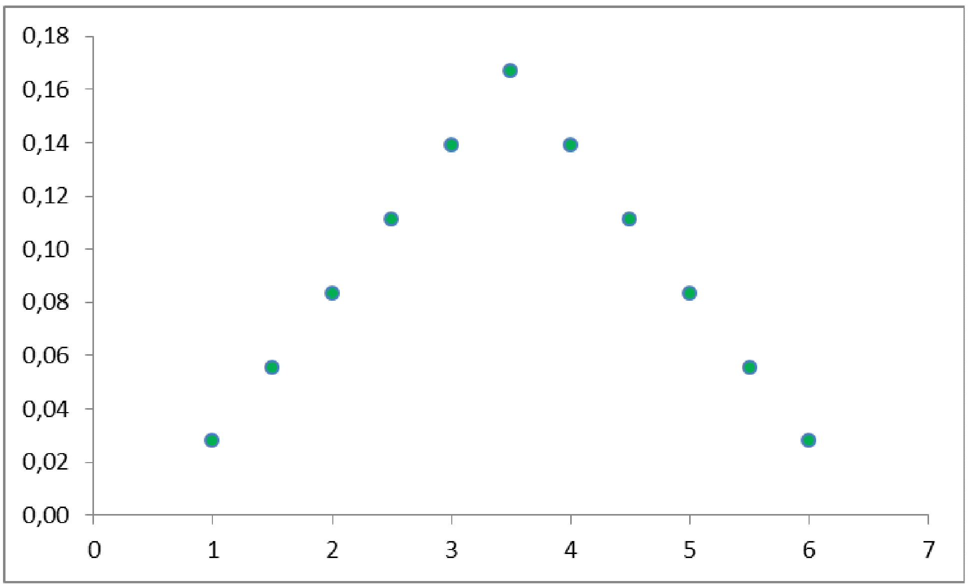
Já vimos anteriormente que o valor esperado de uma variável aleatória X é dado por $E(X) = \sum f(x_n)x_n$ Podemos verificar que $E(X)$ nada mais é que uma média ponderada dos diferentes valores de X com pesos dados pelas respectivas probabilidades.

Mostraremos a seguir dois exemplos clássicos de distribuição de probabilidade de uma variável aleatória discreta, a saber: lançamento de dados e lançamento de moedas.

Exemplo 1: Voltemos ao experimento de lançamento de dados

A distribuição de probabilidade da variável resultante do lançamento de um dado segue a distribuição uniforme, ou seja, qualquer valor (1,2,3,4,5,6) tem a mesma probabilidade (1/6) de ocorrer.
No entanto, se ao invés de lançar um dado, sejam lançados dois dados e calculada a média, a média dos dois dados seguirá a seguinte distribuição:

2 Dados



1º DADO	2º DADO	SOMA	MEDIA
1	1	2	1,0
1	2	3	1,5
2	1	3	1,5
1	3	4	2,0
2	2	4	2,0
3	1	4	2,0
1	4	5	2,5
2	3	5	2,5
3	2	5	2,5
4	1	5	2,5
1	5	6	3,0
2	4	6	3,0
3	3	6	3,0
4	2	6	3,0
5	1	6	3,0
1	6	7	3,5
2	5	7	3,5
3	4	7	3,5

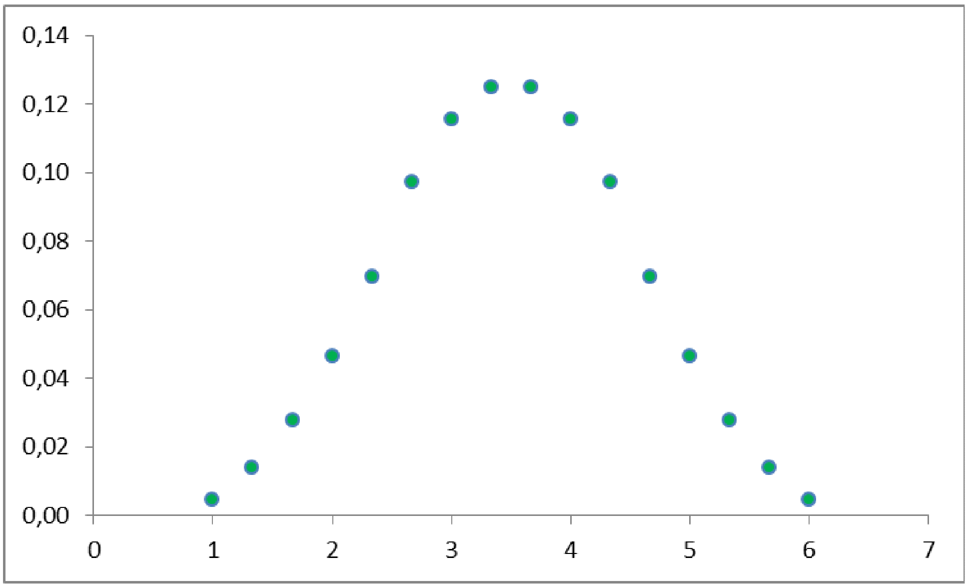
1º DADO	2º DADO	SOMA	MEDIA
4	3	7	3,5
5	2	7	3,5
6	1	7	3,5
2	6	8	4,0
3	5	8	4,0
4	4	8	4,0
5	3	8	4,0
6	2	8	4,0
3	6	9	4,5
4	5	9	4,5
5	4	9	4,5
6	3	9	4,5
4	6	10	5,0
5	5	10	5,0
6	4	10	5,0
5	6	11	5,5
6	5	11	5,5
6	6	12	6,0

Tabela de freqüência da média dos dois dados

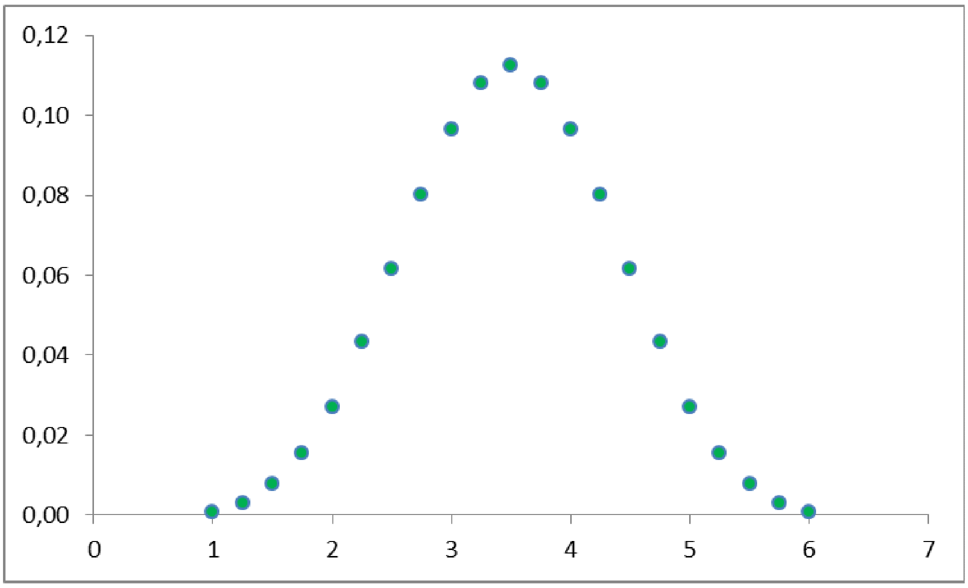
MÉDIA DE DOIS DADOS	FREQUÊNCIA
1,0	1
1,5	2
2,0	3
2,5	4
3,0	5
3,5	6
4,0	5
4,5	4
5,0	3
5,5	2
6,0	1

Repetindo o mesmo experimento com aumentos sucessivos no número de dados teremos as seguintes configurações abaixo:

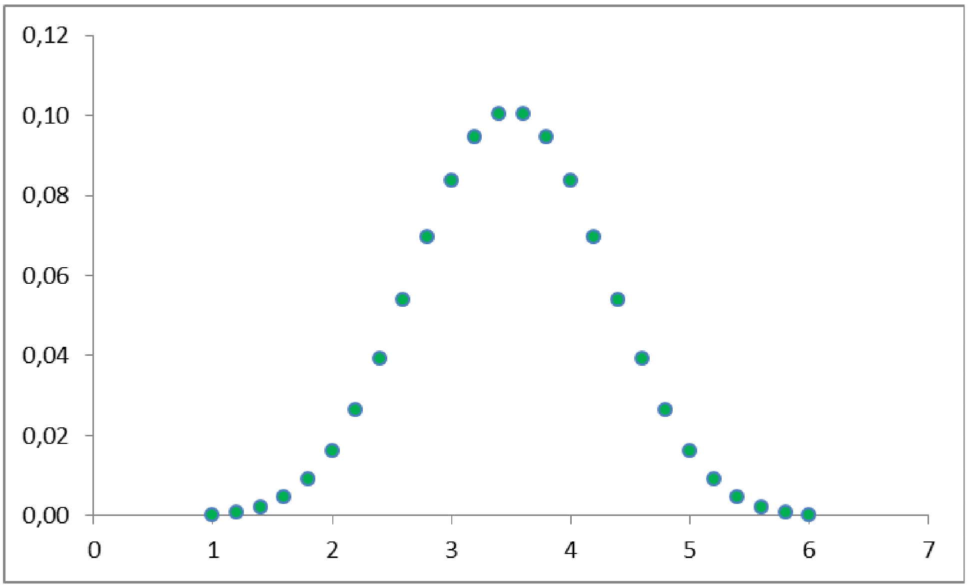
3 Dados



4 Dados



5 Dados

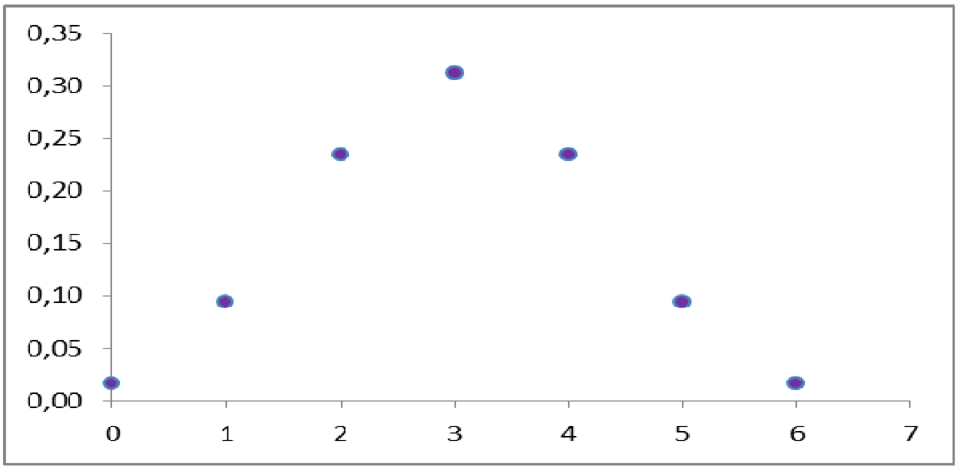


Exemplo 2: número de caras obtidas no lançamento de *n* moedas

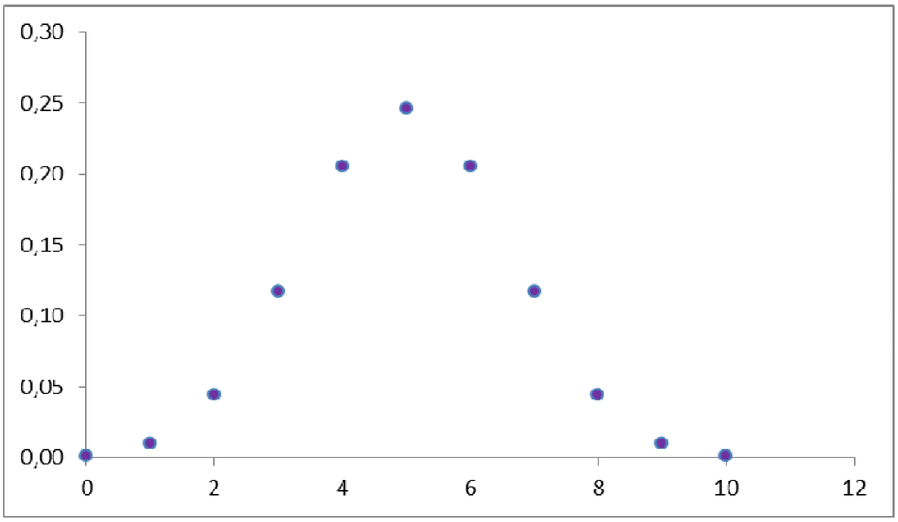
A tabela abaixo mostra o número de caras obtidas ao lançarmos a moeda 6 vezes:

