

Biologia Quantitativa

Populações de Organismos Distribuições Testes de Hipóteses

Depto de Zoologia
Módulo 02 2024-01
02 de abril de 2024

Roteiro da Aula

- Vamos apresentar 3 temas
- Um tema de biologia - populações e distribuições
- Um tema de estatística - teste de hipóteses
- Um tema de R e Rstudio - o R Notebook e os comandos para testar hipóteses com o teste de t

Teste de Hipóteses - Método Popper

- Segue a filosofia de Karl Popper, fundador do positivismo lógico
- Não é possível provar definitivamente uma hipótese (pode sempre haver um dado futuro que contradiga a hipótese)
- É possível rejeitar a hipótese se encontrarmos dados que sejam diferentes dos hipotetizados
- O método estatístico é baseado em testar hipóteses, rejeitar, ou aceitar (não é provar)
- Definir hipótese nula - não há efeitos, variação.
- Testar hipótese com dados e teste estatístico usando distribuição esperada
- Rejeitar hipótese nula e aceitar hipótese alternativa caso os dados mostrem diferenças
- Ou então aceitar a hipótese nula caso os dados não mostrem diferenças

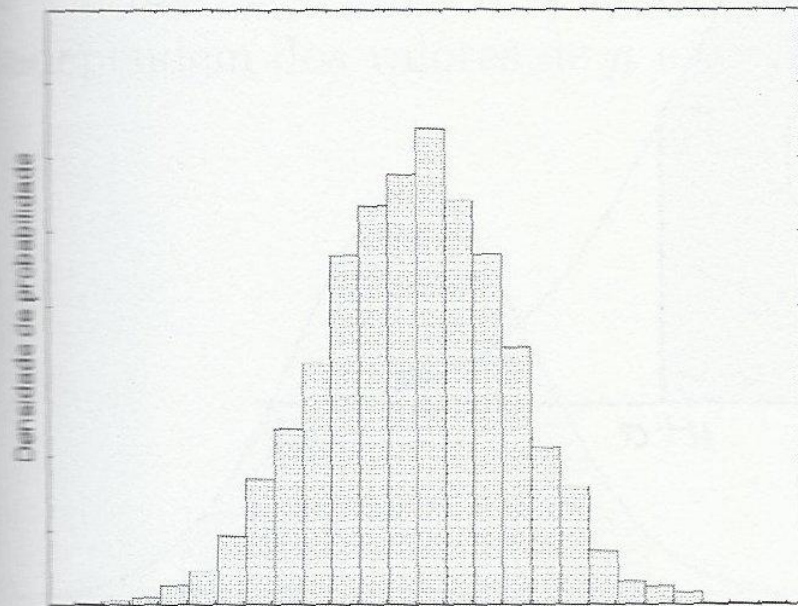
Teste de Hipóteses

- 1 Definir o desenho lógico para testar as hipóteses,
 - verificar se de fato são testáveis
- 2 Definir os tipos de variáveis que estamos medindo
- 3 Definir o tipo de distribuição com que lidamos
- 4 Definir a estatística a ser usada para o teste
- 5 Definir o método (= tipo de teste estatístico) para testar a hipótese
- Nota: existe uma categoria de testes chamados não paramétricos que dispensam o conhecimento da distribuição estatística das amostras. Mesmo assim calculam a distribuição estatística do parâmetro a testar

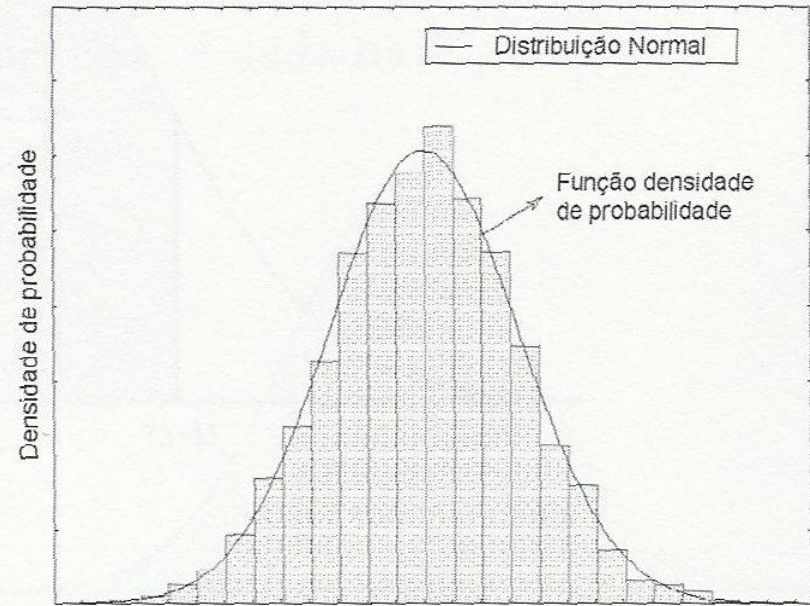
Distribuição Normal

- Surgiu há aproximadamente 200 anos
- Forma de sino
- Parâmetros: média, variância, desvio padrão
- Integral = 1
- Função densidade sempre > 0
- Simétrica em torno da média
- Para cada combinação de média e desvio padrão há uma curva normal

