

# Utilidade para testes de significância

Nathália Demetrio Vasconcelos Moura

DISSERTAÇÃO APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
MESTRE

Programa: Estatística

Orientador: Prof. Dr. Sergio Wechsler

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, maio de 2014

## Utilidade para testes de significância

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 16/05/2014. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Sergio Wechsler (orientador) - IME-USP
- Prof. Dr. Carlos Alberto de Bragança Pereira - IME-USP
- Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Regina Madruga Tavares - UFPA

# Agradecimentos

Registrar a gratidão que sinto por tantos... Cada contribuição à sua maneira... Fico feliz em tentar...

À minha mãe, que está comigo desde sempre, mostrando, por exemplos, o quanto tudo é possível. Muito obrigada por me ensinar a encarar a vida de maneira otimista, humilde e independente.

Ao meu irmão Paulinho, presente nos meus melhores momentos e sentimentos. Aos Demétrios, sempre ao meu lado das mais diferentes formas. Ao meu avô e Smurff (Muff para íntimos), que me permitiram conhecer as formas mais puras de amor. Ao Gustavo, vulgo Guga, que contempla todas as opções anteriores.

A Paulo Moura, pessoa em que as definições de carinho, amizade, admiração e amor sempre se confundem. Como você bem já sabe: “Nós dois fomos feitos muito pra nós dois”.

A todos os meus amigos, por floream minha vida. Em especial a Roberto Mangui, pelo apoio incondicional, bem como à nossa japa preferida, Paula Shinozaki. A Andrés Arias e Felipe Behring, por conversas, filmes e madrugadas, mas principalmente pela amizade, que me sinto extremamente feliz de saber que é recíproca. As “meninas do CRUSP”, que transformaram alguns metros quadrados em um amado e barulhento lar, e a Lina Thomas, que deu continuidade a este lar, em uma metragem um pouco maior.

A todas as pessoas maravilhosas que tornaram a experiência do mestrado uma etapa tão nostálgica: colegas do curso de verão, disciplinas, churrascos e cafés. Aos funcionários do IME-USP, sempre prestativos e pacientes. Aos professores da UFPE, por me permitirem viver esta experiência. Aos professores do IME-USP, pela dedicação nos ensinamentos. Particularmente ao Prof. Luis Gustavo, pelas melhores aulas, e ao Prof. Carlinhos, pelo melhor exemplo. Aos educadores, exemplos e amigos: Alexandre Patriota e Victor Fossaluzza, muito obrigada pelo imenso aprendizado que me permitiram, por tantos caminhos diferentes, sinto-me honrada de tratar vocês pelo primeiro nome.

E acima de qualquer pessoa, entidade, ou agremiação alvinegra, a Sergio Wechsler. Apoiou-me de forma incondicional durante todas as curvas não previstas que, diga-se de passagem, não foram poucas. Sempre fazendo do relacionamento horizontal a base, e dos questionamentos os pilares. Um orientador no sentido mais bonito que se pode atribuir à palavra. Se nosso objetivo em vida for deixar uma marca no mundo, você sem dúvida o fez, pelo meio mais admirável: pessoas! E por isso serei eternamente grata (tá sabendo né?).

“E a gente agradece, com gentileza e compaixão por todos os caminhantes, porque somente quem caminha sabe o valor, o tamanho, a conquista, de que é feita a história de cada único passo.”

Ana Jácomo



# Resumo

Moura, N. D. V. **Utilidade para testes de significância**. 2014. 41 f. Dissertação (Mestrado)- Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

Neste trabalho discutimos os principais argumentos da inferência bayesiana subjetivista. Posteriormente, a partir de uma revisão da literatura dos testes de hipóteses, os principais testes são analisados sob a ótica da teoria da decisão, particularmente no que tange às hipóteses precisas. Adicionalmente, funções de perda para testes de significância, seguindo a proposta de Fisher e do FBST, são analisadas e comparadas.

**Palavras-chave:** Bayes, FBST, Fisher, função de perda, Jeffreys, Popper, teoria da decisão, hipótese precisa, testes de significância.



# Abstract

Moura, N. D. V. **Utility for significance tests**. 2014. 41 f. Dissertação (Mestrado)- Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

This work discusses the main points of the bayesian subjectivist inference. Posteriorly, from a literature review of hypothesis testing, the main approaches are interpreting from a decision-theoretic viewpoint, particularly regarding the precise hypotheses. Additionally, loss functions for tests of significance, following the proposal of Fisher and FBST, are analyzed and compared.