Nº DE CARAS	FREQUÊNCIA
0	1
1	6
2	15
3	20
4	15
5	6
6	1

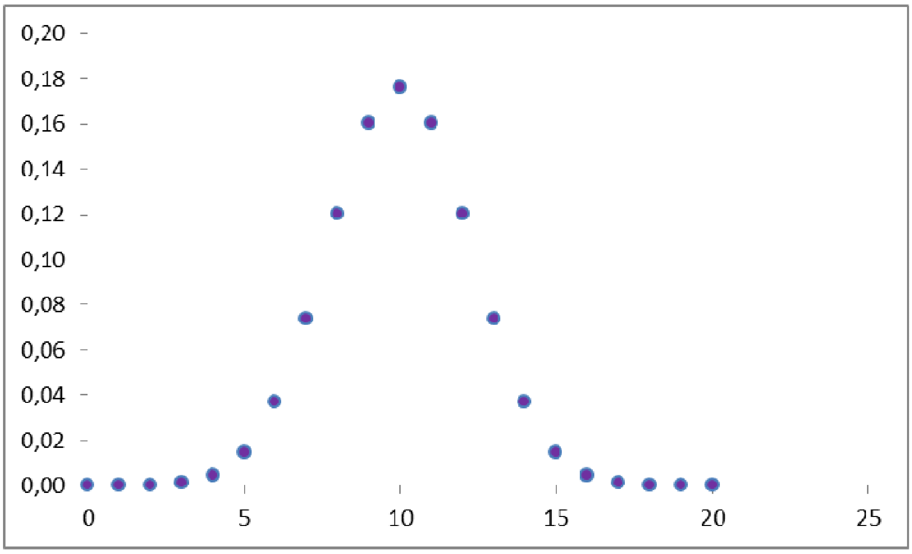
***n*=6**



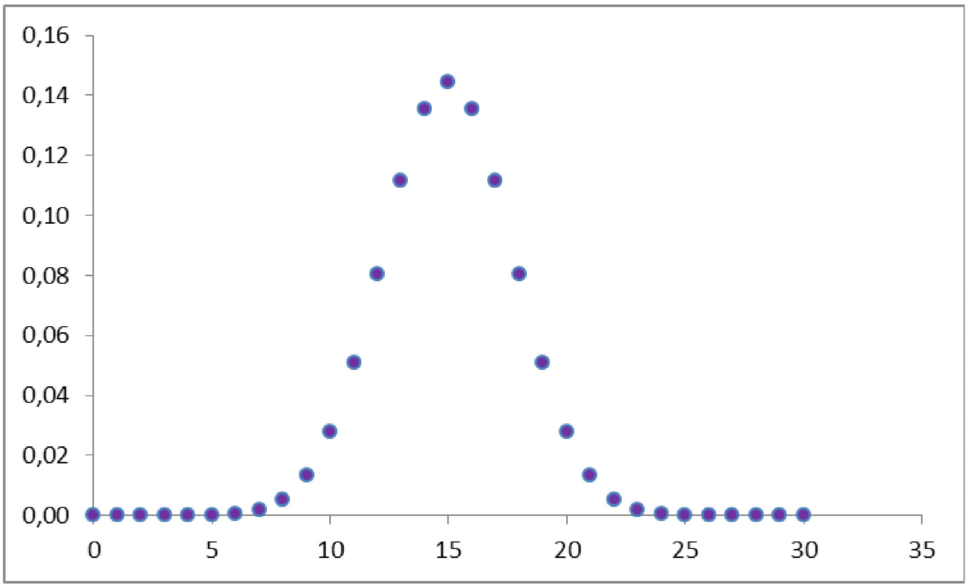
n=10


0,30
0,25
0,20
0,15
0,10
0,05
0,00
0
2
4
6
8
10
12

n=20


0,20
0,18
0,16
0,14
0,12
0,10
0,08
0,06
0,04
0,02
0,00
0
5
10
15
20
25

***n*=30**



Distribuição Normal

NOTAÇÃO: $X \sim N\,(\mu,\ \sigma^2)$, sendo X variável aleatória com média μ e variância σ^2

A curva normal tem uma história bastante longa e está ligada à história da descoberta das probabilidades em matemática, no século XVII, que surgiram para resolver inicialmente questões de apostas em jogos de azar.

O responsável mais direto pela curva normal foi Abraham de Moivre, matemático francês exilado na Inglaterra, que a definiu em 1730, dando seqüência aos trabalhos de Jacob Bernoulli (teorema ou lei dos grandes números) e de seu sobrinho Nicolaus Bernoulli, matemáticos suíços. Moivre notou que quando o número de eventos (ex: jogar uma moeda para o alto, e observar se é cara ou coroa) aumenta, o formato da distribuição binomial se aproxima cada vez mais de um determinado tipo de curva. Então, caso ele pudesse encontrar uma expressão matemática para essa curva, ele seria capaz de resolver problemas (jogos de azar) mais facilmente. Moivre publicou seus trabalhos em 1733 na obra *The doctrine of the chances*. O sucesso da descoberta foi rápido e grandes nomes passaram a trabalhar sobre a curva normal, inclusive Gauss. Daí a curva normal ser também conhecida como distribuição de Gauss.

Uma das primeiras aplicações da distribuição normal foi na análise de erros de medição realizados em observações astronômicas. Esses erros ocorriam devido à imprecisão dos instrumentos e imperfeição dos observadores. Galileu, no século XVII, notou que esses erros eram simétricos e que pequenos erros ocorriam com mais freqüência do que grandes erros. Porém, somente no início do século XIX foi descoberto que esses erros possuíam uma distribuição normal.

A mesma distribuição foi descoberta por Laplace em 1778 quando ele obteve o importante teorema central do limite.

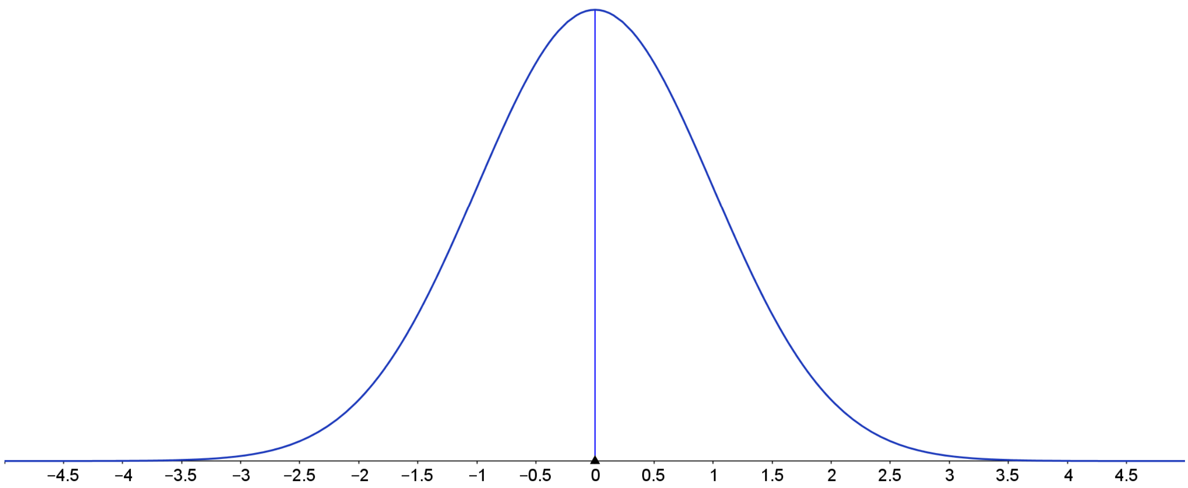
A importância da curva normal deriva basicamente do fato de que a distribuição de muitos fenômenos naturais é aproximadamente uma distribuição normal. Coube a Quételet ser o primeiro a utilizar a distribuição normal nas características humanas. Ele notou que características como altura, peso, força e QI (quociente de inteligência) de uma população eram normalmente distribuídas.

A Função Densidade da Distribuição Normal é dada pela fórmula

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

com $-\infty < x < \infty$, $-\infty < \mu < \infty$ e $\sigma > 0$

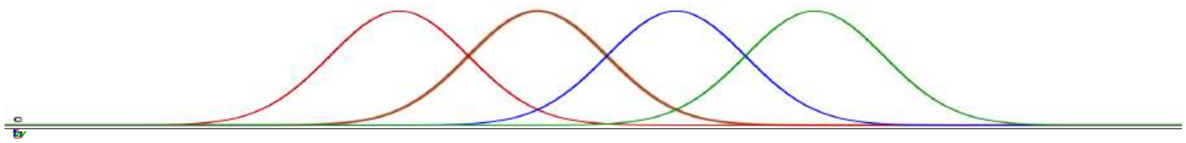
e sua representação gráfica é dada na forma de sino:



Para $\mu = 0$ e $\sigma = 1$

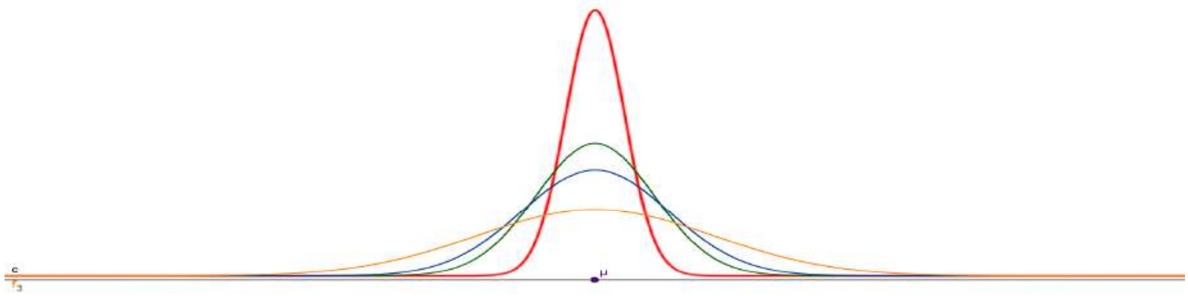
A figura 1 mostra 4 distribuições normais que possuem a mesma variância, porém médias diferentes.

Figura 1



Já a figura 2 mostra 4 distribuições normais que possuem a mesma média, porém variâncias diferentes.

Figura 2



A área abaixo de cada curva é 1, pois estamos tratando de uma distribuição de probabilidade contínua. As caudas da distribuição nunca tocam o eixo dos x. Assim, a variável aleatória normal poderá assumir qualquer valor, incluindo valores distantes da média. Entretanto, podemos verificar que a área nas caudas é pequena, significando que a chance de um valor distante da média ser observado é remota.