Distribuição Normal - Origem



(a) Distribuição de frequência em forma de sino



(b) Uma curva contínua que aproxima a distribuição de frequências observadas

Figura 5.6 – A curva normal

Distribuição Normal

Distribuição

Formula

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

$f(x)$ = probability density function

σ = standard deviation

μ = mean

Desvio Padrão

Formula

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ = population standard deviation

N = the size of the population

x_i = each value from the population

μ = the population mean

Distribuição Normal - Estatísticas

- Variância: soma dos quadrados / n
- Desvio padrão: raiz quadrada variância (unidade)
- Coeficiente de variação: $(dp/media)*100$

- Necessidade de ter estimador não tendencioso
- Como gerar estimadores da pop a partir de amostra
- Média amostral = não tendencioso
- Variância amostral = ok se dividir por $n-1$ e não n
- Desvio padrão amostral: precisa de fórmula de correção
-
- Erro padrão da média: $dpa/raiz(n)$

Distribuição Normal

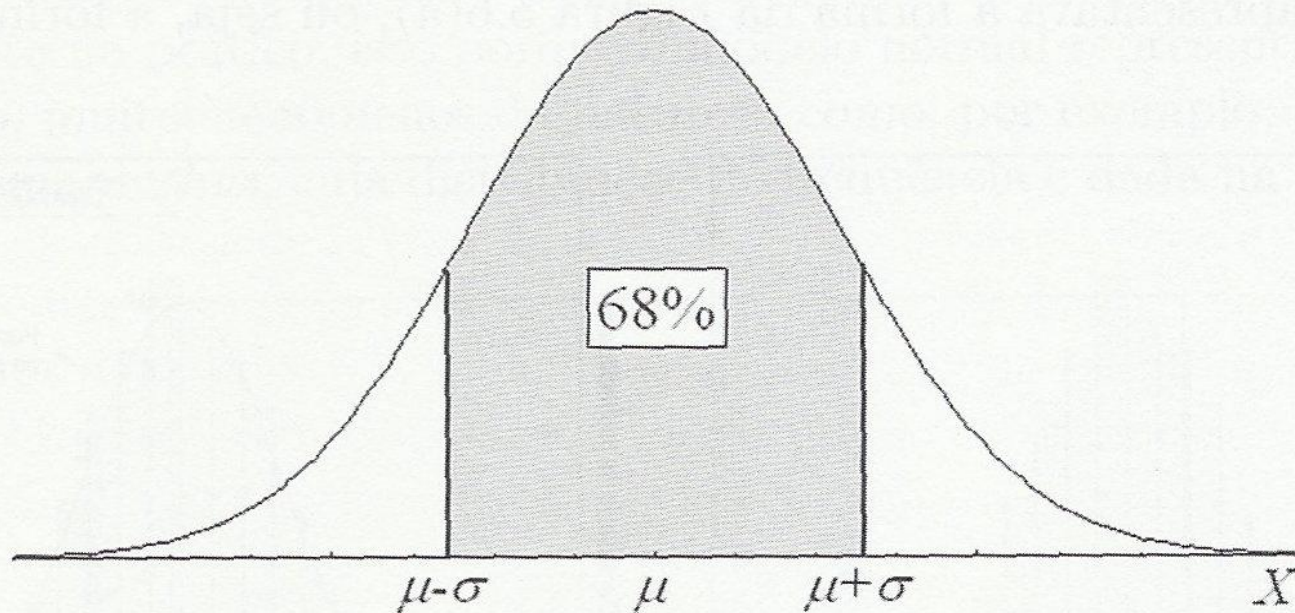


Figura 5.7 – A distribuição normal com parâmetros μ e σ

Distribuição Normal 3 casos

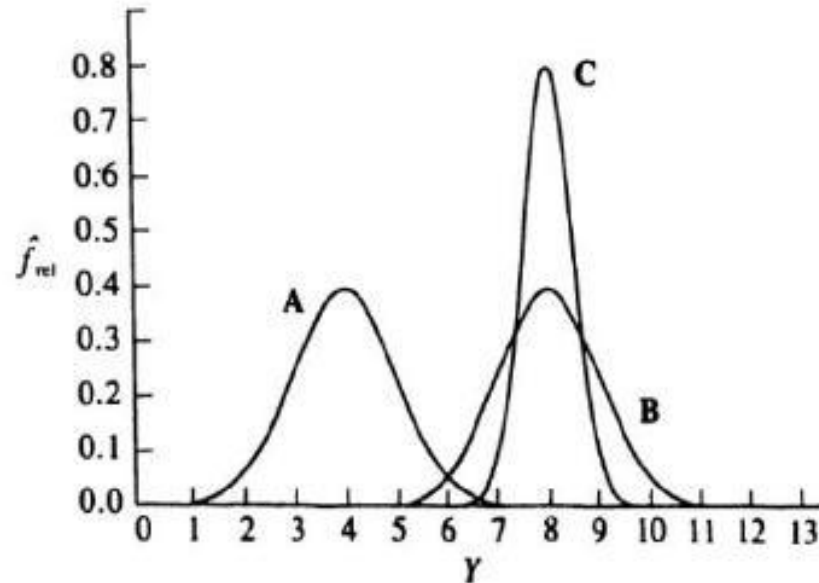
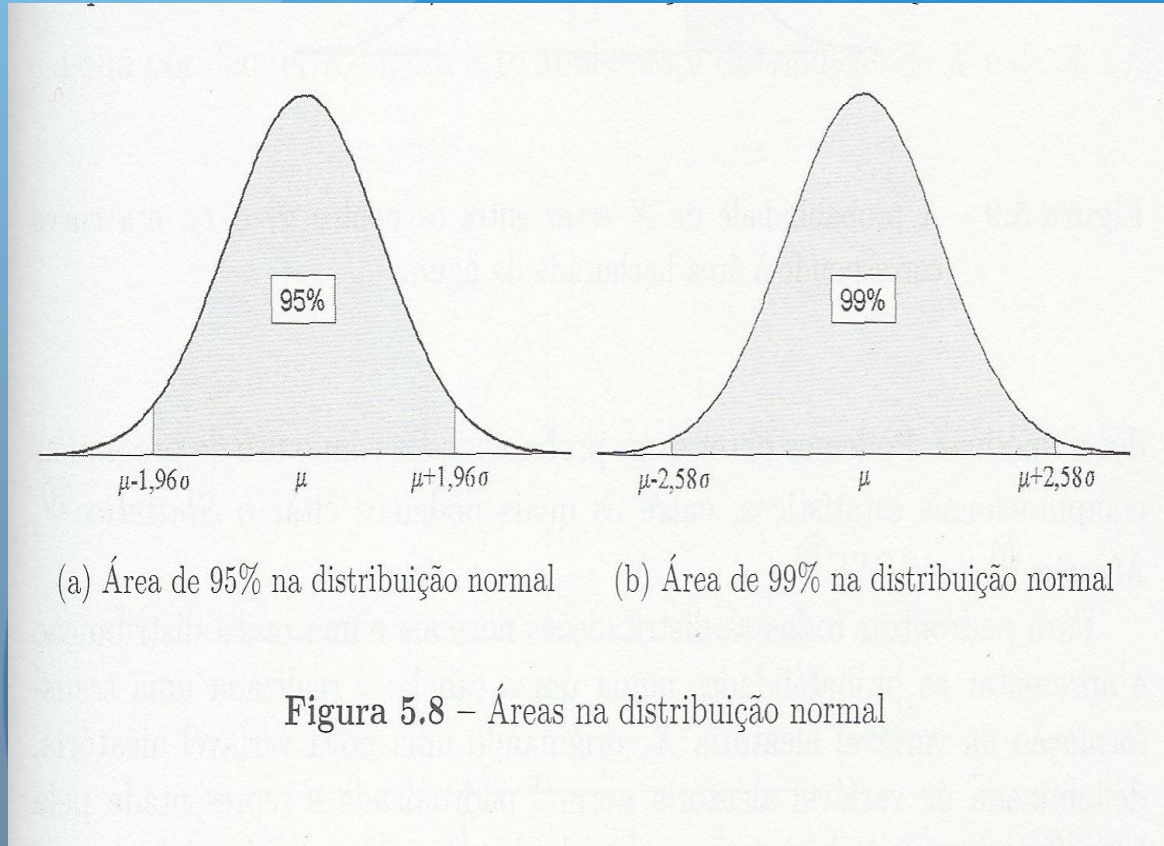


FIGURE 6.2 Changes in the two parameters of the normal distribution affect the shape and location of the normal probability density function. A. $\mu = 4$, $\sigma = 1$; B. $\mu = 8$, $\sigma = 1$; C. $\mu = 8$, $\sigma = 0.5$.