**Keywords:** Bayes, decision theory, FBST, Fisher, Jeffreys, loss function, Popper, precise null hypothesis, significance tests.





# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Teoria da Decisão</b>	<b>3</b>
<b>3</b>	<b>Testes de Hipóteses</b>	<b>7</b>
3.1	Fisher & $p$ -valor . . . . .	7
3.2	Neyman-Pearson & Razão de Verossimilhanças . . . . .	9
3.3	Bayes & Medidas Condicionais . . . . .	10
<b>4</b>	<b>Hipóteses Precisas</b>	<b>13</b>
4.1	Testes Frequentistas . . . . .	13
4.2	Testes Bayesianos . . . . .	14
4.2.1	Jeffreys . . . . .	14
4.2.2	FBST . . . . .	14
<b>5</b>	<b>Funções de Perda</b>	<b>17</b>
5.1	Testes de Hipóteses . . . . .	18
5.1.1	Razão de Verossimilhanças . . . . .	18
5.1.2	Jeffreys . . . . .	19
5.2	Testes de Significância . . . . .	20
5.2.1	FBST . . . . .	20
<b>6</b>	<b>Outra Opção Genuinamente Bayesiana para Testes de Significância</b>	<b>23</b>
6.1	Popper & Rice . . . . .	23
6.2	FBST $\times$ Rice . . . . .	24
<b>7</b>	<b>Conclusão</b>	<b>27</b>
	<b>Referências Bibliográficas</b>	<b>29</b>



# Capítulo 1

## Introdução

Responder sobre fenômenos aleatórios a partir das informações disponíveis é o objetivo básico da Inferência Estatística. Para tal questão, é possível trabalhar com diferentes interpretações: verossimilhança, estrutural, *fuzzy*, entre outras. Entretanto, apesar da variedade de alternativas, a abordagem frequentista é a opção mais utilizada. Para esta leitura, a probabilidade de um evento é dada pelo limite das respectivas frequências relativas. Sendo tal limite representado por uma entidade denominada parâmetro, obtido segundo infinitas e hipotéticas repetições do experimento associado. Particularmente relevante, o parâmetro é responsável por especificar o comportamento da variável aleatória de referência. Contudo, ao encarar este limite como uma quantidade fixa, apesar de desconhecida, várias dificuldades são impostas às análises. Resumidamente:

- utilização da Lei dos Grandes Números como corroboração teórica para a definição de probabilidade. Ainda que a aplicabilidade deste teorema dependa da análise de uma sequência infinita de repetições do experimento, realizadas sob as mesmas condições. Circunstância pouco plausível em termos práticos;
- interpretação circular de probabilidade, uma vez que utiliza o conceito de independência, e este é definido a partir da própria probabilidade;
- distribuições de probabilidade vistas como universais (parâmetros não aleatórios). Situação esta que salvo alguns contextos “controlados”, como simulações ou jogos de azar, é pouco razoável;
- impossibilidade de modelar um experimento que foi interrompido, antes do previsto, por razões não definidas;
- violação do Princípio da Verossimilhança, apesar dos princípios da Condicionalidade e da Suficiência serem básicos<sup>1</sup>;
- uso indiscriminado da aleatorização, tornando a “moeda” responsável por decisões que caberiam a um analista racional;
- aceitação da existência e busca da objetividade, apesar do conceito por si só ser utópico. Além desta “premissa” ser ignorada na especificação do modelo, ou no desenho do experimento. Uma vez que tais etapas englobam a tradição, crença e experiência do analista. Sendo, portanto, uma escolha subjetiva (Berger e Berry, 1988); e
- análise da amostra, do conhecimento do especialista e das implicações práticas feitas separadamente. Inviabilizando a consideração do todo.

