

Metodologia Ecológica

Aula 3 - Elementos de Modelagem Estatística

Problemas com o uso corrente da estatística em Ecologia:

- Ritual estatístico de “achar o teste correto”
- Pressupostos de “testes”

Ignorância filosófica e conceitual

A realidade

- Inferência estatística é baseada em modelos
(os tais “testes” são modelos pré-definidos)
- Dados e objetivos distintos requerem abordagens distintas, modelos distintos

O que é feito quando se usa um teste “enlatado”

- Um modelo único, pré-definido pelo teste, é escolhido. Qual a hipótese sendo testada?
- Dados são coletados apropriados (ou não) ao modelo sendo testado
- Se os dados não se ajustam bem ao modelo de teste?
- Há alguma medida de ajuste do modelo aos dados?

O que é preciso ser feito

- Traduzir hipóteses em modelos
- Coletar dados apropriados aos modelos
- Ajustar modelos aos dados
- Comparar as previsões dos modelos

O que é um modelo?

“Todos os modelos estão errados, mas alguns são úteis” (Box 1976)

Modelos matemáticos: $y = b * x$

Exemplo: Faturamento em loja de sorvete

Cada sorvete = 2 reais

Se vendeu 3 sorvetes, faturamento = 6 reais

Se vendeu 7 sorvetes, faturamento = 14 reais

Se vendeu 13 sorvetes, faturamento = 26 reais

Generalizando: $Y = 2 * X$

O que é um modelo estatístico?

Modelos estatísticos: $y = b \cdot x + \varepsilon$

Cada sorvete = *em média* 2 reais (depende do freguês!)

Se vendeu 3 sorvetes, faturamento = *em média* 6 reais

Se vendeu 7 sorvetes, faturamento = *em média* 14 reais

Se vendeu 13 sorvetes, faturamento = *em média* 26 reais

Generalizando: $Y = 2 \cdot X + \varepsilon$

Passos para construção de um modelo estatístico

- O quê se quer estimar?
 - O número de espécies em função da área
- Defina um modelo:
 - $\text{LogEspécies} = z * \text{LogÁrea} +$
- Defina uma função de densidade/distribuição de probabilidade para a variação residual

Modelos (estatísticos) preferidos

- Devem ser parcimoniosos. Ou seja, preferimos:
- Modelos com $n-1$ parâmetros em relação a outro com n parâmetros
- Modelos com $k-1$ variáveis explanatórias em relação a outro com k var.
- Modelos lineares em relação a modelos que sejam curvos
- Modelos sem “corcova” em relação a modelos com “corcova”
- Modelos sem interação em relação a modelos com interação

Relação
espécies-
área
(Preston
1962)

Iha	Area	Nespecies	LogArea	LogEspecies
Albemarle	5824.9	325	3.765	2.512
Charles	165.8	319	2.219	2.504
Chatham	505.1	306	2.703	2.486
James	525.8	224	2.721	2.350
Indefatigable	1007.5	193	3.003	2.286
Abingdon	51.8	119	1.714	2.076
Duncan	18.4	103	1.265	2.013
Narborough	634.6	80	2.802	1.903
Hood	46.6	79	1.669	1.898
Seymour	2.6	52	0.413	1.716
Barrington	19.4	48	1.288	1.681
Gardner	0.5	48	-0.286	1.681
Bindloe	116.6	47	2.067	1.672
Jervis	4.8	42	0.685	1.623
Tower	11.4	22	1.057	1.342
Wenman	4.7	14	0.669	1.146
Culpepper	2.3	7	0.368	0.845

Definindo uma linha reta e seus dois parâmetros:
inclinação e intercepto

Fig. 9.1

Interpolação e Extrapolação: Fig. 9.2

Ajustando dados a um modelo linear:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$d_i = (Y_i - \hat{Y}_i)^2$$

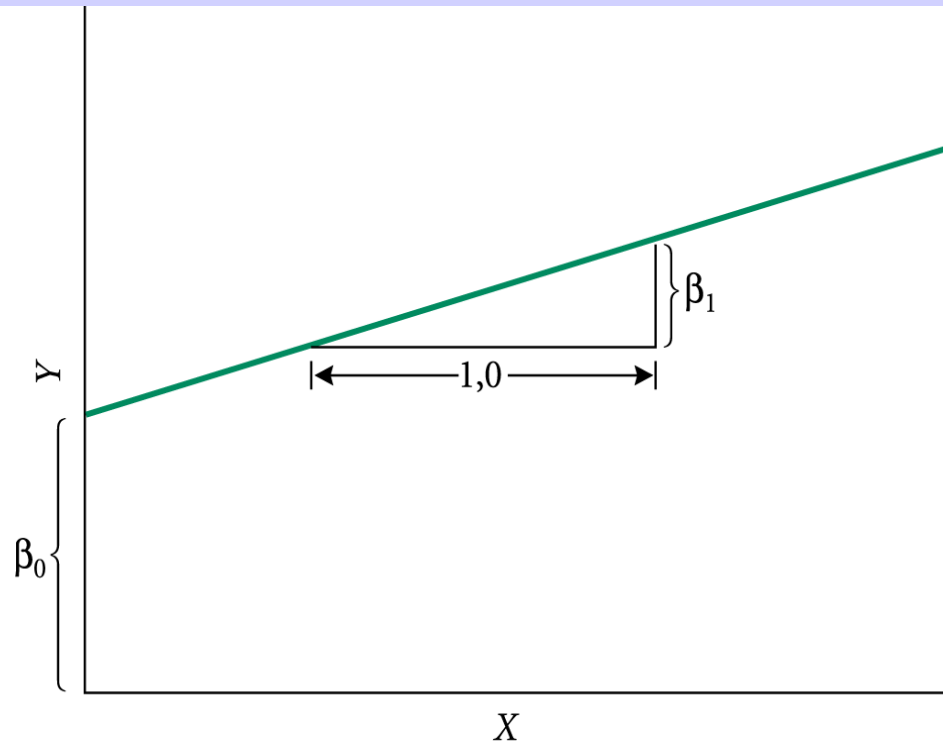


Figura 9.1 Relação linear entre as variáveis X e Y . A linha é descrita pela equação $Y = \beta_0 + \beta_1 X$, onde β_0 é o intercepto e β_1 é a inclinação da linha. O intercepto β_0 é o valor predito da equação quando $X = 0$. A inclinação da linha β_1 é o aumento na variável Y associado com o de uma unidade da variável X ($\Delta Y / \Delta X$). Se o valor de X é conhecido, o valor predito de Y pode ser calculado multiplicando X pela inclinação e somando o intercepto (β_0).