Distribuição Normal - Áreas



Distribuição Normal - Lim Conf

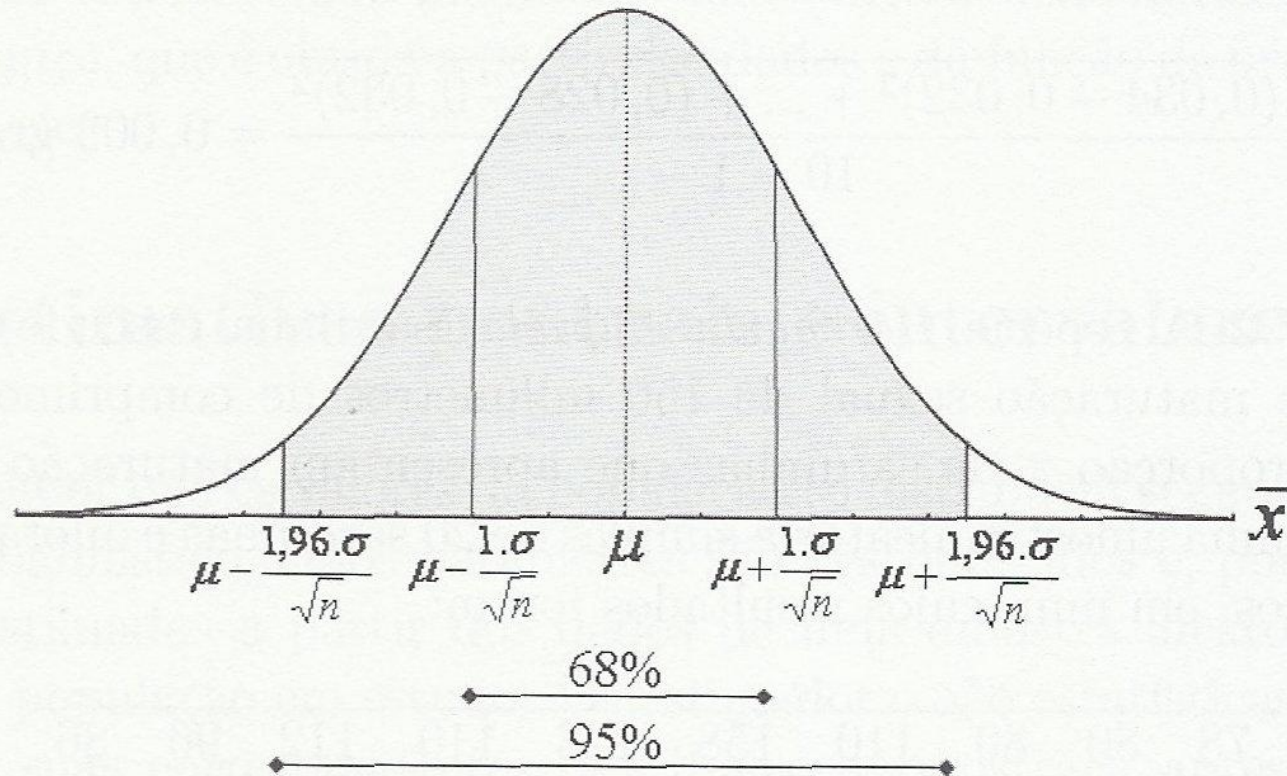
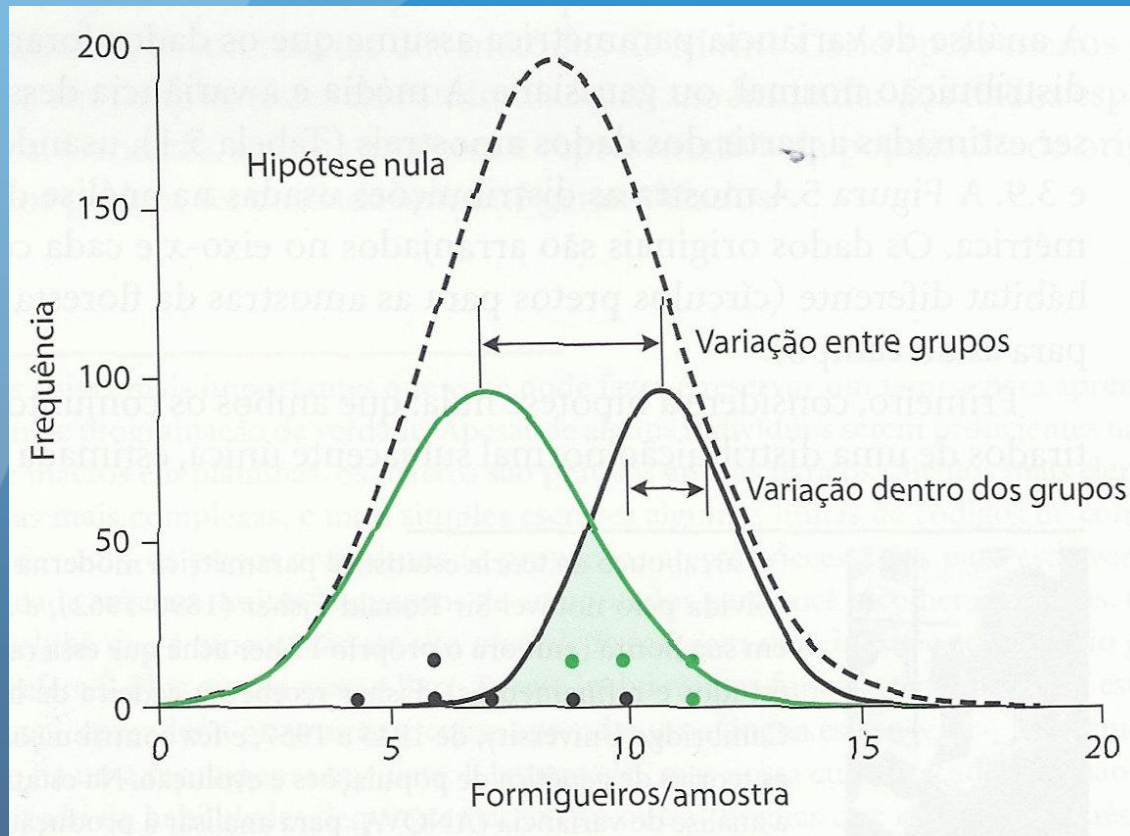