Propriedades da distribuição normal:

A distribuição normal fica determinada especificando dois parâmetros: a média e a variância. A média está localizada no pico da distribuição. A variância determina a dispersão da distribuição – se a distribuição irá se alongar ou se a maior parte da área se concentrará perto do pico.

1) A distribuição é simétrica em torno da média. O que significa que a média, a mediana e a moda são iguais. A média divide a área sob a curva normal em duas partes iguais.

2) O domínio da distribuição estende-se de $-\infty$ a $+\infty$. Portanto, a distribuição é ilimitada.

3) A altura máxima da curva normal é atingida no ponto $x=\mu$, e os pontos de inflexão, isto é, os pontos em que a distribuição torna-se mais achatada, ou em que a curva muda de concavidade, em $x=\mu \pm \sigma$. Logo, o desvio-padrão mede a distância do centro da distribuição ao ponto de inflexão.

Examinando os gráficos gerados anteriormente a partir do número de caras obtidas no lançamento de n moedas, verificamos que, de fato, a distribuição binomial tende a se aproximar da distribuição normal, à medida que n (tamanho da amostra) aumenta.

Sejam:

n = número de tentativas independentes (Bernoulli)
p = probabilidade de sucesso em cada tentativa
q = probabilidade de insucesso em cada tentativa

Se $np < 10$ e $nq > 10$, a distribuição normal é uma boa aproximação da distribuição binomial, com $\mu = np$ e $\sigma^2 = npq$

Distribuição Normal Padrão

NOTAÇÃO: $X \sim N\,(0,1)$, sendo X variável aleatória com média 0 e variância 1

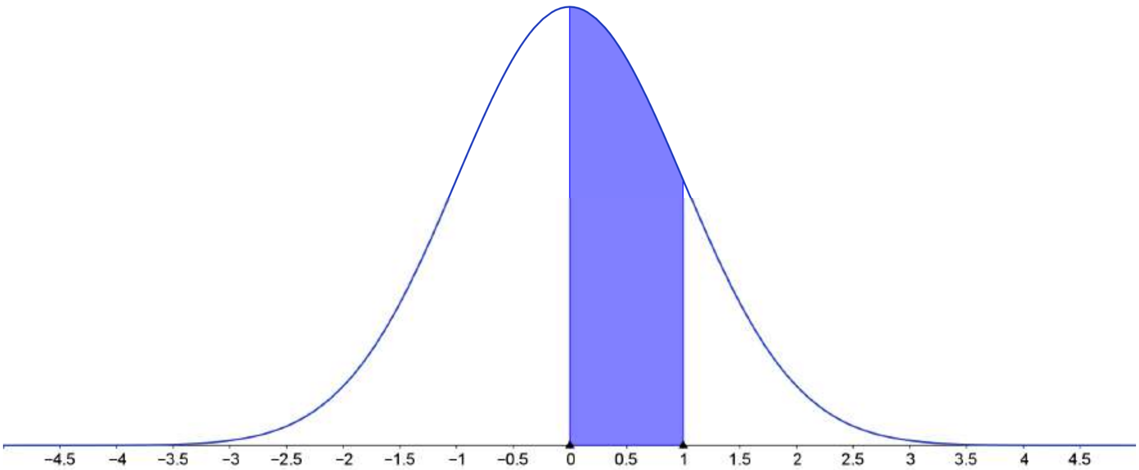
Em muitos casos para sabermos a área abaixo da curva normal é necessário consultarmos uma tabela, ou usar um computador ou calculadora. Seria virtualmente impossível prepararmos uma tabela para cada distribuição normal com diferentes médias e variâncias. Felizmente, podemos encontrar a área abaixo da curva para qualquer distribuição normal comparando-a com a distribuição normal cuja média seja 0 e variância igual a 1. Essa distribuição normal é chamada de distribuição normal padrão.

Seja Z uma variável aleatória com uma função de densidade normal padrão.

Se desejarmos calcular a probabilidade de Z estar entre 0 e 1 é necessário calcular a área abaixo da curva entre 0 e 1.

Utilizando a tabela da Curva Normal Padrão (APÊNDICE A),

$P\,(0 < z < 1) = 0{,}3413$



Devemos destacar que há duas concepções erradas amplamente difundidas entre os estudantes:

1) O valor de Z deve estar entre -3 e +3; deve-se ao fato dos estudantes utilizarem uma tabela de valores que possui essa amplitude de variação (Apêndice A).

2) Não há limite para os valores máximo e mínimo de Z; alguns estudantes entendem que as extremidades da curva normal são assintóticas ao eixo das abscissas.

Veremos agora um importante teorema que diz que se uma variável X tem distribuição normal, uma função linear de X será também normal. A prova deste teorema pode ser encontrada em muitos livros de estatística matemática.

X(normal) $\rightarrow aX + b$ (normal) com a, b constantes

Se encontrarmos a, b tais que a média de $aX + b$ seja zero e sua variância unitária, teremos uma variável normal padrão.

Seja o sistema de equações dado por:

$E(aX + b) = 0$ e $Var(aX + b) = 1$

Sabendo que $Var(aX+b) = E[(aX + b)^2] - (E[aX + b])^2$

$= E[a^2X^2 + 2abX + b^2] - (aE(X) + b)^2$

$= a^2E(X^2) + 2abE(X) + b^2 - a^2E^2(X) - 2abE(X) - b^2$

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

Logo,

$a\mu + b = 0$ e $a^2\sigma^2 = 1$

Resolvendo o sistema para a e b obteremos

$a = 1/\sigma$ e $b = -\mu/\sigma$

Assim temos $aX + b = (X - \mu)/\sigma$

Como $(X - \mu)/\sigma$ tem média zero e variância igual a um, é uma variável normal padrão, isto é,

$(X - \mu)/\sigma = Z$

Assim qualquer variável normal com média μ e variância σ^2 pode ser transformada em uma variável normal padrão.

Vamos considerar $P(x_1 < x < x_2)$, com $x_1 < x_2$.

$(X - \mu)/\sigma = Z$ implica que $X = \sigma Z + \mu$

Portanto,

$x_1 = \sigma z_1 + \mu$ e $x_2 = \sigma z_2 + \mu$

Após as substituições,

$P(x_1 < x < x_2) = P(\sigma z_1 + \mu < \sigma z + \mu < \sigma z_2 + \mu)$

Cancelando os termos positivos comuns do segundo membro da equação temos:

$P(x_1 < x < x_2) = P(z_1 < z < z_2)$

Onde,

$z_1 = (x_1 - \mu)/\sigma$ e $z_2 = (x_2 - \mu)/\sigma$

Assim, a probabilidade de x cair entre x_1 e x_2 é igual à probabilidade de a variável normal padrão cair entre $(x_1 - \mu)/\sigma$ e $(x_2 - \mu)/\sigma$.

EXEMPLO:

Consideremos uma variável X normalmente distribuída que tenha média 5 e desvio-padrão 2. O problema é achar a probabilidade $P(2 < x < 3)$. Para fazer isto temos de achar uma proposição de probabilidade equivalente em termos da variável normal padrão Z. Aqui o limite inferior $x_1 = 2$ e o limite superior $x_2 = 3$. Como $z_1 = (x_1 - \mu)/\sigma$

Então, $z_1 = (2-5)/2 = -3/2$

Do mesmo modo, $z_2 = (x_2 - \mu)/\sigma = (3-5)/2 = -1$

Portanto, $P(2 < x < 3) = P\ (-3/2 < z < -1)$

Utilizando a tabela da Curva Normal Padrão (APÊNDICE A),

$P\ (-3/2 < z < -1)$ é equivalente a $P\ (0 < z < 3/2) – P\ (0 < z < 1)$

Logo, a $P\ (2 < x < 3) = P\ (0 < z < 3/2) – P\ (0 < z < 1) = 0{,}4332 – 0{,}3413 = 0{,}0919$

Função de Distribuição Acumulada

Uma função de distribuição acumulada possui as seguintes propriedades:

1) $F(a)$ se encontra entre 0 e 1.
2) Quanto maior a, $F(a)$ se aproxima de 1.
3) Quanto menor a, $F(a)$ se aproxima de 0.
4) $F(a)$ é não decrescente.

Assim, se desejarmos calcular a probabilidade de X ser maior que um determinado valor de a, podemos usar a fórmula:

$P(X > a) = 1 – P(X < a) = 1 – F(a)$

É importante destacar que a fórmula só se aplica para variáveis aleatórias contínuas.

3 DESENVOLVIMENTO

3.1 Inferência Estatística

A Inferência Estatística consiste de procedimentos para fazer generalizações sobre as características de uma população a partir da informação contida na amostra. No processo de busca pelo conhecimento de certas características quantitativas ou qualitativas de uma determinada população são feitos estudos ou levantamentos onde são realizadas medições ou contagens provenientes de todos os elementos dessa população - levantamento censitário - ou de um subconjunto ou amostra dela - levantamento amostral.

Em princípio, o senso comum nos diz que o levantamento censitário é mais preciso que um levantamento amostral. Entretanto, isso nem sempre é verdade, pois erros ocorrem devido ao despreparo da equipe de coleta de dados etc.

Normalmente, utilizamos o levantamento amostral e, a partir de dados fornecidos por essas amostras fazemos afirmações probabilísticas sobre a população em estudo. Destacamos que o processo de seleção de uma amostra exerce um papel primordial na determinação da composição dessa amostra. Há situações, porém, em que a amostra se constitui por mera observação.

Nas populações infinitas, a amostra será randômica se cada observação for independente da outra, isto é, a ocorrência de um evento não influencia de modo algum na ocorrência do outro.

As características numéricas de uma população chamam-se parâmetros enquanto as características de uma amostra chamam-se estatísticas. Tais características podem ser medidas de tendência central (por exemplo: média ou moda), dispersão (por exemplo: desvio-padrão) ou, no caso de fenômenos qualitativos, a proporção de observações de determinada espécie. Naturalmente, os parâmetros de uma população infinita nunca são observados. Os parâmetros de uma população finita poderiam ser observados, mas por questões como custo e tempo de processamento tornam-se inviáveis de serem observados na prática.

Assim, se a população for infinita, nunca poderá ser observada como um todo, e qualquer julgamento sobre ela só poderá vir da amostra. Mas mesmo no caso da população ser finita pode haver boa razão para observarmos apenas a amostra, como no caso de se querer medir a duração média de lâmpadas.

O objetivo da inferência estatística é, portanto, ajuizar sobre parâmetros populacionais com base em estatísticas amostrais. Na verdade, são tentativas de adivinhação revestidas de certo grau de segurança e que podem ser de dois tipos: ou se relacionam à estimação de parâmetros ou ao teste de alguma hipótese sobre o parâmetro de interesse.

A estimação é feita com o auxílio de um estimador, ou seja, de uma fórmula que descreve o modo de calcularmos o valor de determinado parâmetro populacional. O valor específico de um estimador chama-se estimativa. Juízos na forma de testes de hipótese consideram a priori certos pressupostos sobre o valor de um parâmetro. Se a informação amostral fornecer evidência contra a hipótese, nós a rejeitamos. Caso contrário, a hipótese será aceita.

É evidente que diferentes amostras conduzem a diferentes avaliações. Algumas estarão mais perto da verdade, ou seja, do valor verdadeiro do parâmetro, do que outras. Na prática, geralmente só temos uma amostra e, portanto, somente uma avaliação.

Convém ressaltar, que a redução no erro de amostragem é substancial para aumentos iniciais no tamanho da amostra, tornando-se, porém, desprezível após determinado tamanho da amostra.