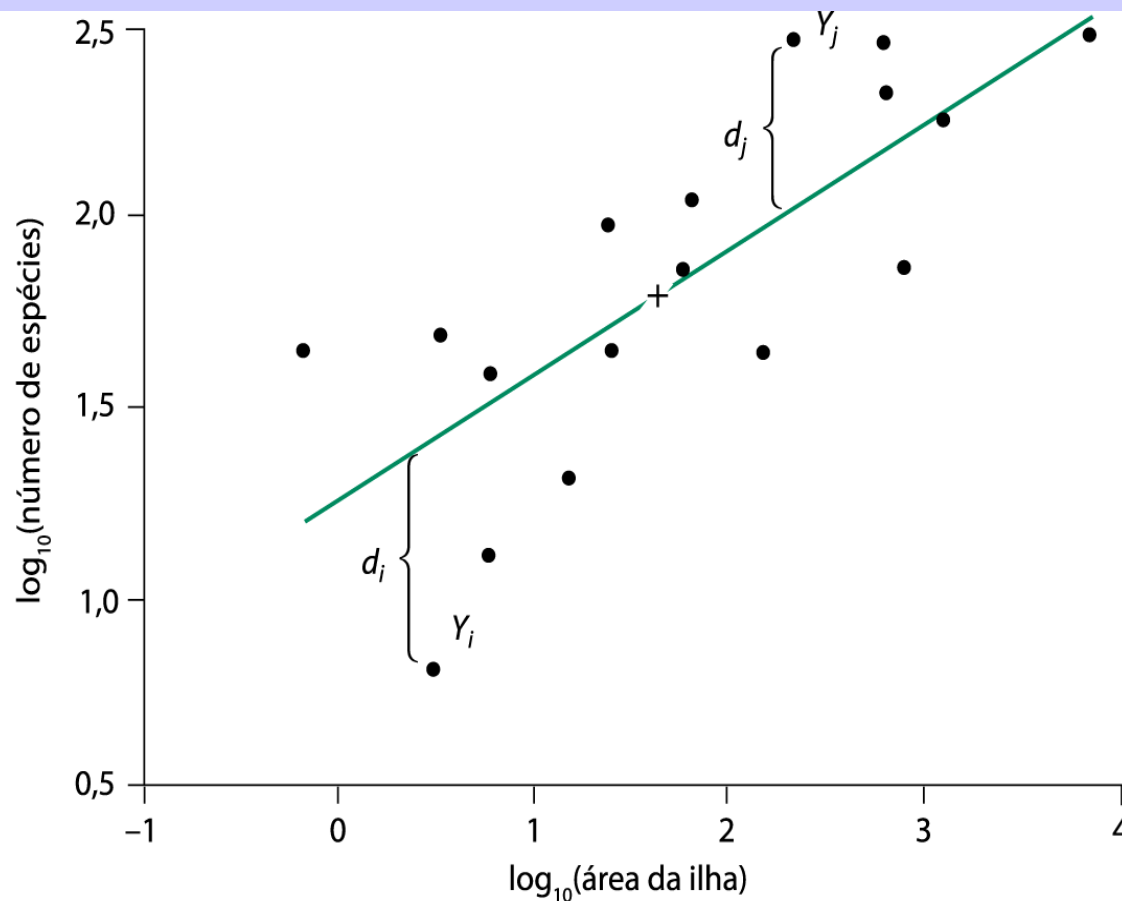


Figura 9.3 A soma dos quadrados dos resíduos é obtida somando os desvios quadrados (d_i) de cada observação da linha de regressão ajustada. A estimativa do parâmetro dos mínimos quadrados garante que essa linha de regressão minimize a soma dos quadrados dos resíduos. O + marca o ponto central dos dados (\bar{X} , \bar{Y}). Essa linha de regressão descreve, ainda, a relação entre o logaritmo da área das ilhas e o do número de espécies de plantas das Ilhas de Galápagos, dados da Tabela 8.2. A equação da regressão é $\log_{10}(\text{espécies}) = 1,320 + \log_{10}(\text{área}) \times 0,331$; $r^2 = 0,584$.