Figura 7.2 – Intervalo para uma média com 68% e 95% de confiança

Distribuição Normal - Int Conf

- Média da amostra diferente da média populacional
- Diferença é o erro amostral
- O valor do erro não pode ser calculado, mas a partir da variância da amostra podemos o erro amostral máximo e seu intervalo de confiança (sec 7.3.2 Andrade Ogliari)
- A média da amostra é um estimador não tendencioso da média populacional
- A variância da amostra é estimador não tendencioso se dividido por $n-1$ em vez de n (soma dos quadrados / $(n-1)$)
- O desvio padrão (raiz quadrada da variância amostral) é estimador tendencioso, precisa de correção. Diminui com amostras grandes

Qual é a Hipótese Nula?



Teste de Hipóteses

- Conceito de hipótese nula e alternativa cap 8 a&o
- Erros tipo 1 e tipo 2
- Tipo 1: rejeitar hipótese nula quando é verdadeira
- Tipo 2: não rejeitar hipótese nula quando é falsa
- Erro tipo 1: significância do teste – 5%
- Erro tipo 2: geralmente aceitável 20%

Exemplos de Teste de Hipóteses

- Testar se uma observação vem de determinada população com distribuição conhecida
- Testar se duas amostras vem da mesma população (conhecida ou estimada)
- Medir um parâmetro de uma amostra (média, variância, ranking, contagens, etc) e e comparar com o de outra amostra usando um teste conhecido e as propriedades daquele parâmetro para determinar se as amostras vem da mesma população.

Teorema do Limite Central

- A distribuição das médias de qualquer distribuição tende à distribuição normal
- Independe do tipo de distribuição original