---

<sup>1</sup>{Princípio da Verossimilhança}  $\Leftrightarrow$  {Princípio da Condicionalidade}  $\wedge$  {Princípio da Suficiência} (Birnbaum, 1962)

Para contornar tais limitações, temos como opção ampliar a análise para a compreensão bayesiana. Nesta, ao encarar o parâmetro como uma entidade aleatória latente, obtemos uma leitura harmoniosa com o modo que a incerteza é intuída no dia a dia. Apesar das possíveis contradições lógicas, casualmente declaradas em situações informais (Mosteller e Youtz, 1990), os axiomas de coerência (Kadane, 2011) pressupostos para tal abordagem, além de implicarem no cálculo de probabilidades (conhecido como *Dutch Book Argument*), são: simples, interpretáveis e intuitivos (Berry, 1997):

“Inevitably, students (and everyone else except for statisticians [frequentistas] ) give an inverse or Bayesian twist to frequentist measures such as confidence intervals and p-values.” pág. 242.

No caso de acessarmos a informação disponível na amostra por meio da função de verossimilhança, teremos o teorema de Bayes especificando o mecanismo segundo o qual os dados são conectados ao conhecimento a priori,

$$\textit{probabilidade a posteriori} \propto \textit{probabilidade a priori} \times \textit{verossimilhança}.$$

A partir da ótica do movimento subjetivo bayesiano<sup>2</sup>, associado a DeFinetti, Savage, Lindley e DeGroot, dispomos não só da incorporação do conhecimento prévio à análise, mas também da possibilidade de modelar qualquer entidade desconhecida do ponto de vista probabilístico.

Existe também uma opção “objetiva” da inferência bayesiana [sic]. Nesta, apesar de utilizarmos o mesmo processo mecânico, não associamos subjetividade à priori. Sua especificação é feita segundo outros princípios, como histórico amostral, máxima entropia, invariância etc. Entretanto, uma vez que o conhecimento individual não é tratado como fonte de informação, temos como consequência a violação do princípio da verossimilhança e, por tal razão, não trataremos desta abordagem neste trabalho. É interessante ressaltar que apesar desta ser uma alternativa menos problemática do que o entendimento frequentista, optar por refletir sobre a análise, incorporando todas as fontes de informações existentes, ainda sim é a alternativa mais indicada .

No caso das evidências amostrais darem suporte explícito a decisões, é possível realizar uma análise ainda mais abrangente através da Teoria da Decisão Estatística. Nesta, incorporamos não apenas as informações relacionadas à incerteza, mas também a expectativa do pesquisador quanto aos resultados. Tal compreensão é de extrema importância prática, dado que, no “mundo real”, as evidências amostrais configuram apenas uma vertente a ser considerada, competindo com aspectos financeiros, políticos, processuais, etc.

Em termos práticos, tanto a procura por padrões quanto o levantamento de conjecturas se apresentam como necessidades naturais durante o estudo de eventos aleatórios. No contexto estatístico, tais pontos podem ser traduzidos segundo os Testes de Hipóteses. Devido à sua essência de verificação, além da ampla aplicabilidade, os testes de hipóteses configuram papel decisivo em muitas áreas do conhecimento. Tal fato, torna as discussões filosóficas associadas aos seus fundamentos ainda mais relevantes, particularmente no que diz respeito às hipóteses precisas, cenário igualmente praticado e debatido. Neste contexto, revisaremos a literatura dos testes de hipóteses, interpretando as principais abordagens sob a ótica da teoria da decisão, bem como as perdas associadas às hipóteses precisas. Tendo sempre, como base, o entendimento bayesiano subjetivista (perdão ao leitor pela redundância).

---

<sup>2</sup>As leis de probabilidade são a estrutura segundo a qual um indivíduo coerente deve expressar sua incerteza. Corroborada teoricamente pelo Teorema Fundamental de B. DeFinetti (Loschi e Wechsler, 2002).

## Capítulo 2

# Teoria da Decisão

A todo momento somos expostos a decisões, desde questões simples e individuais, como sair de casa levando, ou não, um guarda-chuva, até ações significativamente complexas, podendo influenciar toda uma sociedade, caso de uma medida política por exemplo. Logo, torna-se necessário dispormos de uma ferramenta que nos auxilie a optar pela melhor ação de modo racional, levando em consideração nossos objetivos, circunstâncias e conhecimento. Papel este plenamente executado pela **teoria da decisão**. Nesta, a ação a ser tomada admite valores no espaço de decisões  $\mathcal{D}$ , sendo influenciada pelos resultados de uma entidade que envolva incerteza. Tais resultados são definidos no conjunto  $\Omega$  e denominados “estados da natureza”. Assim, dadas as preferências do agente decisor em relação às possíveis consequências ( $\mathcal{C} = \mathcal{D} \times \Omega$ ), obtemos a escolha ótima segundo o cenário especificado pelo próprio agente.