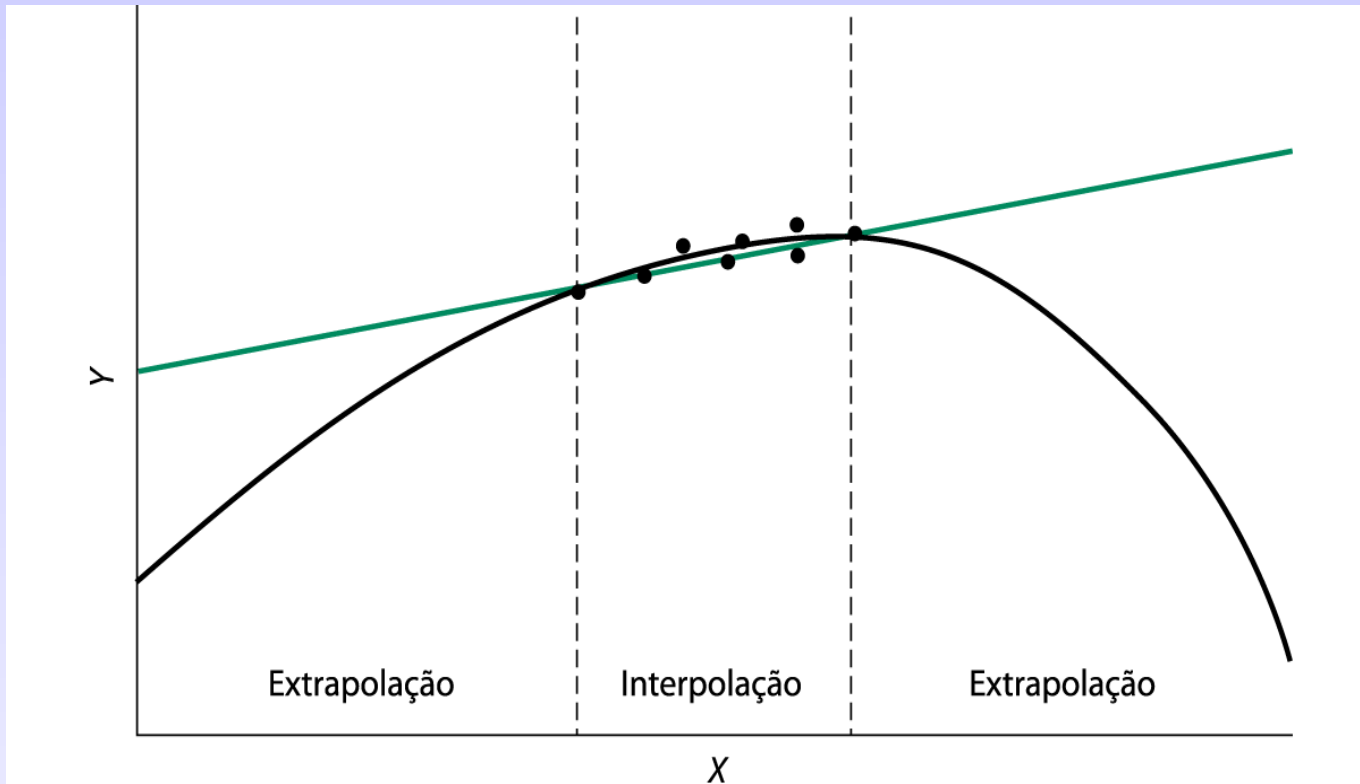


Figura 9.2 Modelos lineares podem aproximar funções não lineares sobre um domínio limitado da variável X . A interpolação dentro desses limites pode ser aceitável e acurada, embora o modelo linear (linha verde) não descreva a verdadeira relação funcional entre Y e X (curva preta). A extrapolação se tornará crescentemente menos acurada conforme as previsões se movem para além da amplitude dos dados coletados. Uma premissa muito importante da regressão linear é que a relação entre X e Y (ou transformações dessas variáveis) é linear.

O significado de uma “variável” ser dita “aleatória”

- Definimos o espaço amostral $\Omega = \{(\text{captura}), (\text{fuga})\}$ para os resultados possíveis dos evento visita de um inseto a uma planta carnívora.
- Para analisar estatisticamente esta informação, precisamos associar a cada elemento do espaço amostral um número.
- Precisamos de uma **função** que atribua um valor a cada elemento do espaço amostral.
- A esta função é dado o nome **variável aleatória**, normalmente representada por letras maiúsculas, como X .
- Como seria uma variável aleatória para Ω ?

Variáveis aleatórias (ii)

- Assim a variável aleatória na verdade é uma função cujos valores não são necessariamente aleatórios (!?).
- É dita “aleatória” porque os valores de X dependem do resultado do experimento, que tem um grau de incerteza nos resultados, de aleatoriedade.

Variáveis aleatórias (iii)

- Variáveis aleatórias podem ser discretas ou contínuas.
 - Variáveis aleatórias discretas têm um número finito de valores dentro de um intervalo.
 - Variáveis aleatórias contínuas podem* ter um número infinito de valores dentro de um intervalo.
- * Na prática o número de valores é limitado pela precisão do método ou instrumento.



Exemplo: amostragem de uma planta rara *Rhexia mexicana* em 349 localidades

- Como é uma planta rara, vamos supor que a probabilidade de *Rhexia* estar presente é baixa, 0,02, isto é
 - $P(X = 1) = p = 0,02$.
 - $X = 1$ *Rhexia* presente
 - e $P(X = 0) = ?$
 - $P(X = 0) = (1-p) = 0,98$

$X = 0$ *Rhexia* ausente



- Qual a probabilidade de *Rhexia* ocorrer em uma cidade E em nenhuma outra?

Em uma cidade: $p = 0,02$

Não ocorrer em cada uma das 348 cidades restantes:

$$\begin{aligned}(1-p)^{348} &= 0,98^{348} \\ &= 0,02 \times 0,98^{348} \\ &= 0,02 \times 0,00084452 = 0,00001769\end{aligned}$$

- Qual a probabilidade de *Rhexia* ocorrer em 10 localidades específicas E em nenhuma outra?

$$\begin{aligned}p^{10} \times (1-p)^{339} &= 0,02^{10} \times 0,98^{339} \\ &= (1,024 \times 10^{-17}) \times 0,00106 = 1,086 \times 10^{-20}\end{aligned}$$

Variável aleatória binomial

- Em ciência raramente será feita apenas um experimento. Sempre haverão réplicas dentro do mesmo experimento.
- O número de resultados com sucesso em n tentativas de Bernoulli caracteriza uma **variável aleatória binomial**.
- $X \sim \text{Bin}(n, p)$
 - Traduzir



- Qual a probabilidade de *Rhexia* ocorrer em 10 localidades QUAISQUER?
- Evento complexo: existem vários resultados que dariam este evento:
- (a, b, c, d, e, f, g, h, i, j) ou (a, b, c, d, e, f, g, h, k) ou (a, b, c, d, e, f, g, h, l), etc.
- Quantas combinações de 10 localidades podem haver?

Variável aleatória binomial (ii)

- A probabilidade de obter X sucessos em n tentativas em uma variável aleatória binomial é

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X}$$

ou

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

pX = probabilidade de X sucessos independentes, cada um com probabilidade p

$(1-p)^{n-X}$ = probabilidade de $n - X$ insucessos, cada um probabilidade $1-p$

- E o termo $\frac{n!}{X!(n-X)!}$ ou $\binom{n}{X}$?