Em síntese, a variabilidade amostral indica que nem todas as amostras são iguais. Já a representatividade amostral sugere que a média amostral terá características similares a da população se o método de amostragem utilizado for adequado. O equilíbrio entre essas duas ideias aparentemente antagônicas é primordial no estudo da Inferência Estatística. Nota-se que os estudantes em geral dão ênfase excessiva na representatividade da amostra e esquecem a sua variabilidade e o efeito do tamanho da amostra sobre a mesma.

Destacamos que, no cálculo da variância de uma amostra de tamanho n, devemos utilizar um fator de correção da fórmula de variância conhecido como Fator de Correção de Bessel

(APÊNDICE B). Assim, ao invés de dividirmos o desvio quadrático de cada valor possível em relação ao valor esperado por n (supondo que cada valor possível possui igual probabilidade de ocorrência), faremos a divisão por n-1.

3.2 Distribuições Amostrais

Na seção anterior vimos que a maneira de sabermos o grau de fidedignidade de uma avaliação é conhecer o comportamento de todas as avaliações que poderiam ser feitas com base em todas as amostras possíveis. Podemos retirar uma amostra após a outra, procurando saber qual o valor da avaliação de cada amostra e arranjar tais cálculos em forma de distribuição. Se tivermos um número infinito de tais amostras, a distribuição resultante será chamada de distribuição amostral. Consideremos, por exemplo, o problema de estimar a renda média familiar do Brasil em determinado mês, com base em uma amostra de 100 famílias. Uma possibilidade seria calcular a renda média familiar de nossa amostra e usá-la como estimativa da média populacional.

Naturalmente, poderíamos usar a moda ou a mediana ou alguma outra medida como nosso estimador. Suponhamos que estimamos a média populacional usando a média amostral. Desejamos então saber o grau de segurança deste estimador. Um modo de descobrir isto seria retirar um número infinito de tais amostras, calculando o valor da média amostral de cada amostra e dispondo estes valores em forma de distribuição. Observe que, embora a população de todas as famílias do Brasil seja finita, o número de amostras que podemos retirar desta população é infinito, enquanto permitirmos que cada família seja incluída em cada uma das amostras. Tal amostragem chama-se amostragem com repetição.

Se cada família contida na amostra for selecionada ao acaso, não saberemos de antemão qual será sua renda. Neste caso a renda familiar será uma variável randômica. Além disso, a renda média observada numa amostra é também uma variável randômica. Isso significa que a distribuição amostral de uma média de amostras, considerando um número infinito de amostras, é realmente uma distribuição de probabilidade.

Tal distribuição pode ser discreta ou contínua, dependendo da natureza da variável populacional. Em nosso exemplo, a distribuição amostral é contínua visto que a renda é uma variável contínua.

Naturalmente, amostras de diversos tamanhos nos darão informações diversas sobre a população de onde são retiradas. Portanto, os estimadores que utilizam amostras de diferentes tamanhos ostentarão diferentes graus de confiabilidade. É importante destacar que qualquer distribuição amostral se refere sempre a amostras do mesmo tamanho.

Quando se usa uma amostra da população sempre haverá uma probabilidade de estar cometendo um erro (justamente por ser usada uma amostra), chamado erro amostral: a diferença entre os métodos estatísticos e os outros reside no fato de que os métodos estatísticos permitem calcular essa probabilidade de erro. E para que isso seja possível a amostra da população precisa ser aleatória.

Distribuição Amostral da Média

Suponha uma população identificada pela variável aleatória Y, cujos parâmetros média populacional $\mu=E(Y)$ e variância $\sigma^2 = Var(Y)$ são supostamente conhecidos. Vamos retirar todas as amostras possíveis de tamanho n dessa população e para cada uma delas, calcular a média $\bar{Y}$.

Vamos supor a seguinte população {2,3,4,5} com média μ=3,5 e variância σ^2 = 1,25.

Vamos relacionar todas as amostras possíveis de tamanho 2, com reposição, desta população. Da análise combinatória verificamos 16 possíveis combinações: 16 = 4x4

(2,2)	(2,3)	(2,4)	(2,5)
(3,2)	(3,3)	(3,4)	(3,5)
(4,2)	(4,3)	(4,4)	(4,5)
(5,2)	(5,3)	(5,4)	(5,5)

Agora, vamos calcular a média de cada amostra. Teremos:

2,0	2,5	3,0	3,5
2,5	3,0	3,5	4,0
3,0	3,5	4,0	4,5
3,5	4,0	4,5	5,0

Por fim, vamos calcular a média das médias, ou seja,

$E(\bar{Y})$ = (2,0 + 2,5 + ...+ 5,0)/16 = 3,5 = μ

Agora, vamos calcular a variância:

$VAR\ (\bar{Y})$ = [(2,0 - 3,5)² + (2,5 - 3,5)² + ... + (5,0 - 3,5)²]/16

$VAR(\bar{Y})$ = 0,625

Agora vamos relacionar todas as amostras possíveis de tamanho 3, com reposição, desta população. Nesse caso, existem 64 possíveis combinações: 64 = 4x4x4

(2,2,2)	(2,2,3)	(2,2,4)	(2,2,5)	(2,3,2)	(2,3,3)	(2,3,4)	(2,3,5)
(2,4,2)	(2,4,3)	(2,4,4)	(2,4,5)	(2,5,2)	(2,5,3)	(2,5,4)	(2,5,5)
(3,2,2)	(3,2,3)	(3,2,4)	(3,2,5)	(3,3,2)	(3,3,3)	(3,3,4)	(3,3,5)
(3,4,2)	(3,4,3)	(3,4,4)	(3,4,5)	(3,5,2)	(3,5,3)	(3,5,4)	(3,5,5)
(4,2,2)	(4,2,3)	(4,2,4)	(4,2,5)	(4,3,2)	(4,3,3)	(4,3,4)	(4,3,5)
(4,4,2)	(4,4,3)	(4,4,4)	(4,4,5)	(4,5,2)	(4,5,3)	(4,5,4)	(4,5,5)
(5,2,2)	(5,2,3)	(5,2,4)	(5,2,5)	(5,3,2)	(5,3,3)	(5,3,4)	(5,3,5)
(5,4,2)	(5,4,3)	(5,4,4)	(5,4,5)	(5,5,2)	(5,5,3)	(5,5,4)	(5,5,5)

Similarmente, calculando a média de cada amostra e a média das médias obteremos:

$E(\bar{Y})$ = 3,5

e

$VAR(\bar{Y}) = 0{,}417$

Seja $VAR(\bar{Y}) = VAR(Y)/n$, em que n é o tamanho das amostras retiradas da população:

Para n=2 temos:

$VAR(\bar{Y}) = (1{,}25)/2 = 0{,}625$

Para n=3 temos:

$VAR(\bar{Y}) = (1{,}25)/3 = 0{,}417$

Podemos verificar que, a distribuição das médias destas amostras tende para uma distribuição com média μ (igual à média da população) e com desvio padrão $\sigma/\sqrt{n}$.

Teorema Central do Limite

O Teorema Central do Limite afirma que quando o tamanho da amostra aumenta a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal. Este resultado é fundamental na teoria da inferência estatística, e sua utilidade vai desde estimar os parâmetros como a média populacional ou o desvio padrão da média populacional a partir de uma amostra aleatória dessa população, ou seja, da média amostral e do desvio padrão da média amostral, até calcular a probabilidade de um parâmetro ocorrer dado um intervalo, sua média amostral e o desvio padrão da média amostral.

O uso generalizado da distribuição normal no universo estatístico deve-se ao Teorema Central do Limite. Sua prova pode ser encontrada em muitos textos de estatística matemática, mas não a apresentaremos aqui por não fazer parte do escopo de nosso trabalho. Esse teorema apresenta três propriedades básicas:

a) A média da distribuição amostral é igual à média da população, e igual à média de uma amostra quando o tamanho da amostra tende ao infinito (Lei dos Grandes Números).

b) A forma da distribuição amostral tende a assumir a forma de sino à medida que se aumenta o tamanho da amostra, e aproximadamente normal, independente da forma da distribuição da população.

c) A forma da distribuição amostral cresce em altura e decresce em dispersão à medida que o tamanho da amostra cresce.

Seja uma população com média μ e desvio padrão σ, e selecionemos várias amostras de tamanho n. Para cada uma dessas amostras é possível calcular a respectiva média.

Pelo Teorema Central do Limite, a distribuição das médias destas amostras tende para uma distribuição normal com média μ (igual à média da população) e com desvio padrão $\sigma/\sqrt{n}$.

Ao desvio padrão da distribuição das médias amostrais dá-se o nome de erro padrão da média.

Se o erro padrão for pequeno as amostras com médias semelhantes à média da população são mais frequentes e assim é mais provável que a amostra que obtivemos seja uma dessas amostras. O erro padrão pode ser controlado com o tamanho da amostra, pois quanto maior for o tamanho da amostra menor será o erro padrão.

Obs: para amostras com tamanho $n>30$ a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal.

Na verdade, já havíamos mostrado anteriormente a aplicação desse teorema (sem mencioná-lo explicitamente) quando discorremos sobre a distribuição de probabilidade no lançamento de n dados. Verificamos, também, sua validade no experimento de lançamento de n moedas se ao invés de classificarmos como sucesso ou insucesso a obtenção de cara ou coroa respectivamente atribuirmos o valor 0 para insucesso e 1 para sucesso.

Exemplo:

As notas num certo exame padronizado têm media 500 e desvio-padrão 60. Uma nota acima de 530 é considerada muito boa. Uma pessoa consegue entrar em uma Universidade de prestígio se ela obtém acima de 530 neste exame. Numa certa sala onde o exame foi aplicado, 36 pessoas fizeram o teste. A nota média destas pessoas foi de 540. É razoável haver algum tipo de investigação para tentar detectar se houve alguma fraude no certame?

Solução: Seja *Y* a nota no teste. Pelo enunciado do problema, *Y* tem media 500 e desvio-padrão 60. Logo, a média amostral das notas das 36 pessoas daquela sala (supondo que as notas são independentes umas das outras e identicamente distribuídas) é uma variável com media 500 e variância $(60)^2/36$. Então, pelo Teorema Central do Limite:

$$\frac{\bar{Y} - 500}{\sqrt{\frac{(60)^2}{36}}} = \frac{\bar{Y} - 500}{\frac{60}{6}} = \frac{\bar{Y} - 500}{10}$$

é aproximadamente N (0,1)

$$P(\bar{Y} > 540) = P\left(\frac{\bar{Y} - 500}{10} > \frac{540 - 500}{10}\right) \cong 0$$

Logo, é absolutamente improvável que a nota media daquelas 36 pessoas tenha sido de 540, um indício claro de fraude no teste e que, portanto, precisará ser investigado.

3.3 Amostragem

Amostragem é o processo de obtenção de amostras. Ele é considerado parte fundamental no estudo da inferência estatística. Nessa parte do trabalho apresentaremos algumas noções sobre esse assunto visto que a Teoria das Amostragens constitui hoje um campo bastante desenvolvido e amplo da Estatística.

Um processo de amostragem diz-se enviesado quando tende sistematicamente a selecionar elementos de alguns segmentos da população, e a não selecionar sistematicamente elementos de outros segmentos da população.

Surge assim a necessidade de fazer um planejamento da amostragem que consiste, entre outras coisas, estabelecer quais elementos da população deverão compor a amostra assim como o método de seleção desses elementos.

De um modo geral, o trabalho do Estatístico deve começar antes de os dados serem recolhidos. Nesse sentido, o planejamento de um estudo estatístico, que começa com a forma de selecionar a amostra, deve ser feito de forma a evitar amostras enviesadas.

A seguir apresentamos exemplos de amostras enviesadas e como elas afetam o resultado da sua aplicação:

.