Inicialmente, é necessário formalizar a ordenação das preferências do agente decisor<sup>1</sup> quanto às consequências. Para tal, utilizaremos a seguinte linguagem:

**Definição 1** *Para quaisquer duas consequências  $c, c' \in \mathcal{C}$*

- $c \prec c'$ :  $c$  é preferível a  $c'$ ;
- $c \succsim c'$ :  $c$  é pelo menos tão preferível quanto  $c'$ ; e
- $c \sim c'$ :  $c$  e  $c'$  são equivalentes.

A partir destas relações, as preferências do agente decisor devem ser especificadas segundo quatro axiomas: completude, transitividade, dominância e convexidade (Souza, 2002). Deste modo, teremos a **função de perda** definida conforme segue,

$$\begin{aligned} L : \mathcal{D} \times \Omega &\rightarrow \mathbb{R} \\ (d, \theta) &\mapsto L(d, \theta), \end{aligned}$$

permitindo trabalhar com números reais, além de preservar a ordem das preferências inicialmente especificada, isto é,  $c \succsim c' \Leftrightarrow L(c) \geq L(c')$ . Em termos práticos, teremos a função de perda representando o prejuízo que incorreremos ao optar por  $d$  quando o estado da natureza for  $\theta$ .

Para determinar a ação a ser tomada, procuramos pela decisão tal que a perda associada seja mínima. Porém, como a escolha precisa ser feita sem o conhecimento do estado da natureza em questão, atribuímos probabilidade ao conjunto  $\Omega$  sendo, portanto, visto como uma variável aleatória. Assim, estimamos o seu comportamento através da compreensão que o agente decisor possui sobre o espaço paramétrico. Sendo tal conhecimento representado por meio de uma distribuição de probabilidade  $\pi(\theta)$ , denominada priori. Supondo que esta satisfaça as suposições de coerência, podemos avaliar a perda esperada para cada decisão  $d$ , por intermédio da **perda esperada a priori**  $r_\pi(d)$ ,

$$r_\pi(d) = \mathbb{E}_\pi[L(d, \Theta)] = \int_\Omega L(d, \theta)\pi(\theta)d\theta, \quad (2.1)$$

---

<sup>1</sup>O agente pode ser caracterizado por: um indivíduo, um grupo de pesquisa, um país, etc. Trataremos apenas do primeiro caso, em que o agente é configurado por um único indivíduo.

também designada como **risco contra a priori**  $\pi$ . Note que, para todo caso em que os estados da natureza são discretos, basta substituir as integrais pelos respectivos somatórios. Por fim, admitindo que a integral apresentada em (2.1) é finita para cada  $d \in \mathcal{D}$ , a melhor decisão será aquela cuja perda a priori é minimizada para todos os elementos do espaço de decisões, ou seja:

**Definição 2** *Uma decisão  $d_\pi^*$  será ótima, em relação à distribuição  $\pi(\theta)$  se,*

$$d_\pi^* = \arg \min_{d \in \mathcal{D}} r_\pi(d).$$

Assim, qualquer ação obtida segundo a Definição 2 é denominada **decisão de Bayes contra a priori**  $\pi(\theta)$ . No mesmo sentido, o risco atingido é qualificado como **risco de Bayes a priori**.

Considerando agora o caso em que o agente decisor tenha acesso à uma amostra  $x$ , sendo a respectiva variável aleatória  $X$  definida segundo um modelo estatístico paramétrico  $(\chi, \mathcal{A}, \mathcal{F}, \Omega, \pi)$ , em que,

- $\chi$  é o espaço amostral;
- $\mathcal{A}$  é a sigma-álgebra de subconjuntos de  $\chi$ ;
- $\mathcal{F}$  especifica uma classe de distribuições sobre  $\mathcal{A}$ ;
- $\Omega$  é o espaço paramétrico que indexa  $\mathcal{F}$ ; e
- $\pi$  é a probabilidade especificada a priori em relação a  $\Omega$ ,

a tomada de decisão passa a contemplar o conhecimento de tais evidências. Como alteração, teremos que a metodologia passa a ser denominada **teoria da decisão estatística**. Assim, a decisão será especificada segundo uma **regra de decisão**  $\delta$ , definida para cada possível valor  $X \in \chi$  e distribuição  $\pi$ . Isto é,

$$\begin{aligned} \delta_\pi : \chi &\rightarrow \mathcal{D} \\ x &\mapsto \delta_\pi(x). \end{aligned}$$