– Coeficiente binomial

- 10 localidades com a planta em 349 é a probabilidade de obter 10 sucessos e 339 insucessos.
- Começando a procura por *Rhexia*, existem 349 localidades onde pode ocorrer
- Então inicialmente existem 349 combinações de localidades onde você poderia ter apenas uma ocorrência, e 348 combinações onde poderia ter a segunda ocorrência.
- Assim, o número de combinações que 10 *Rhexia* poderiam estar distribuídas entre as 349 localidades seria
- $349 \times 348 \times 347 \times 346 \times 345 \times 344 \times 343 \times 342 \times 341 \times 340 = 2,35 \times 10^{25}$
- $= n \times (n-1) \times (n-2) \times \dots \times (n-9)$

- Estão ausentes $(n-X) \times (n-X-1) \times (n-X-2) \times \dots \times 1$
 $= (n-X)!$
- Então $n! / (n-X)!$ temos o número total de combinações obtendo X sucessos em n tentativas.
- Mas falta ainda explicar a divisão por $X!$
- É que não queremos contar como dois resultados distintos a ocorrência de *Rhexia* na localidade A e depois B, e primeiro na localidade B e depois A.
- Para cada combinação possível de X tentativas, existem sempre X resultados em que a diferença é apenas na ordem e não na composição.
- Estes têm que ser então descontados do total (são **permutações**).

- Podemos gerar então uma distribuição binomial, um histograma contendo a frequência de cada um dos possíveis resultados.



Variáveis Aleatórias de Poisson

- Quando n é muito grande e p muito pequeno, a distribuição binomial torna-se difícil de ser aplicada. É o caso da distribuição de plantas e animais raros.
- Também requer saber n , o número de tentativas, o que nem sempre é conhecido.
- Exemplo: dispersão de sementes
- Amostragem no tempo também pode ter estas características.
- Nestes casos usamos a **distribuição de Poisson** (Siméon-Denis Poisson, 1781-1840).



- Uma **variável aleatória de Poisson** é o número de ocorrências de um evento em uma amostra de tamanho fixo. Em geral este número é baixo e a maioria das amostras = 0.
- Caracterizada por um único parâmetro, λ , o número médio de ocorrências do evento em cada amostra
- $X \sim \text{Poisson}(\lambda)$

Função de distribuição de probabilidades de Poisson

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$$\lim_{n \rightarrow \infty} \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} = \frac{(np)^X e^{-(np)}}{X!}$$

- Fazendo $P(X) = \frac{m^X e^{-m}}{X!}$

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

- O número médio de plantas jovens de orquídeas encontradas em um quadrat de 1m² é 0,75
- Quais as chances de que um único quadrat tenha 4 plantas jovens?

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$P(4 \text{ plântulas}) = \frac{0,75^4}{4!} e^{-0,75} = 0,0062$$

Exemplos de distribuição de Poisson

- primeiras gotas de chuva sobre uma rodovia
- erros de impressão em um livro
- itens defeituosos em uma indústria
- Outros?
- A distribuição de Poisson é simétrica?

Valor Esperado

- Se a variável aleatória X pode ter os valores a_1, a_2, \dots, a_n , o valor esperado de X é

$$E(X) = \sum_{i=1}^n a_i p_i$$

- Assim, $[(0 \times 0,1) + (50 \times 0,9)] = 45$
- A variância de uma variável aleatória discreta é tão importante quanto a média

- Apenas a média não caracteriza e distingue uma distribuição

- $\sigma^2 (X) = E[X - E(X)]^2$
$$\sigma^2 = \sum_{i=1}^n p_i \left(a_i - \sum_{i=1}^n a_i p_i \right)^2$$

- O desvio de cada valor i é ponderado pela sua probabilidade de ocorrência, p_i
- Por que elevar ao quadrado os desvios do valor esperado?

Valores esperados e variâncias das distribuições Bernoulli, Binomial e Poisson

- Tabela 2.3 Gotelli & Ellison

Variáveis aleatórias contínuas

- Qual o espaço amostral?
- Ainda pode ser definido, mas não tem mais resultados discretos
- Os eventos são definidos pelo subintervalo onde ocorrem

Variável aleatória uniforme

- A probabilidade em subintervalos de igual tamanho é a mesma
- A probabilidade de ocorrer em dois intervalos quaisquer é a soma das probabilidades de cada um
- Esta função é chamada de uma **função de densidade de probabilidade** (*probability density function*)