Amostra 1 – Utilização de alguns alunos de uma turma para tirar conclusões sobre o aproveitamento de todos os alunos da escola.

Resultado – Poderíamos concluir que o aproveitamento dos alunos é pior ou melhor do que na realidade é. As turmas de uma escola não são todas homogêneas, pelo que a amostra não é representativa dos alunos da escola. Poderia servir para tirar conclusões sobre a população constituída pelos alunos da turma.

Amostra 2 – Utilização dos jogadores de uma equipa de basquete de uma determinada escola para avaliar as alturas dos alunos dessa escola.

Resultado – O estudo concluiria que os estudantes são mais altos do que na realidade são.

Os exemplos que apresentamos anteriormente são exemplos de amostras enviesadas porque tiveram a intervenção do fator humano. Veremos logo adiante que, mesmo uma amostra aleatória pode não ser representativa da população em estudo.

A amostragem divide-se em métodos probabilísticos e métodos não probabilísticos.

Os métodos não probabilísticos de seleção de amostras não podem ser medidos objetivamente em contraposição aos métodos probabilísticos. Uma amostra é dita probabilística se a sua seleção é feita de maneira que cada elemento da população tem probabilidade conhecida de ser selecionado. No presente trabalho apresentaremos os principais métodos probabilísticos de seleção de amostras:

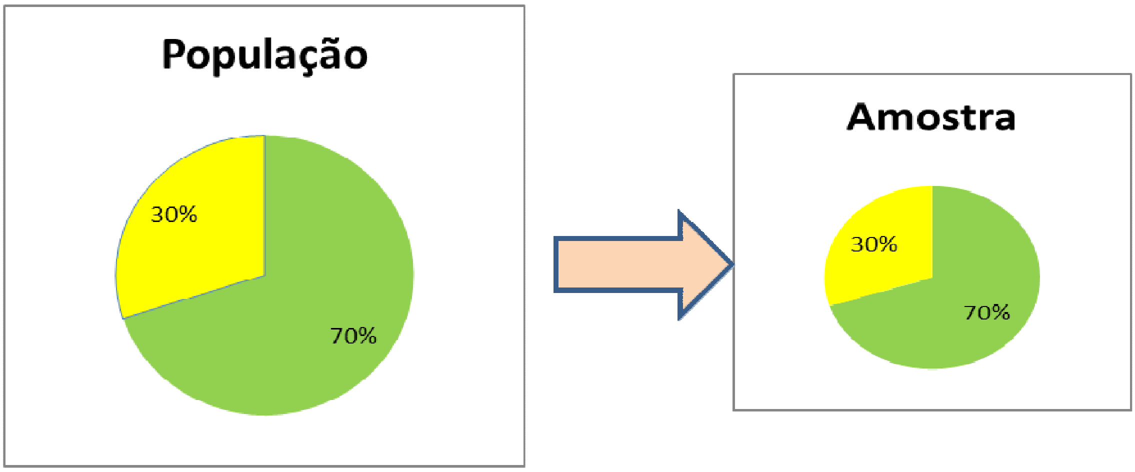
1) Amostragem aleatória simples: todos os elementos da população têm a mesma chance de serem selecionados. Este tipo de amostragem probabilística somente é recomendável se a população for homogênea em relação à variável de interesse.

   Exemplo: Queremos realizar uma pesquisa de opinião sobre a qualidade de um curso universitário, que tem cerca de 1000 alunos, perguntando aspectos relativos ao encadeamento das disciplinas no currículo.

   No presente caso estamos interessados na opinião dos alunos sobre o currículo. É razoável imaginar que um aluno do quarto ano tenha um conhecimento diferente do currículo do que outro do segundo ano, podendo, portanto, acarretar em diferentes opiniões e comprometer o resultado da pesquisa. Assim, como não há homogeneidade na população acerca da variável de interesse, a amostragem aleatória simples não é apropriada para este caso.

2) Amostragem estratificada: consiste em dividir a população em grupos (chamados estratos) por um processo denominado estratificação e de cada grupo selecionar uma amostra aleatória simples. A divisão da população em grupos tem por finalidade juntar num mesmo grupo elementos da população mais homogêneos entre si quanto à característica em estudo do que na população como um todo. Após a determinação dos estratos, seleciona-se uma amostra aleatória simples de cada estrato. Divide-se em:

   a) Amostragem Estratificada Proporcional: A proporcionalidade do tamanho de cada estrato da população é mantida na amostra. Exemplo: Se um estrato abrange 30% da população, ele também deve abranger 30% da amostra.



Exemplo:

Em uma população de 1000 pessoas, há 600 homens e 400 mulheres. Extraia uma amostra representativa, de 10%, dessa população.

Nesse exemplo, há uma característica que permite identificar 2 subconjuntos, a característica sexo. Considerando essa divisão, a amostra será constituída da seguinte forma:

SEXO	POPULAÇÃO	AMOSTRA (10%)
Masculino	600	60
Feminino	400	40
Total	1000	100

Portanto, a amostra deve conter 60 pessoas do sexo masculino e 40 do sexo feminino, totalizando 100 pessoas, que correspondem a 10% da população.

Para selecionar os elementos da população para formar a amostra, podemos executar os seguintes passos:

1º) Numerar as pessoas de 1 a 1000, sendo os homens numerados de 1 a 600 e as mulheres, de 601 a 1000;

2º) Escrever os números de 1 a 600 em pedaços de papel e colocá-los em uma urna A;

3º) Escrever os números de 601 a 1000 em pedaços de papel e colocá-los em uma urna B;

4º) Retirar 60 pedaços de papel, um a um, da urna A, e 40 da urna B, formando a amostra da população.

São exemplos desta técnica de amostragem as pesquisas eleitorais por região, cidades pequenas e grandes, área urbana e área rural, sexo, faixa etária, faixa de renda, etc. Mais adiante, quando falarmos sobre pesquisa eleitoral, destacaremos a importância da existência de um cadastro.

b) Amostragem Estratificada Uniforme: Selecionamos o mesmo número de elementos em cada estrato. É o processo usual quando se deseja comparar os diversos extratos.

3) Amostragem sistemática: quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente até compor o total da amostra, sendo o primeiro elemento sorteado aleatoriamente. Nesse tipo de amostragem a população deverá ser homogênea em relação à variável de interesse.

Exemplo: Uma operadora telefônica pretende saber a opinião de seus assinantes comerciais sobre seus serviços em determinada cidade. Supondo que há 25037 assinantes comerciais, e a amostra precisa ter no mínimo 800 elementos, mostre como seria organizada uma amostragem sistemática para selecionar os respondentes.

A operadora dispõe de uma lista ordenada alfabeticamente com todos os seus assinantes, o intervalo de retirada será:

$k = N/n = 25037/800 = 31{,}2965$

Como o valor de k é fracionário algo precisa ser feito. Aumentar o tamanho da amostra não resolverá o problema, porque 25037 é um número primo. Como não podemos reduzir o tamanho de amostra, devendo permanecer igual a 800, se excluirmos por sorteio 237 elementos da população, e refizermos a lista teremos:

k = N/n = 24800/800 = 31

A cada 31 assinantes, um é retirado para fazer parte da amostra. Sorteando como ponto de partida um número de 1 a 31 (do 1º ao 31º assinante), e tendo como sorteado o número 5, então a amostra será da forma:

{5, 36, 67, 98, ..., 24774}

3.4 Estimação

No processo de estimação, diferentemente do que ocorre no teste de hipóteses, não fazemos asserções apriorísticas cuja credibilidade deva ser disputada. No processo de estimação fazemos perguntas sobre o valor de um parâmetro particular, e mediante o uso de um estimador calculamos uma estimativa para aquele parâmetro.

A estimação pode ser dividida em duas partes, estimação por pontos e estimação por intervalos. Na estimação por ponto o objetivo é usar a informação amostral e apriorística para se calcular um valor que seria, em certo sentido, nossa melhor avaliação quanto ao valor de fato do parâmetro em questão. Na estimativa por intervalo usa-se a mesma informação com o propósito de se produzir um intervalo que contenha o valor verdadeiro do parâmetro com algum nível de probabilidade. Como um intervalo está plenamente caracterizado por seus limites, a estimação de um intervalo equivale à estimação de seus limites.

Para que o estimador atenda os objetivos do estatístico/pesquisador, ou seja, para que se alcance boas estimativas de um determinado parâmetro populacional, ele deverá apresentar as seguintes propriedades:

1. Consistência

Consistência é uma propriedade por meio da qual a acurácia de uma estimativa aumenta quando o tamanho da amostra aumenta.

2. Não tendenciosidade

Em uma particular amostra, o valor calculado pelo estimador pode desviar para mais ou para menos do valor do parâmetro, mas espera-se que, em média, ele determine o verdadeiro valor do parâmetro populacional. Não tendenciosidade é uma propriedade que assegura que, em média, o estimador é adequado.

3. Erro quadrático médio

É um conceito relacionado ao conceito de variância. A diferença entre a variância de um estimador e o erro quadrático médio é que, enquanto a variância mede a dispersão da distribuição em torno da media, o erro quadrático médio mede a dispersão em torno do verdadeiro valor do parâmetro. Quanto menor o seu valor, melhor é o estimador.

Inicialmente faremos algumas considerações a respeito da aplicabilidade da estatística em um processo eleitoral.

A ideia básica da estimação baseia-se na suposição de que se a nossa amostra é representativa da população, então a proporção de indivíduos na amostra que possuem a mesma intenção de voto em um determinado candidato deverá nos fornecer uma estimativa razoável do percentual de indivíduos da população em geral que pretendem votar nesse candidato. E para quantificarmos esse resultado utilizamos a teoria da probabilidade.

Os institutos de pesquisa no Brasil não fazem pesquisas probabilísticas puras, sem cotas, como é comum nos Estados Unidos. O objetivo das cotas é garantir a representatividade do universo que se pretende estudar. No caso das pesquisas eleitorais, a amostra deve refletir a distribuição do eleitorado segundo dados atualizados pelo TSE. O tempo necessário para aplicar o método de forma correta no Brasil não permitiria acompanhar o caráter dinâmico do processo eleitoral. Além disso, os altos custos envolvidos nas pesquisas domiciliares com método probabilístico puro inviabilizam a técnica no país. Nos Estados Unidos, a maioria dos estudos eleitorais são feitos por telefone , com controle do histórico de comportamento do eleitorado (se o entrevistado foi votar em eleições anteriores, por exemplo). Mas mesmo lá pesquisas feitas por telefone têm sido questionadas, já que uma parcela significativa dos eleitores está trocando as linhas fixas por celulares.

Uma dúvida bastante comum entre as pessoas é com relação aos dados divulgados pela imprensa. Quando algum instituto de pesquisa informa que o candidato A possui 44% das intenções de voto e que a margem de erro é de 2,0 pontos percentuais para mais ou para menos com intervalo de confiança de 95% isso significa que, em 100 amostras, 95 delas contêm o verdadeiro valor. Invariavelmente os institutos de pesquisa omitem o intervalo de confiança. Geralmente o intervalo adotado é de 95%.

Vejamos, então, a aplicação do método de estimação para calcularmos o percentual provável de indivíduos que irão votar em um determinado candidato em um processo eleitoral:

Primeiramente, escolhemos um indivíduo aleatoriamente. Uma vez escolhido não queremos escolher o mesmo indivíduo novamente. Supondo que a população supera em muito qualquer número de indivíduos que possamos escolher para nossa pesquisa, isso significa que o percentual de indivíduos que compartilham de uma mesma opinião não se alterará, ao removermos esse indivíduo da população. Isso acontece ao removermos o segundo, o terceiro, o quarto ou o quinto indivíduo, e daí sucessivamente. Portanto, ao selecionarmos n indivíduos aleatoriamente da população estaremos reproduzindo na verdade o experimento de Bernoulli n vezes.