Teorema do Limite Central

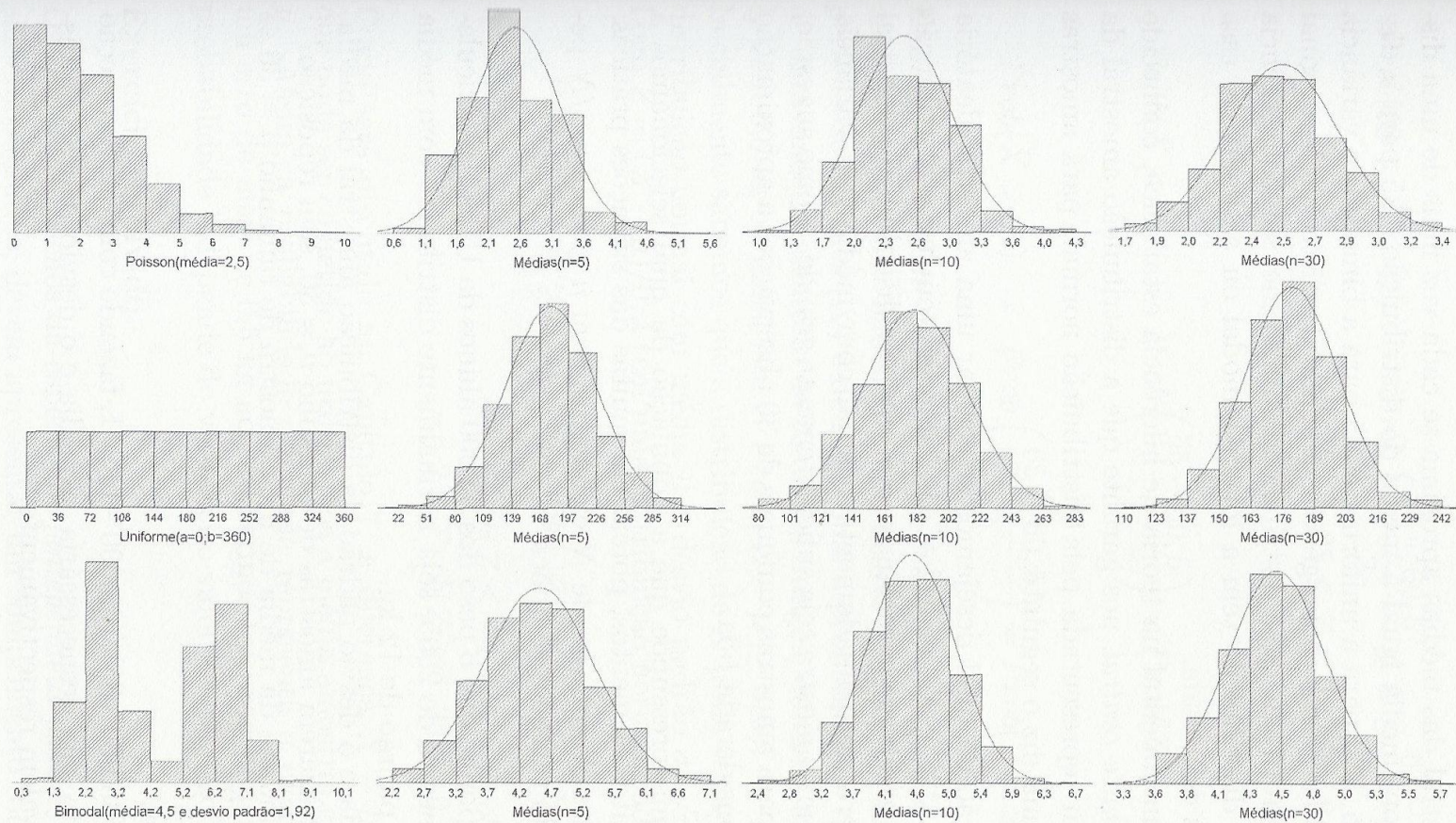


Figura 6.6 – Histogramas da distribuição amostral das médias obtidas de amostras de três tamanhos ($n = 5$, $n = 10$ e $n = 30$) para três distribuições (Poisson, uniforme e bimodal)

Distribuição Student t

- Andrade e Ogliari sec 7.3.3
- Na maioria das situações o erro padrão da população não é conhecido
- Entretanto podemos calcular o erro padrão da amostra
- Usar o erro padrão da amostra como estimador da população pode alterar a distribuição de probabilidade
- Usa-se ao invés da distribuição normal a distribuição t de student com parâmetro $n-1$
- Este parâmetro é conhecido como graus de liberdade

Distribuição de Student

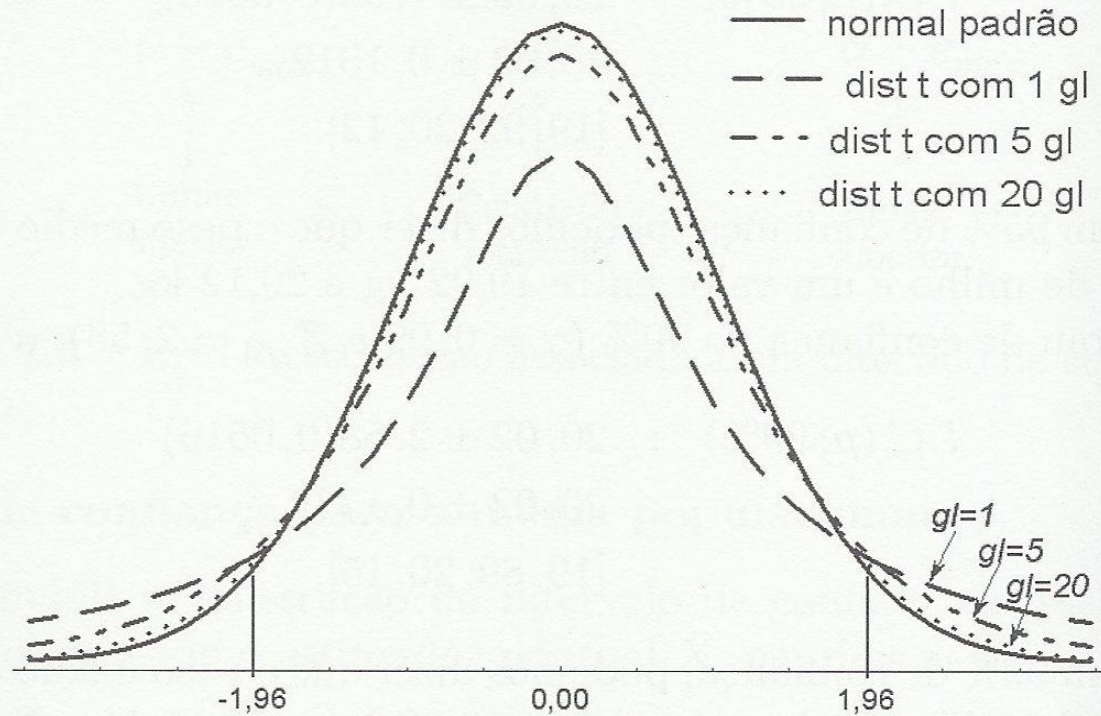


Figura 7.4 – Formas das distribuições t de Student e normal padrão

Abordagem Frequencista

- Desenho experimental / observacional pré-definido Bolker p.10
- Calcular probabilidade de resultado particular, definido como a frequência média daquele resultado em uma sequência a longo prazo de experimentos repetidos.
- Calcular o valor p , ou seja a probabilidade do resultado particular ou de resultados mais extremos dada uma hipótese nula
- Caso a probabilidade for baixa, rejeitar a hipótese nula
- Críticas: mesmo com valores de p baixos ($< 5\%$), rejeita-se a hipótese nula em alguns casos - um em 20 para 5%. Se o $p > 5\%$ aceitamos a hipótese nula com o risco da mesma ser falsa.