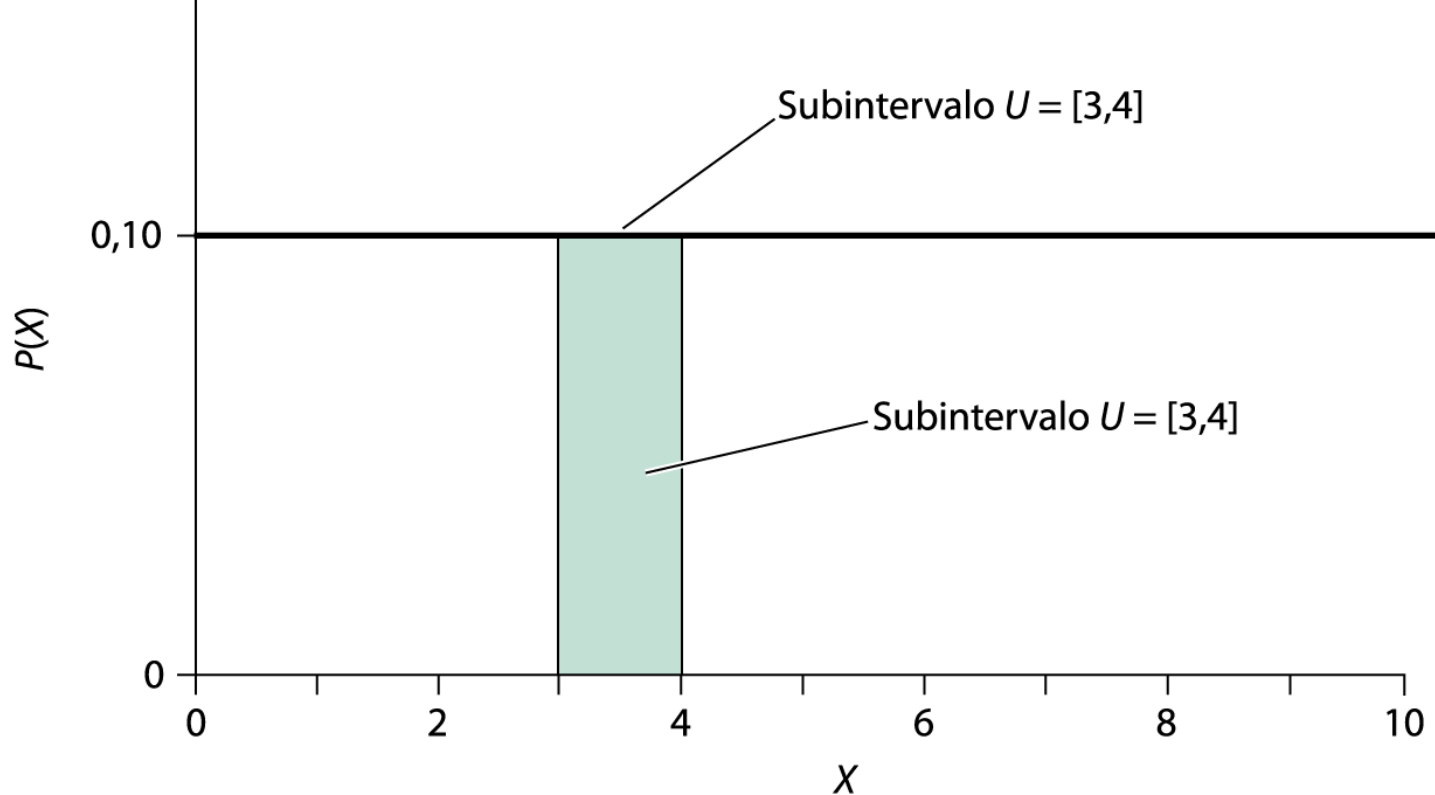


Figura 2.4 Distribuição uniforme com intervalo $[0,10]$. Em uma distribuição uniforme contínua, a probabilidade de um evento ocorrer em um subintervalo particular depende da área relativa do subintervalo; ela é a mesma independente de onde o subintervalo está inserido dentro dos limites da distribuição. Por exemplo, se a distribuição é delimitada por 0 e 10, a probabilidade de que um evento ocorra no subintervalo $[3,4]$ é a área relativa delimitada por aquele subintervalo, que neste caso é 0,10. A probabilidade é a mesma para qualquer outro subintervalo com o mesmo tamanho, como $[1,2]$ ou $[4,5]$. Se o subintervalo escolhido é maior, a probabilidade de um evento ocorrer naquele subintervalo será proporcionalmente maior. Por exemplo, a probabilidade de um evento ocorrer no subintervalo $[3,5]$ é de 0,20 (desde que 2 dentre as 10 unidades do intervalo sejam transpassadas), e é de 0,6 para o subintervalo $[2,8]$.

- De forma semelhante ao caso discreto, a esperança será a soma do valor de cada observação ponderada pela sua probabilidade, assim será

$$\sum_{i=1}^n x_i f(x_i) \Delta x$$

- Se fizermos Δx ficar cada vez menor, o somatório anterior se aproximará um valor único, no limite de $\Delta x \rightarrow 0$.
- ou melhor dizendo

$$E(X) = \int x f(x) dx$$

Variáveis aleatórias normais

- Figura 2.6
- Às vezes chamadas também de variáveis aleatórias Gaussianas, ou de Movre-Gauss-Laplace.
- Propriedades úteis da distribuição normal:
 1. Distribuições normais podem ser somadas
- $E(X + Y) = E(X) + E(Y)$
- $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$

2. São facilmente transformáveis por operações de deslocamento e mudança de escala.
- Multiplicando X por uma constante como a é uma mudança de escala porque uma unidade de X torna-se a unidades de Y .
 - Figura 2.7