Ainda não sabemos a probabilidade de sucesso em uma tentativa. Porém, se escolhermos um número significativo de indivíduos o modelo probabilístico se aproximará da distribuição normal. Como o experimento é aleatório, tudo pode acontecer. Mas esperamos que o número de sucessos em nossa amostra esteja próximo à media do experimento, e no intervalo de 1 ou 2 desvios-padrão da media. Portanto, o resultado esperado deverá ficar em algum ponto próximo da parte central de nossa curva de distribuição normal.

Destacamos que, a proporção de eleitores da amostra que pretendem votar no candidato A certamente será diferente se selecionarmos outra amostra ao acaso. Utilizando a distribuição binomial e o fato de que para uma amostra suficientemente grande a distribuição de p^* é aproximadamente normal com média igual a p e variância dada por $p(1\text{-}p)/n$, podemos mostrar que a probabilidade de que o intervalo abaixo contenha o verdadeiro valor de p é aproximadamente igual a 95%.

$$p^* \pm 1{,}96\sqrt{\frac{p(1-p)}{n}}$$

Demonstração:

Sejam:

p– proporção de eleitores que pretendem votar no candidato A
n– número de eleitores (tamanho da amostra)

p^* é variável, dependendo da amostra colhida

Com o auxílio da distribuição binomial mostraremos que a média de p^* é igual a p e que a variância é dada por $p(1\text{-}p)/n$

Sabemos como calcular o valor esperado e a variância de uma variável aleatória binomial (pag.16). Seja X o número de sucessos e p a proporção de sucessos em n tentativas, então $p=X/n$. Podemos encontrar $E(p)$ e $Var(p)$:

$E(p)=E(X/n)= E(X)/n= np/n=p$

e

$Var(p)=Var(X/n)=Var(X)/n^2 = np(1\text{-}p)/n^2 = p(1\text{-}p)/n$

Como a aferição desse intervalo depende do valor desconhecido de p, se substituirmos $p(1\text{-}p)$ pelo seu valor máximo o problema estará resolvido. O produto $p(1\text{-}p)$ atinge seu máximo quando p= 0,5. Portanto, pode-se afirmar que a probabilidade de que o intervalo supracitado contenha o verdadeiro valor de p é no mínimo de 95%.

Assim, se desejarmos uma margem de confiança de 95% e uma margem de erro de 2,0 pontos percentuais (para mais ou para menos) o tamanho de nossa amostra *n* deverá satisfazer:

$$\frac{1{,}96}{2\sqrt{n}} = \frac{2{,}0}{100}$$

E, portanto, *n* deverá ser igual a 2401 eleitores.

Em nossa pesquisa eleitoral entrevistaremos 2401 pessoas aptas a votar. Suponhamos que o número de pessoas que responderam que iriam votar no candidato A seja de 1675 pessoas de 2401 pesquisadas. Se a amostra for realmente aleatória ela deverá refletir a população como um todo.

Assumindo que a estimativa obtida em nossa amostra seja próxima à probabilidade p_0 (nº de eleitores na população que apoiam o candidato A) então $p_0 \approx 0{,}698$.

Utilizando as fórmulas abaixo:

$$\mu = np_0$$

$$\sigma^2 = np_0(1 - p_0)$$

$$\sigma = \sqrt{np_0\,(1 - p_0)}$$

Obtemos,

$$m = ns = 2401\,.\,\frac{1675}{2401} = 1675$$

$$d^2 = ns(1 - s) = 2401\,.(0{,}698).(1 - 0{,}698) = 506{,}12$$

$$d = \sqrt{ns(1 - s)} = \sqrt{506{,}12} = 22{,}50$$

Logo, há 95% de probabilidade da media μ se encontrar entre $m - 1{,}96d$ e $m + 1{,}96d$.

Ou seja,

$$1.630{,}90 \leq \mu \leq 1.719{,}10$$

Como μ = np_0 então,

$$1.630{,}90 \leq 2.401p_0 \leq 1.790{,}10 \rightarrow 0{,}67926 \leq \mu \leq 0{,}71599$$

Concluímos, portanto, que o candidato A tem aproximadamente 70% de intenção de voto com margem de erro de 2% para mais ou para menos.

Assim, se o tamanho da amostra permanecer fixo, um aumento da precisão implica necessariamente numa diminuição da confiança. A única forma de melhorar a precisão sem alterar a confiança é aumentando o tamanho da amostra.

É importante observar também que a confiança e a precisão estão relacionadas com o tamanho da amostra. Nesse sentido, se quisermos manter a confiança e reduzir o intervalo pela metade, precisaremos de uma amostra quatro vezes maior que a proposta inicialmente. Logo, o preço a ser pago pela melhoria da precisão nem sempre será compensado pelo tempo de coleta.

MARGEM DE ERRO (%)	TAMANHO DA AMOSTRA
1,0	9604
2,0	2401
2,5	1537
3,0	1068
3,5	784
4,0	601

Finalmente, devemos frisar que as pesquisas de intenção de voto são bem mais sofisticadas do que a apresentada nesse trabalho e que eventuais variações nos resultados das pesquisas eleitorais realizadas pelos diferentes institutos de pesquisa devem-se basicamente à metodologia adotada.

Testes de Hipótese

Nos testes de hipótese, faz-se uma afirmação referente à população, e o intuito é saber se a proposição é verdadeira ou falsa. Geralmente fazemos mais de uma afirmação, mas nem todas elas devem ser testadas. Aquelas informações que não se pretende que sejam expostas a testes chamam-se *hipóteses subjacentes.* Compõem-se de todos os pressupostos sobre os quais nos apoiamos e nos quais acreditamos. Naturalmente, nunca estamos totalmente certos de que tais pressupostos sejam válidos, caso contrário não seriam pressupostos. Acreditamos que eles possuem validade provável, de modo que as hipóteses subjacentes se encontram muito próximas das corretas. As suposições remanescentes que devem ser testadas chamam-se *hipóteses testáveis.* Como afirmações específicas são mais fáceis de serem rejeitadas do que afirmações vagas, é desejável formular problemas de testes de hipótese de modo a fazer com que a hipótese nula seja a mais específica possível. Isso significa que frequentemente utilizamos como hipótese nula a proposição que de fato queremos rejeitar.

Destacamos que, o critério para a rejeição ou não rejeição da hipótese nula com base em uma amostra não é uma garantia de chegarmos a uma conclusão correta.

O teste de hipótese compõe-se essencialmente de três passos básicos: definição das hipóteses, estabelecimento dos limites entre as regiões de aceitação e rejeição e a obtenção do valor amostral do teste estatístico. No processo de obtenção desse valor duas ocorrências são possíveis: ou a estatística/teste estatístico cai na região de aceitação ou não. Vamos considerar, em primeiro lugar, a segunda ocorrência. Neste caso o valor do teste estatístico é tal que, se a hipótese nula for realmente verdadeira, a probabilidade de ela ocorrer, por acaso, será muito pequena, por exemplo, 5% (intervalo de significância). Isto significa que, se o teste for repetido um número infinito de vezes e se a hipótese nula for realmente verdadeira, rejeitaríamos a hipótese nula 5% das vezes. Chamamos este erro de *erro do tipo I.* Nos testes estatísticos a probabilidade de se cometer tal erro é dada precisamente pelo nível de significância escolhido. Consideremos agora a segunda ocorrência possível do teste, isto é, o caso em que não rejeitamos a hipótese nula, isto é, continuamos julgando-a verdadeira. Mas a probabilidade de que cheguemos a uma conclusão incorreta, isto é, de que a hipótese nula seja realmente falsa não pode ser excluída. Um erro desta espécie chama-se *erro do tipo II*. Nos testes estatísticos, a probabilidade exata desta espécie de erro é geralmente desconhecida, pois depende da diferença entre o valor da hipótese e o verdadeiro parâmetro da população.

	Se H_0 é	
	VERDADEIRA	FALSA
ACEITAR H_0	DECISÃO CORRETA	ERRO TIPO II
REJEITAR H_0	ERRO TIPO I	DECISÃO CORRETA

Exemplo:

Considere uma turma de 10 alunos de uma escola em que se é aplicado dois testes. O objetivo é verificar se houve significativo progresso da turma ao longo do tempo mediante aferição dos resultados obtidos nos dois testes. Suponhamos que os dois testes aplicados possuem dificuldade equivalente.

Para essa medição, atribuímos os números 5, 4, 3, 2, 1 para os diferentes graus A, B, C, D, E, respectivamente.

Sejam os resultados obtidos discriminados na tabela abaixo:

Aluno	1	2	3	4	5	6	7	8	9	10
1ª avaliação	B	C	E	B	C	B	E	C	E	E
2ª avaliação	A	C	D	A	A	D	C	B	B	B
Diferença	+1	0	+1	+1	+2	-2	+2	+1	+3	+3

Após a aplicação dos dois testes, apuramos a diferença total de +12, sinalizando, a princípio, que houve significativo progresso da turma. Entretanto, uma dúvida persiste. A diferença verificada é fruto de uma melhoria de fato ou um mero acaso?

Vamos, por hipótese, assumir que não há razão para crer que houve de fato um progresso da turma. Essa será a nossa hipótese a ser testada (H_0). A hipótese alternativa é a de que, de fato, houve um progresso da turma (H_1).

GRAU	A	B	C	D	E
Número de Ocorrências	3	6	5	2	4
Percentagem	15	30	25	10	20
Percentagem Cumulativa	15	45	70	80	100

Podemos estabelecer uma correspondência entre os números aleatórios conforme a tabela abaixo:

Número Aleatório	GRAU
01 - 15	A
16 - 45	B
46 - 70	C
71 - 80	D
81 - 00	E

Agora, reproduza esse experimento diversas vezes, utilizando para cada experimento 20 números aleatórios e, para cada experimento, faça o registro da diferença total obtida. Verifique quantas vezes a diferença foi maior que +12. Se o número de vezes for pequeno, a hipótese está provavelmente errada, sugerindo que houve um progresso significativo da turma. Caso contrário, a hipótese está provavelmente correta.

Obs: o experimento que acabamos de realizar segue uma distribuição t de Student. A distribuição t de Student é uma das distribuições mais utilizadas na estatística. A sua função densidade tem a mesma forma em sino da distribuição normal, mas reflete uma maior variabilidade (com curvas mais alargadas) que é de se esperar em amostras pequenas.

Para auxílio do professor preparamos uma sequência de passos para a construção de um Teste de Hipótese:

Passo 1: Determinar qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1.

Passo 2: Usar a teoria estatística e as informações disponíveis para decidir qual o estimador será usado para testar a hipótese H_0. Obter as propriedades dessa estatística (distribuição, média, desvio-padrão).

Passo 3: Fixar a probabilidade α de cometer o erro Tipo I e usar este valor para construir a região crítica (regra de decisão).

Passo 4: Use as observações da amostra para calcular o valor da estatística do teste.

Passo 5: Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite H_0; caso contrário, rejeite H_0.

Na construção de hipóteses, sempre vamos estabelecer a hipótese nula como uma igualdade, de forma que a probabilidade do Erro Tipo I, α, pode ser controlada em um valor específico.

Já a hipótese alternativa pode ser tanto unilateral como bilateral, dependendo da conclusão a ser retirada se H_0 é rejeitada.

Se o objetivo é fazer alegações como "maior que", "menor que", "superior a", "excede", "no mínimo", e assim por diante, uma alternativa unilateral é apropriada.

Se nenhuma direção é implicada pela alegação, ou tivermos "diferente de", "não igual a", uma alternativa bilateral deve ser usada.