Abordagem Frequencista Slide 2

- Problemas com hipótese nula:
- Pode ser rejeitada mesmo com diferenças pequenas em relação à alternativa, em situações de grandes amostras que dão significancia estatística.
- Hipótese nula pontual (ex: inclinação da reta=0) não é realista. O teste acaba por depender de termos dados suficientes para rejeitar a hipótese nula.
- Recomendação (Bolker). Melhor estimar os valores dos parâmetros biologicamente significativos e obter seus intervalos de confiança do que centrar nos valores de p.

Abordagem Bayesiana - Bolker

- Os dados observados são a realidade. Os parâmetros ou hipóteses tem distribuições probabilísticas
- Frequencista: existe um conjunto de parâmetros reais. Os dados experimentais fazem parte de uma distribuição de possíveis resultados.
- Vantagens da abordagem bayesiana: respostas são derivadas dos dados e não de uma sequência hipotética de repetições; permite fazer afirmativas sobre probabilidades de diferentes hipóteses ou valores de parâmetros.
- Dificuldades: na estatística Bayesiana as probabilidades das hipóteses tem de ser definidas a priori.
- Vantagens bayesianas: se há dados anteriores para incorporar à análise; modelos complexos multivariáveis; lacunas de dados; apoio à decisão (eventos raros/catastróficos)

Abordagem por Estimadores- Bolker

- Estimativa por máxima verossimilhança (Andrade e Ogliari sec 7.2). Combinado com análise frequencista.
- Dado um modelo estatístico, este método obtém o conjunto de parâmetros que maximizam a probabilidade de ocorrência dos dados observados.
- Geralmente usa-se o logaritmo da máxima verossimilhança (log-likelihood)
- Aceita-se um intervalo de 2 unidades $\log e^{-2} \sim 1/7.4 = 14\%$
- Outra abordagem: em amostragem repetida, distribuição do log negativo segue qui-quadrado. Usa se 95% da distribuição qui-quadrado ou 1.92 unidades log ou razão $e^{1.92} = 6.82$ como fator de decréscimo. Fora do intervalo apenas 5% no tempo

Desenho Lógico

- Andrade e Ogliari sec 1.2 e 1.3
- Vieira Bioestatística cap 1,
- Todas as hipóteses contempladas nas categorias experimentais ou observacionais
- Pré- definição das alternativas que testam cada hipótese
- Tamanho amostral compatível com os valores das variáveis e a variação esperada
- Evitar pseudo-replicação

Exemplos de Distribuições

- Vieira introdução à bioestatística caps 9,10,11,12,13
- Sokal & Rohlf caps 5,6,7,8
- Normal
- Binomial
- t
- Qui-quadrado
- F
- Poisson

Exemplos de Distribuições

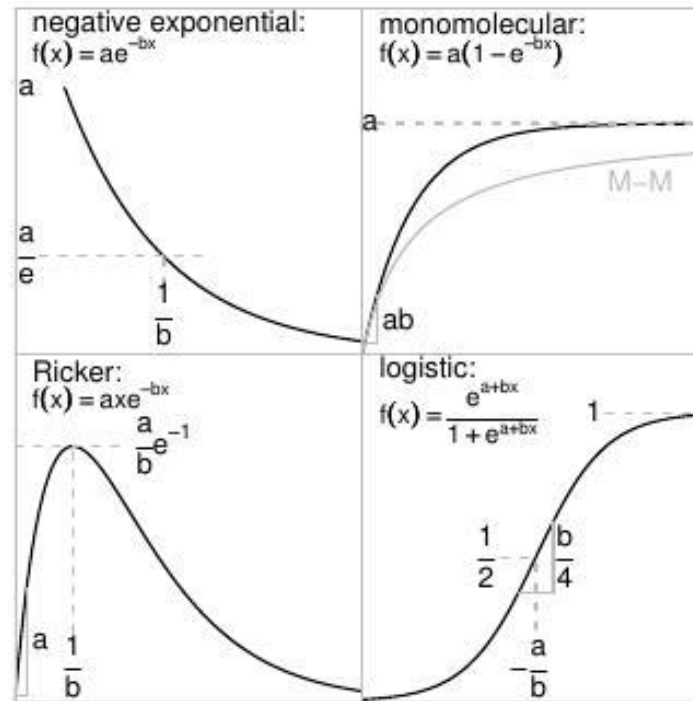


Figure 3.9 Exponential-based functions. "M-M" in the monomolecular figure is the Michaelis-Menten function with the same asymptote and initial slope.