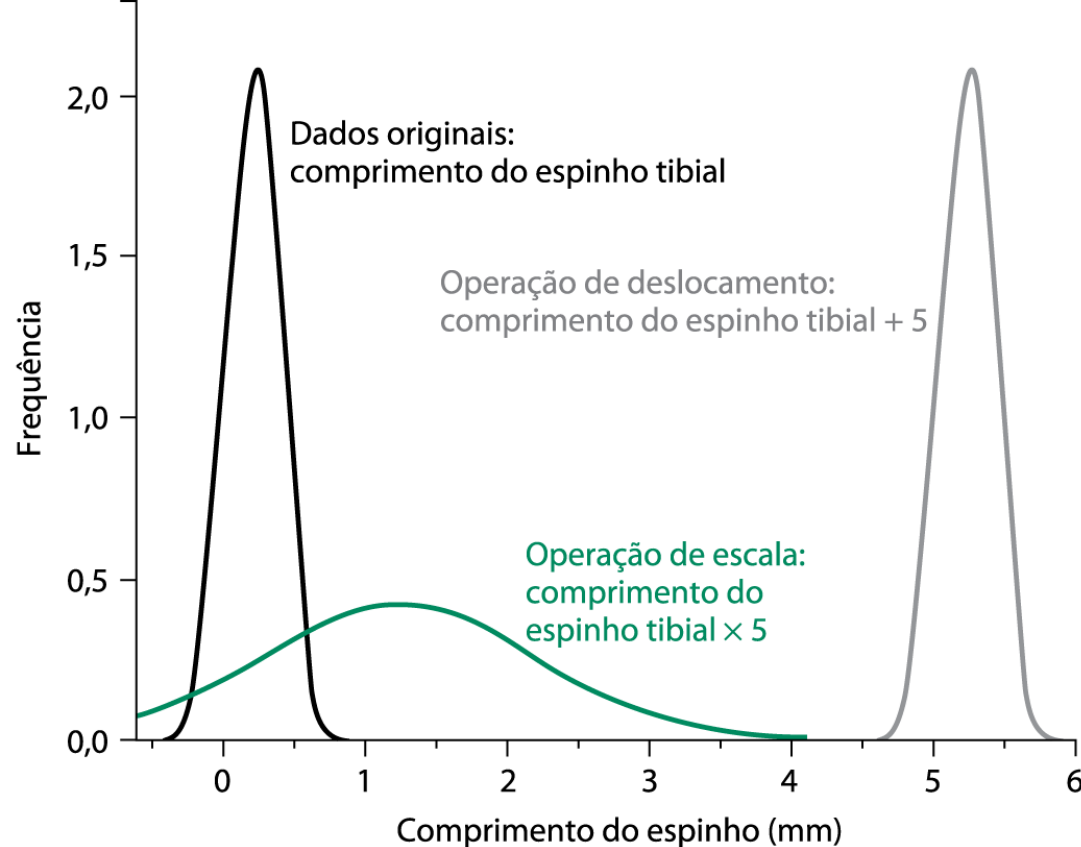


Figura 2.7 Operações de deslocamento e escala sobre uma distribuição normal. A distribuição normal possui duas propriedades algébricas convenientes. A primeira é uma operação de deslocamento: se a constante b é adicionada a um conjunto de medidas com média μ , a média da nova distribuição será deslocada para $\mu + b$, mas a variância não é afetada. A curva preta é o ajuste da distribuição normal a um conjunto de 200 medidas do comprimento do espinho tibial de aranhas (Figura 2.6). A curva cinza mostra a distribuição normal deslocada após o valor 5 ser acrescido a cada uma das observações originais. A média se deslocou 5 unidades para a direita, mas a variância não é alterada. Em uma operação de escala (curva verde), multiplicar cada observação por uma constante a causa um acréscimo na média por um fator de a^2 . Esta curva é o ajuste da distribuição normal aos dados depois de terem sido multiplicados por 5. A média é deslocada para um valor 5 vezes maior que o original, e a variância aumenta por um fator de $5^2 = 25$.

3. No caso especial em que o deslocamento

$b = -1(\mu/\sigma)$ e a mudança de escala é

$a = 1/\sigma$, temos

$$Y = (1/\sigma)X - \mu/\sigma = (X - \mu)/\sigma$$

- $E(Y) = 0$ e $\sigma^2(Y) = 1$

=> Variável aleatória normal padrão

Variáveis aleatórias log-normais

- A distribuição pode refletir processos de crescimento exponencial sujeitos a muitos fatores que agem independentemente, como populações biológicas.
- Ex.: Distribuição de abundâncias relativas de espécies em uma comunidade ou montagem.
- mas também a distribuição de riqueza econômica entre os países, de “sobrevivência” de classes de tempo de bebida em um restaurante movimentado

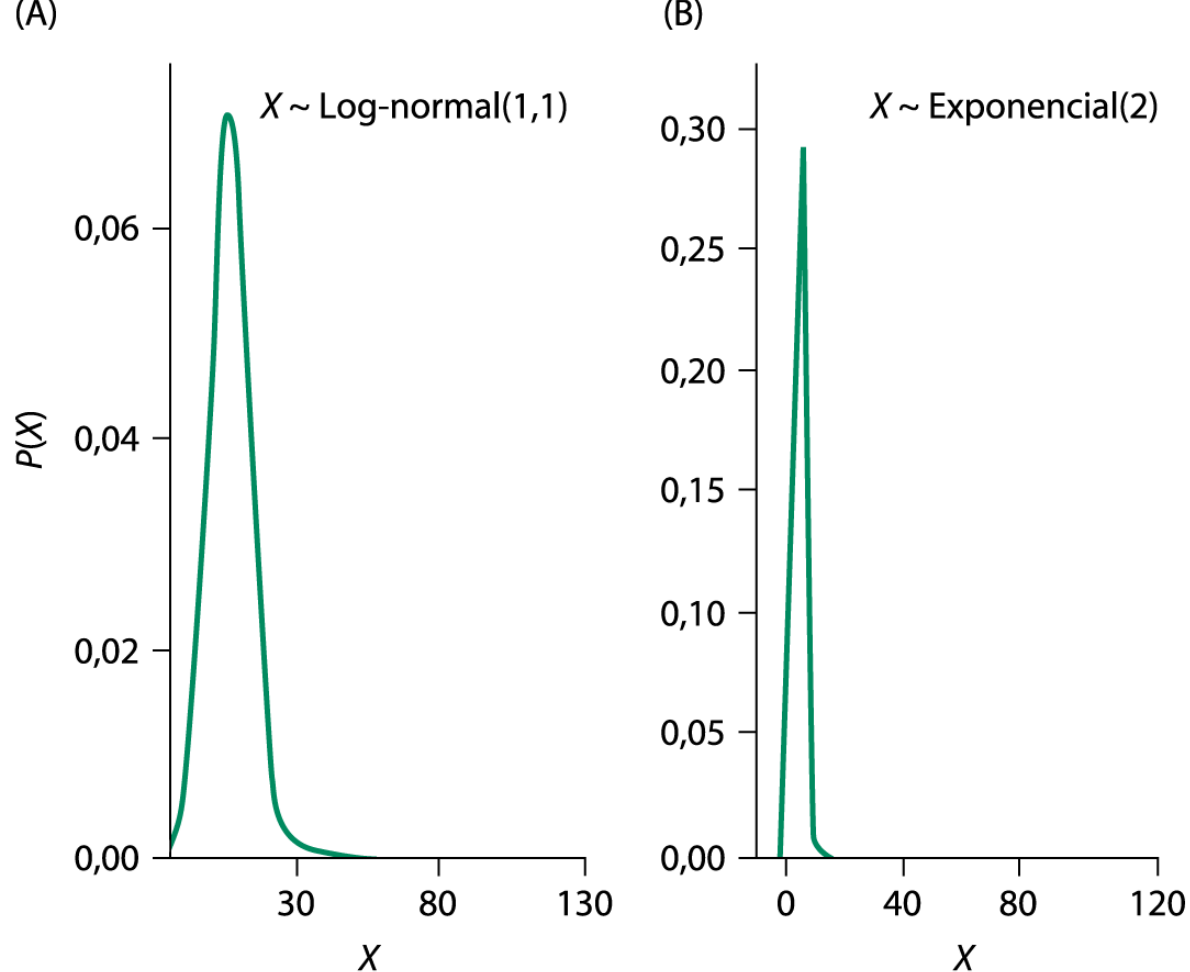


Figura 2.8 Distribuições log-normal e exponencial se ajustam a certos tipos de dados ecológicos, como a distribuição de abundâncias de espécies e distâncias de dispersão de sementes. (A) A distribuição log-normal é descrita por dois parâmetros, média e variância, ambos são 1 neste exemplo. (B) A distribuição exponencial é descrita por um único parâmetro b , que é 2 nesse exemplo. Ver a Tabela 2.4 para as equações usadas com as distribuições log-normal e exponencial. Ambas as distribuições, log-normal e exponencial, são assimétricas, com uma longa cauda a direita que desvia a distribuição à direita.