Portanto, se a hipótese nula e alternativa de um teste de hipótese são:

$$\mathbf{H_0 : \mu = \mu_0}$$

$$\mathbf{H_1 : \mu < \mu_0}$$

Onde μ_0 é uma constante conhecida, o teste é chamado de teste unilateral esquerdo. Para

$$\mathbf{H_0 : \mu = \mu_0}$$

$$\mathbf{H_1 : \mu > \mu_0}$$

O teste é chamado de teste unilateral direito.

No caso em que

$$\mathbf{H_0 : \mu = \mu_0}$$

$$\mathbf{H_1 : \mu \neq \mu_0}$$

o teste é chamado de teste bilateral.

4 SEQUÊNCIA DIDÁTICA – CONSTRUÇÃO DA CURVA NORMAL e INFERÊNCIA ESTATÍSTICA

Nesta parte do trabalho propomos uma sequência didática para a construção da Curva Normal e uma sequência didática para o ensino da inferência estatística nas escolas de ensino médio do nosso país. O objetivo é, principalmente, adequar o ensino da matemática ministrado em nossas escolas às melhores práticas de ensino desenvolvidas ao redor do mundo. Nesse sentido, consideramos importante a utilização de recursos computacionais para melhor ilustrar os assuntos expostos, ainda que não sejam considerados imprescindíveis para a sua compreensão.

Segundo Zabala, as sequências didáticas são um conjunto de atividades ordenadas, estruturadas e articuladas para a realização de certos objetivos educacionais, que têm um princípio e um fim conhecidos, tanto pelos professores quanto pelos alunos. (ZABALA, 1998)

Planejamento da Sequência

DISCIPLINA: Matemática

SÉRIO: Ensino Médio

CONTEÚDO: Curva Normal e Inferência Estatística

OBJETIVO:

Geral: Apresentar a Curva Normal e demonstrar o que significa fazer uma Inferência Estatística

Específicos: Construir a Curva Normal e descrever suas propriedades. Mostrar, também, a conexão existente entre o estudo das probabilidades e a inferência estatística.

TEMPO ESTIMADO: 4 a 5 aulas para realização das atividades relacionadas à Curva Normal e 2 aulas para o ensino da inferência estatística

Plano de Aula – Curva Normal

Atividade 1

Nessa atividade o aluno fará uma análise dos dados apresentados e responderá aos seguintes questionamentos. O objetivo desse exercício é mostrar ao aluno a quantidade de informações que podemos extrair ao organizar os dados segundo métodos apropriados:

Dado o rol de medidas das alturas (dadas em cm) de uma amostra de 100 indivíduos de uma faculdade:

151	152	154	155	158	159	159	160	161	161
161	162	163	163	163	164	165	165	165	166
166	166	166	167	167	167	167	167	168	168
168	168	168	168	168	168	168	168	169	169
169	169	169	169	169	170	170	170	170	170
170	170	171	171	171	171	172	172	172	173
173	173	174	174	174	175	175	175	175	176
176	176	176	177	177	177	177	178	178	178
179	179	180	180	180	180	181	181	181	182
182	182	183	184	185	186	187	188	190	190

Calcule:

a) a amplitude amostral;

b) o número de classes;

Comentário: para facilitar o desenvolvimento dessa atividade o professor pode estabelecer a priori o número de classes.

c) os limites de classes;

d) as freqüências absolutas da classes;

e) as freqüências relativas;

f) os pontos médios da classes;

g) as freqüências acumuladas;

h) o histograma e o polígono de freqüência;

i) o polígono de freqüência acumulada;

j) faça um breve comentário sobre os valores das alturas desta amostra através da distribuição de frequência.

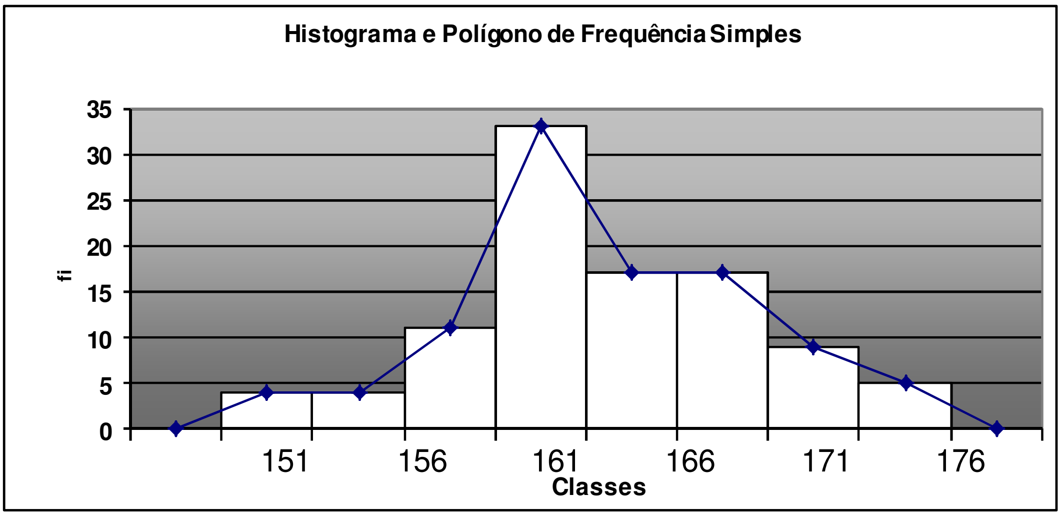
Solução

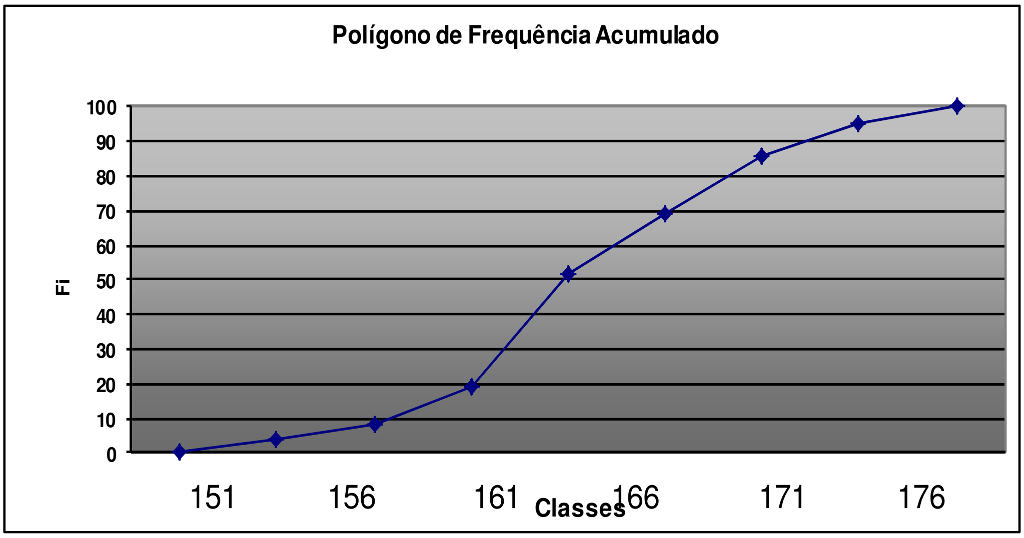
A_i = 190-151 = 40

K = 1 + 3,32 log (100) = 7,64 ≈ 8 (Fórmula de Sturges)

K = 40/8 = 5

Classes	fi	fri	Fi	Fri	xi
151 |- 156	4	0,04	4	0,04	153,5
156 |- 161	4	0,04	8	0,08	158,5
161 |- 166	11	0,11	19	0,19	163,5
166 |- 171	33	0,33	52	0,52	168,5
171 |- 176	17	0,17	69	0,69	173,5
176 |- 181	17	0,17	86	0,86	178,5
181 |- 186	9	0,09	95	0,95	183,5
186 |- 191	5	0,05	100	1,00	188,5
Total	100	1,00	-	-	-





Atividade 2

Será apresentado o experimento de Galton na construção de uma distribuição normal e sua associação com o triângulo de Tartaglia.

Essa atividade poderá ser realizada sem o auxílio de recursos computacionais. Entretanto, para melhor ilustrar o experimento de Galton recomendamos a utilização do aplicativo disponibilizado no site http://www.mathsisfun.com/data/quincunx.html ou qualquer outro disponível em sites especializados de matemática. Por meio desse aplicativo é possível simular a aproximação do desenvolvimento binomial da Curva Normal à medida que se aumenta o tamanho da amostra.

Atividade 3 – construção da Curva Normal

I: Tarefa

Os estudantes serão instruídos a coletar dados e usar os dados para achar a media e o desvio padrão para construção da curva normal. Os estudantes também responderão questões envolvendo desvio padrão, valor Z (valor padronizado) e variância.

II: Objetivos

O estudante será capaz de

1) Organizar e coletar dados sobre o assunto pesquisado;
2) Esboçar o gráfico da Curva Normal;
3) Usar calculadora e/ou planilha excel para achar a media e o desvio padrão; e
4) Analisar os dados e tecer conclusões sobre o experimento.

III: Requisitos

Os estudantes deverão:

1) Possuir conhecimento básico de medidas de tendência central.
2) Saber coletar dados e construir um histograma.
3) Possuir conhecimento básico da fórmula do valor Z (valor padronizado) e reconhecer as variáveis correspondentes.

Antes de iniciarmos a atividade propriamente dita teremos uma 1 aula de revisão de conceitos estatísticos. Esta primeira etapa tem como objetivo equalizar o conhecimento estatístico mínimo necessário para a construção da Curva Normal.
Essa revisão contemplará os seguintes assuntos: cálculo da média, desvio padrão e variância de uma amostra, construção de um histograma e conversão de qualquer variável normal com média μ e variância σ^2 em uma variável normal padrão (consultar o assunto no texto-base).

Caso o professor considere necessário, poderá fazer uso do aplicativo Geogebra, onde será mostrado que a Curva Normal fica plenamente determinada com apenas dois parâmetros, a saber: média e desvio-padrão. Poderá, inclusive, mediante o uso do recurso “controle deslizante” fazer a transformação gráfica de uma distribuição normal qualquer na distribuição normal padrão correspondente.

Construção da Curva Normal (sequência de passos)

Para construção da Curva Normal os estudantes serão divididos em grupos. Cada grupo irá recolher os dados (p.ex. amostra de tamanho 15) que servirão para sua construção. Abaixo, segue orientação pormenorizada da atividade a ser desenvolvida:

1) 1 aula para discutir e planejar a coleta de dados. Escolha uma das opções listadas abaixo:

   a) Notas de testes realizados por estudantes de uma turma;
   b) Altura das pessoas (escolher um gênero – masculino ou feminino);
   c) Outro assunto qualquer.

2) A coleta de dados deverá ser realizada em até uma semana a partir da apresentação da tarefa pelo professor (essa atividade será desenvolvida em período fora do horário de aula)

3) 1 aula para que os alunos, utilizando os dados coletados, encontrem a media e o desvio padrão para construção da curva normal e respondam os outro itens do questionário distribuído pelo professor.

4) Os trabalhos serão devolvidos ao professor e, os resultados obtidos e as respectivas conclusões debatidas em classe. A classe discutirá os aspectos positivos e negativos dessa tarefa (1 aula)

Ao longo da tarefa, o professor deverá acompanhar os números obtidos a fim de se certificar que eles fazem sentido. Isso permitirá que o professor discuta questões que possam surgir como, por exemplo, pontos fora da curva.