Distribuição Normal 3 casos

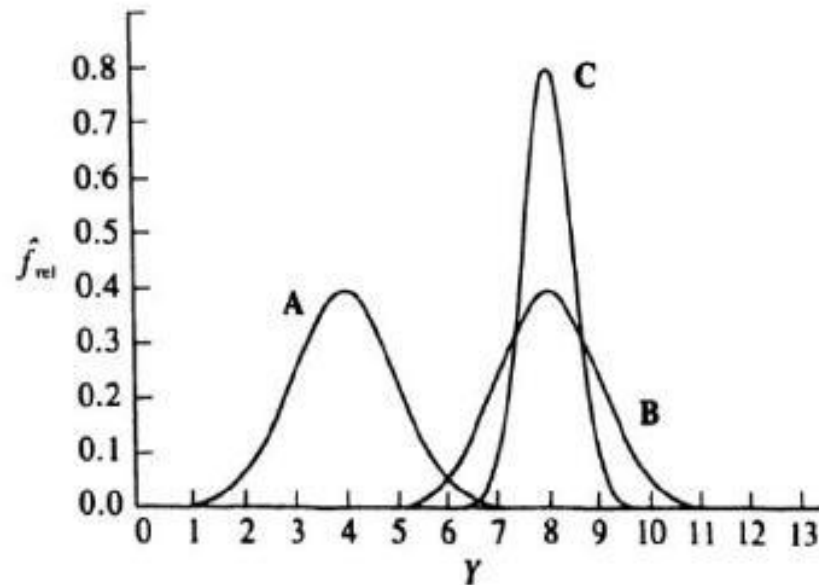


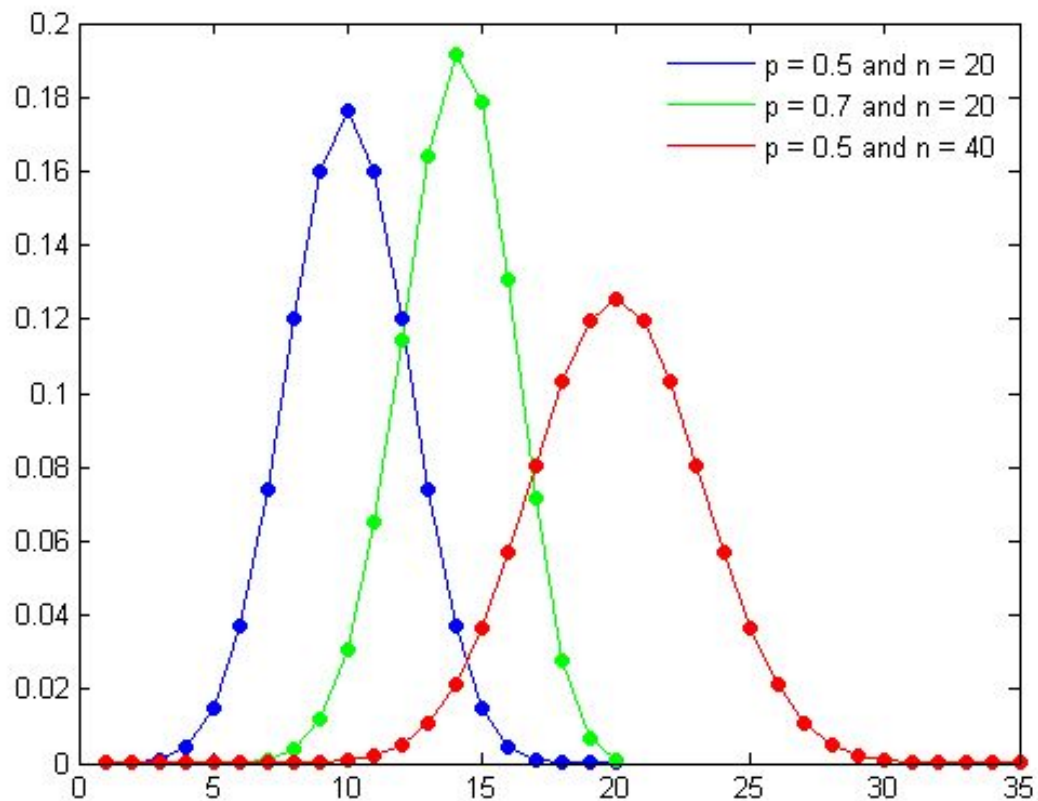
FIGURE 6.2 Changes in the two parameters of the normal distribution affect the shape and location of the normal probability density function. **A.** $\mu = 4$, $\sigma = 1$; **B.** $\mu = 8$, $\sigma = 1$; **C.** $\mu = 8$, $\sigma = 0.5$.

Distribuição Binomial

$$P_x = \binom{n}{x} p^x q^{n-x}$$

- P = binomial probability
- x = number of times for a specific outcome within n trials
- $\binom{n}{x}$ = number of combinations
- p = probability of success on a single trial
- q = probability of failure on a single trial
- n = number of trials

Distribuição Binomial



Exemplo de Estatísticas

- Bioestat 2
- Mediana (divide distribuição em duas)
- Moda (frequência ou valor mais abundante)
- Variância e desvio padrão (medida de dispersão da pop)
- Erro padrão da média (medida de dispersão na estimativas da média)
- Intervalo de confiança - intervalo que cobre x% da distribuição

Exemplo de Métodos

- Teste t
- Análise de variância
- Regressão
- Teste de Fisher 2 x2
- Mann-Whitney comparação de ranking
- Kolmogorov-Smirnov comparação de distribuições