Variáveis aleatórias exponenciais

- Os espaços entre variáveis de Poisson (discretas) são por sua vez contínuos, como a distância entre plantas jovens nos quadrats, ou o tempo entre amostragens.
- Tabela 2.4

Soma dos quadrados dos resíduos
(=Residual Sum of Squares, RSS):

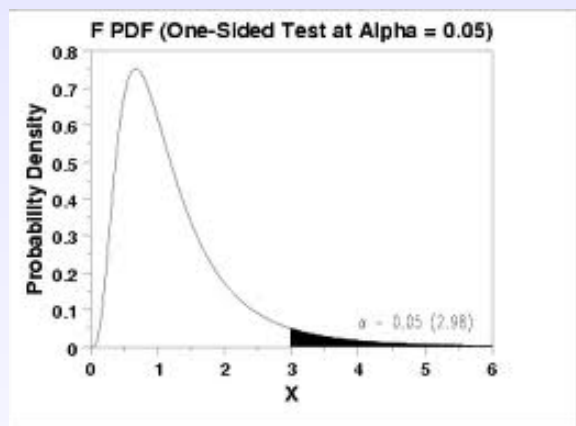
$$SQR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$SQ_Y = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$s^2_Y = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

TABLE 9.1 Complete ANOVA table for single factor linear regression

Source	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	Expected mean square	F-ratio	P-value
Regression	1	$SS_{reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$\frac{SS_{reg}}{1}$	$\sigma^2 + \beta_1^2 \sum_{i=1}^n X^2$	$\frac{SS_{reg}/1}{RSS/(n-2)}$	Tail of the F distribution with 1, $n-2$ degrees of freedom
Residual	$n-2$	$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$\frac{RSS}{(n-2)}$	σ^2		
Total	$n-1$	$SS_Y = \sum_{i=1}^n (Y_i - \bar{Y})^2$	$\frac{SS_Y}{(n-1)}$	σ_Y^2		



Métodos científicos: combinações de raciocínio indutivo e dedutivo

- O que é um método científico?
 - Um método que compare e decida entre hipóteses (não teorias!) baseando-se em observações e previsões baseadas nas hipóteses.
 - Não há apenas um único método de fazer isso.

- Dedução: parte do geral para o específico.
 - Em ciência, parte afirmações ou teorias gerais para formular hipóteses específicas para uma situação.



- Indução: parte do específico para o geral.
- Na sua forma mais simples, a partir do acúmulo de observações em casos específicos formula-se uma hipótese geral para explicar os dados.
- Seria o Inducionismo de Francis Bacon (1561-1626).

- Dedução:

1. Todas as aranhas papa-moscas e saltadoras pertencem à família Salticidae.
2. Coletei uma aranha saltadora.
3. Esta aranha é da família Salticidae.

- Indução:

1. Todas estas 25 aranhas são da família Salticidae.
2. Todas estas 25 aranhas são papa-moscas e saltadoras.
3. Todas as aranhas papa-moscas e saltadoras pertencem à família Salticidae.

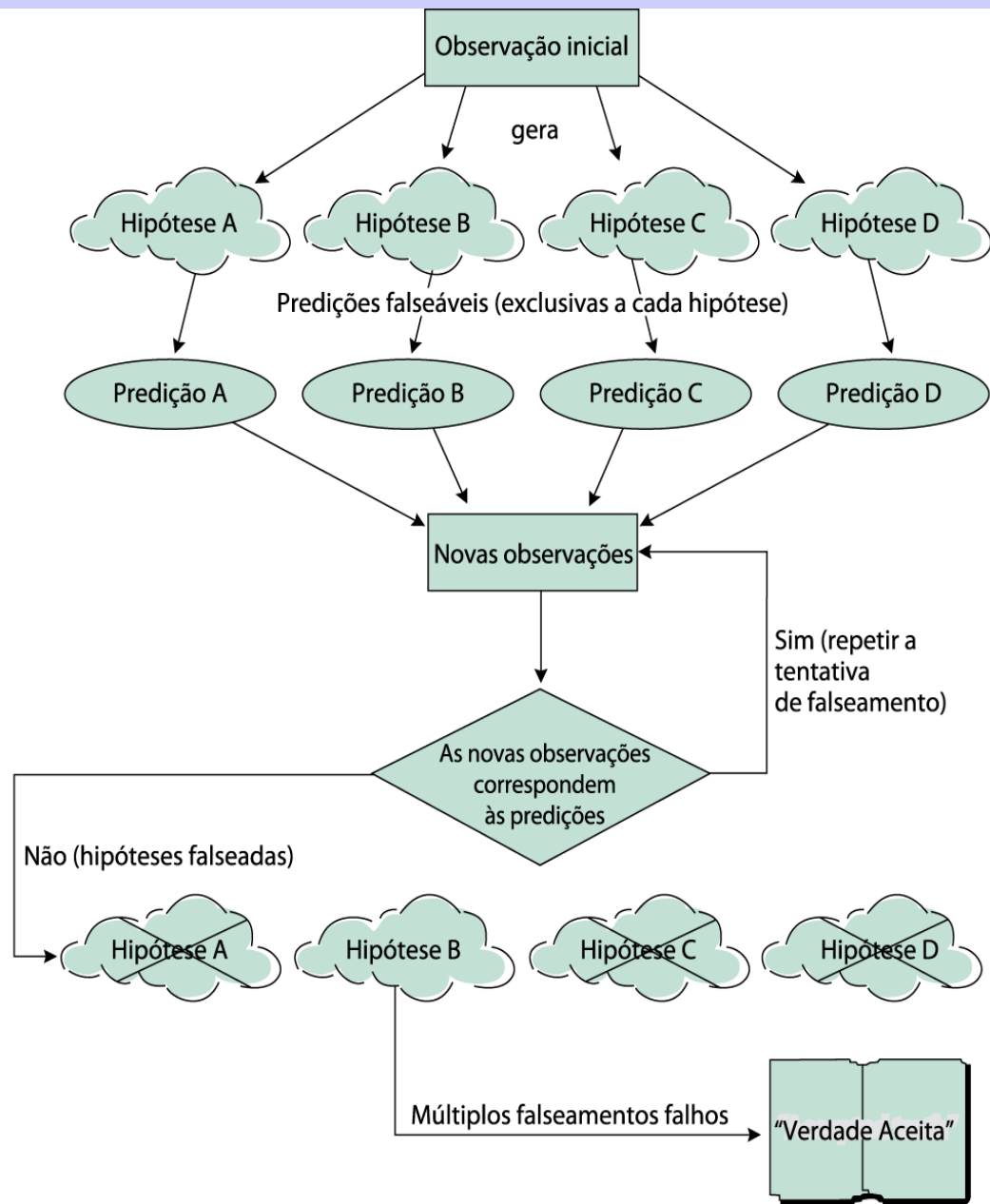
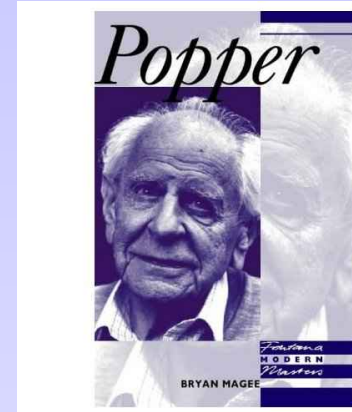