Questionário

1)desenhar a curva normal

2)qual o valor associado a z=0,5

R:______________

3)qual o valor associado a z= -2.4

R:______________

4)calcular a variância do conjunto de dados

R:_______________

5)qual é o intervalo de valores de dados para que 99% dos dados se encontrem em torno da média?

R:_______________

6)existe algum valor atípico no conjunto de dados? (valor atípico refere-se ao valor que esteja além de 3 desvios-padrão da media)

R:_______________

7) por que nem sempre todos os dados se encontram em até 3 desvios-padrão da media?

R:______________

Avaliação do docente

O professor deverá atentar para os seguintes aspectos da tarefa:

1) os dados foram corretamente tabelados?

2) os dados estão organizados e de fácil entendimento?

3) a curva normal foi desenhada corretamente?

4) os elementos-padrão da curva normal foram identificados corretamente?

5) os cálculos foram feitos corretamente?

Planilha de Coleta de Dados

	Altura dos alunos do sexo masculino de uma mesma série escolar
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Plano de Aula – Inferência Estatística

Atividade 1

Nessa atividade o aluno resolverá a questão abaixo justificando sua resposta:

1) Maria possui um casal de cães (de mesma raça). Ela verificou que geralmente nascem mais filhotes machos do que fêmeas. Para confirmar sua suposição, ela decidiu cruzar novamente os cães e obteve 15 filhotes machos de 18 nascidos nessa cria.

   Vamos testar a hipótese de que cada filhote tem igual chance de nascer macho ou fêmea e, como hipótese alternativa, a chance de nascer macho seja maior que a chance de nascer fêmea.

   A tabela abaixo mostra o resultado obtido em 1000 experimentos (simulações) com cães dessa mesma raça:

1) De acordo com a simulação, qual a probabilidade de nascer 15 ou mais filhotes machos do total de 18?

R:______________

Obs: caso o resultado observado possuir probabilidade de ocorrência menor que 1%, a hipótese será rejeitada.

Nº de Filhotes Machos	Frequência
0	0
1	0
2	1
3	3
4	12
5	33
6	71
7	121
8	167
9	184
10	167
11	121
12	71
13	33
14	12
15	3
16	1
17	0
18	0

1) O que podemos concluir a respeito da hipótese?

Escolha uma das respostas abaixo:

a) Não podemos rejeitar a hipótese.
b) Devemos rejeitar a hipótese.

Resposta: b

Justificativa:

De acordo com a tabela, de 1000 experimentos (simulações):

- Em 3 simulações se obteve 15 filhotes machos

- Em 1 simulação se obteve 16 filhotes machos
- Não houve experimento em que houvesse nascido 17 ou 18 filhotes machos.

No total, houve 4 experimentos em 1000. Isso implica que a probabilidade de nascer 15 ou mais filhotes machos é de 0,4%. Portanto, é mais razoável rejeitar a hipótese de que a chance de nascer macho seja a mesma de nascer fêmea.

Em outras palavras, o resultado observado é bastante improvável. Logo, devemos rejeitar a hipótese.

Atividade 2

Nessa atividade os alunos construirão um teste de hipótese com base nos dados colhidos para o trabalho de construção da Curva Normal. O objetivo dessa atividade é familiarizar o aluno com a estrutura de um teste de hipótese. O aluno deverá seguir os passos abaixo:

1) Determinar a hipótese H_0 a ser testada e a hipótese alternativa H_1

   Sugestão: a hipótese a ser testada poderá ser a média de altura dos alunos do sexo masculino de uma mesma série escolar.

   Comentário: a escolha das hipóteses H_0 e H_1 ficarão a critério do aluno. O professor deverá verificar se a escolhas foram adequadas.
2) Utilizar o z-teste (teste de hipótese baseado na estatística Z)
3) Fixar o nível de significância α

   Sugestão: utilizar $\alpha = 5\%$ ou $\alpha = 1\%$

4) Usar a amostra (tamanho da amostra = 15) para o cálculo do valor da estatística do teste
5) Verificar se a hipótese deverá ser rejeitada ou aceita para o nível de significância utilizado ($\alpha = 5\%$ ou $\alpha = 1\%$)

5 CONSIDERAÇÕES FINAIS

Ao longo do trabalho, houve relativa dificuldade quanto à disposição em que os tópicos deveriam ser apresentados, visto ser necessário o domínio de vários conceitos estatísticos para o desenvolvimento do tema propriamente dito.

Acreditamos que o contato do aluno do ensino médio com a Curva Normal e com a estatística inferencial é de fundamental importância para uma análise crítica de diversos outros tópicos ministrados em outras disciplinas do ensino médio. Além disso, um primeiro contato com esses assuntos servirá como base para eventuais estudos futuros na Universidade, independentemente da área escolhida. Em determinadas partes do trabalho o rigor matemático foi dispensado para uma maior fluidez do assunto.

Destaco a fundamental relevância do Teorema Central do Limite sem a qual o desenvolvimento do tema proposto careceria de uma maior aplicabilidade.

Espero que esse trabalho possa contribuir para que o ensino da estatística em nossas escolas de ensino médio não fique limitado à apresentação da estatística descritiva básica e, principalmente, conscientizar nossos professores do momento mais adequado para desenvolver esses assuntos em sala de aula.

REFERÊNCIAS

Livro.

- Elementos de Econometria – Teoria Estatística Básica
  Jan Kmenta – Professor de Economia
  Michigan State University
  Editora Atlas S.A.

- Business Statistics
  by Douglas Downing, Ph.D.
  School of Business and Economics (Seatle Pacific University)

  and Jeffrey Clark, Ph.D.
  Mathematics Department (Elon University)
  Barron's Educational Series, Inc.

Documento eletrônico.

http://www.biblioteca.pucminas.br/teses/EnCiMat_DuarteLR_1.pdf

Trabalho apresentado em evento.

- Didáctica de lá Estadística
  Grupo de Educación Estadística
  Universidad de Granada
  Departamento de Didáctica de la Matemática
  Carmem Batanero

Apêndice A – Tabela da Curva Normal

Áreas de uma distribuição normal padrão. Cada casa na tabela dá a proporção sob a curva entre z=0 e um valor positivo de z. As áreas para os valores de z negativos são obtidas por simetria.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Apêndice B – Fator de Correção de Bessel

No cálculo da variância amostral utilizamos n-1 no lugar de n, sempre que a média da população for desconhecida, onde n é o número de observações de uma amostra. Caso não se faça essa correção, a variância amostral tornar-se-á um estimador tendencioso da variância populacional, com valor inferior ao valor esperado.

Dado o escopo do nosso trabalho não apresentaremos a demonstração da Correção de Bessel, nos restringindo somente às razões pelas quais a variância amostral se torna um estimador tendencioso da variância populacional.

Suponha que a média populacional seja 1050, porém desconhecida pelo investigador/pesquisador. Considere uma amostra (n=5) aleatória da população cujos valores são:

1051, 1053, 1055, 1050, 1051

Logo, a média amostral será:

(1051 + 1053 + 1055 + 1050 +1051)/5 = 1052

Calculando a variância populacional temos:

$[(1051 - 1050)^2 + (1053 - 1050)^2 + (1055 - 1050)^2 + (1050 - 1050)^2 + (1051 - 1050)^2]/5 = 7{,}2$

Mas a nossa estimativa para a média populacional é a média amostral cujo valor é de 1052.

Portanto temos:

$[(1051 - 1052)^2 + (1053 - 1052)^2 + (1055 - 1052)^2 + (1050 - 1052)^2 + (1051 - 1052)^2]/5 = 3{,}2$

Considere, também, a identidade abaixo:

$(a + b)^2 = a^2 + 2ab + b^2$

$[1053 - 1050]^2 = [(1053 - 1052) + (1052 - 1050)]^2$

$= (1053 – 1050)^2 + 2(1053 - 1052)(1052 - 1050) + (1052 – 1050)^2$

a^2	2ab	b^2
$(1051 - 1052)^2$	2(1051 - 1052)(1052 - 1050)	$(1052 - 1050)^2$
$(1053 - 1052)^2$	2(1053 - 1052)(1052 - 1050)	$(1052 - 1050)^2$
$(1055 - 1052)^2$	2(1055 - 1052)(1052 - 1050)	$(1052 - 1050)^2$
$(1050 - 1052)^2$	2(1050 - 1052)(1052 - 1050)	$(1052 - 1050)^2$
$(1051 - 1052)^2$	2(1051 - 1052)(1052 - 1050)	$(1052 - 1050)^2$

- A soma da primeira coluna (a^2) é a soma dos quadrados dos desvios em relação à média amostral;

- A soma das colunas (a^2 e b^2) é a soma dos quadrados dos desvios em relação à média populacional; e

- A soma da coluna intermediária (2ab) deve ser zero.

Concluímos, então, que a soma dos desvios quadráticos em relação à média populacional será sempre maior que a soma dos desvios quadráticos em relação à média amostral, exceto quando a média populacional for igual à média amostral. Nesse caso singular não haverá necessidade de aplicarmos a correção de Bessel.

Apêndice C – Glossário

Amostra – grupo de itens escolhido de uma população.

Desvio Padrão – medida de dispersão que mostra o quanto de variação ou "dispersão" existe em relação à média, ou valor esperado. O desvio padrão define-se como a raiz quadrada da variância. É simbolizado por σ.

Distribuição Binomial - distribuição de probabilidade discreta aplicável a um experimento realizado *n* vezes em que cada experimento possui uma probabilidade p de sucesso e que cada tentativa seja independente das outras.

Valor Esperado – valor médio que se espera obter ao observarmos uma variável aleatória inúmeras vezes. Também chamado de média , e é simbolizado por μ ou *E(X).*

Experimento Aleatório – experimento cujo resultado não pode ser previsto até que o experimento ocorra.

Função probabilidade - função que associa a cada possível ocorrência de uma variável aleatória discreta uma probabilidade.

Lei dos Grandes Números (LGN) - é um teorema fundamental da teoria da probabilidade, que descreve o resultado da realização da mesma experiência repetidas vezes. De acordo com a LGN, a média aritmética dos resultados obtidos ao se realizar por sucessivas vezes um mesmo experimento tende a se aproximar do valor esperado à medida que mais tentativas se sucederem. Em outras palavras, quanto mais tentativas são realizadas, mais a probabilidade da média aritmética dos resultados observados se aproxima da probabilidade real.

População - totalidade de todas as observações possíveis sobre medidas ou ocorrências, podendo ser finita ou infinita. Uma população finita é aquela na qual o número de observações possíveis é limitado (menor que o infinito).

Variância – medida de dispersão para uma variável aleatória. É simbolizada por σ^2 ou $VAR\,(X)$, e seu valor é dado por $VAR(X) = E(X^2) - [E(X)]^2$

Variável aleatória – variável cujo valor depende do resultado de um experimento aleatório, ou seja, aquela cujo valor não pode ser plenamente controlado ou determinado antes da observação. Também chamada de variável randômica ou estocástica.

Variável aleatória contínua - variável aleatória que pode assumir qualquer número real em um intervalo. É caracterizada pela curva Função Densidade de Probabilidade (FDP) em que a área abaixo da curva entre dois números quaisquer representa a probabilidade da variável aleatória se situar entre esses dois números. Podemos citar como exemplos típicos deste tipo de variável a temperatura e a renda familiar. As variáveis aleatórias contínuas são usualmente mais fáceis de manipular matematicamente do que variáveis aleatórias discretas (variável que pode assumir somente valores específicos em um intervalo). Se uma distribuição discreta possui vários valores possíveis próximos uns dos outros, então pode ser aproximada por uma distribuição contínua. A Distribuição Normal é a mais importante distribuição de variável aleatória contínua.

Variável aleatória discreta – variável aleatória que pode assumir um conjunto finito ou enumerável de números.