Figura 4.4 O método hipotético-dedutivo. Hipóteses múltiplas de trabalho são propostas e suas predições são testadas com o objetivo de falsear as incorretas. A explicação correta é aquela que se mantém depois de repetidos testes que falham em falseá-la.



O método hipotético-dedutivo de Karl Popper



- Ênfase na dedução como método científico se inicia partir de Isaac Newton e outros cientistas do séc. XVII
- Atinge seu maior status atual com os trabalhos do filósofo Karl Popper e do “círculo de Vienna” em torno de 1935, formulando o método hipotético-dedutivo.
- Talvez a maior inovação de Popper tenha sido a formulação de hipóteses múltiplas como passo inicial da investigação científica.
- Estas hipóteses seriam todas elaboradas para explicar observações iniciais, sendo testáveis e gerando previsões distintas umas das outras.
- Haveria então tentativas de falsificá-las através de experimentos que permitiriam testar suas previsões.
- Mas para cada hipótese, um teste específico, uma hipótese nula

Quais as limitações dos métodos indutivo e hipotético-dedutivo?

- É preciso haver uma hipótese “correta” entre as alternativas, isto é, uma não-falsificável. No método indutivo pode-se começar com hipóteses incorretas, que serão alteradas com novas observações.
- As hipóteses têm efeitos distintos. Não é possível testar hipóteses como efeitos redundantes em algum grau. Com a indução é possível incorporar efeitos múltiplos numa hipótese complexa.

- O teste de hipóteses a partir da tentativa de falsificação de hipóteses nulas é a implementação do método hipotético-dedutivo de Popper.
- De forma geral, a hipótese nula é uma hipótese estatística, que tentamos refutar, ou melhor, falsificar. Idealmente, deve haver apenas uma hipótese alternativa, que seria então aceita.
- Se houver mais de uma hipótese, estas devem fazer previsões distintas sobre os resultados, permitindo falsificar todas menos a “verdadeira”.

Testes de hipóteses



Estimativa de parâmetros

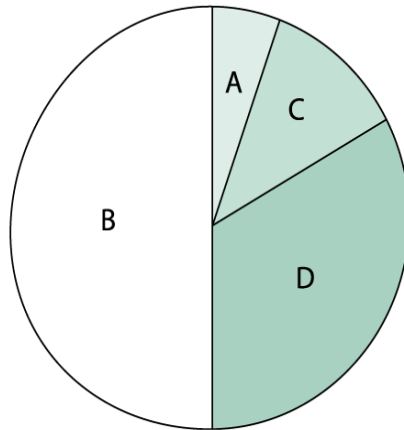


Figura 4.6 Teste de hipóteses *versus* estimativa de parâmetros. A estimativa de parâmetros acomoda com mais facilidade mecanismos múltiplos e pode permitir uma estimativa da importância relativa dos diferentes fatores. A estimativa de parâmetros pode envolver a construção de intervalos de credibilidade (ver Capítulo 3) para estimar a força de um efeito. Uma técnica relacionada, na análise de variância, é decompor a variação total dos dados em proporções que são explicadas por diferentes fatores no modelo (ver Capítulo 10). Ambos os métodos quantificam a importância relativa de diferentes fatores, enquanto o teste de hipóteses enfatiza uma decisão binária de sim/não sobre se um fator tem um efeito mensurável ou não.

Arcabouços para análise estatística

Frequentista:

Paramétrico

Não-paramétrico: Aleatorização e Monte Carlo

Seleção de modelos

Bayesiana



<http://www.rasch.org/rmt/rmt1237.htm>

Suponha que em 5 jogadas da moeda foram obtidas 5 caras. Qual o deve ser o viés da moeda?

- Para estimarmos o viés da moeda, olhamos a probabilidade de obter 5 caras em todas as hipóteses. A hipótese onde esta frequência é a mais provável fornece uma estimativa do viés.
- No caso, 0,9 é a estimativa mais próxima do viés.
- O somatório dos valores de uma coluna é sempre 1,0. São as probabilidades dos resultados em cada hipótese, cada coluna uma hipótese.
- Como o somatório dos valores de uma linha é sempre diferente de 1, foi necessário diferenciar estes valores de probabilidades, adotando-se o termo verossimilhança (*likelihood*).