Desta forma a função de risco é re-escrita em termos da regra de decisão e, conseqüentemente, irá possuir dois elementos aleatórios,  $\Theta$  e  $X$ , isto é,

$$R_\pi(\delta) = \mathbb{E}_\pi[L(\delta(X), \Theta)] = \int_\Omega \int_\chi L(\delta(x), \theta) f(x|\theta) \pi(\theta) dx d\theta. \quad (2.2)$$

Logo, similarmente ao caso inicial, procuramos a estratégia que minimiza o risco, ou seja:

**Definição 3** *Uma regra de decisão  $\delta_\pi^*$  será ótima, em relação a  $\pi(\theta)$  se,*

$$\delta_\pi^* = \arg \min_{\delta \in \mathcal{D}} R_\pi(\delta).$$

Porém, no caso da ordem de integração em (2.2) ser alterável, o que seguramente é possível se a perda for limitada e não-negativa, a obtenção de  $\delta_\pi^*$  é equivalente a encontrar a regra que minimiza a **perda esperada a posteriori**. Ou seja, o resultado da minimização do risco definido em (2.2) é o mesmo encontrado pela minimização do **risco contra a posteriori**  $\pi$ , definida conforme segue,

$$r_{\pi(\cdot|x)}(\delta) = \mathbb{E}_{\pi(\cdot|x)}[L(\delta(X), \Theta)] = \int_\Omega L(\delta(x), \theta) \pi(\theta|x) d\theta. \quad (2.3)$$

Deste modo a decisão ótima  $\delta_\pi^*$  será dada pela minimização do risco  $r_{\pi(\cdot|x)}(d)$ . Tal simplificação é extremamente útil, dada a complexidade envolvida na equação (2.2), em que é necessário considerar todas as possíveis regras de decisão associadas ao problema.

Por fim, no caso de o procedimento ser obtido segundo a Definição 3, tal ação será denominada **regra de decisão de Bayes contra a posteriori**  $\pi(\theta|x)$ , e o risco relacionado, que pode não ser único, **risco de Bayes a posteriori**.

Adicionalmente, temos que para uma função linear de  $L(d, \theta)$ , dada por uma constante positiva  $a$  e uma função  $\lambda(\theta) : \Omega \rightarrow \mathbb{R}$ , tal que a integral  $\int_{\Omega} \lambda(\theta)\pi(\theta)d\theta$  seja finita, isto é,

$$L^*(d, \theta) = aL(d, \theta) + \lambda(\theta), \quad \forall \theta \in \Omega \text{ e } d \in \mathcal{D},$$

é possível redefinir a função de perda  $L$  sem alterar o respectivo  $d$  mínimo. Desta forma, para quaisquer duas decisões  $d_1$  e  $d_2 \in \mathcal{D}$ , teremos  $r_{\pi}(d_1) \leq r_{\pi}(d_2) \Leftrightarrow r_{\pi}^*(d_1) \leq r_{\pi}^*(d_2)$ , em que  $r^*$  é o risco relacionado à perda  $L^*$ .

Para o caso da função  $\lambda^*$  ser dada por  $\inf_{d \in \mathcal{D}} L(d, \Theta)$ , satisfazendo a condição inicialmente imposta, e  $a = 1$ , podemos substituir a perda original  $L$  por  $L^*$ . Também conhecida como função de perda *regret*, tal penalidade será dada por,

$$L^*(d, \Theta) = L(d, \Theta) - \inf_{d \in \mathcal{D}} L(d, \Theta),$$

garantindo, portanto, uma perda não-negativa. Tal artifício permite a simplificação dos cálculos, sem qualquer alteração nas ordenações iniciais, ou nas decisões ótimas. Por tal razão, consideraremos apenas as funções de perda *regret* neste trabalho, mesmo sem a indicação explícita destas. Adicionalmente, existe a possibilidade de trabalhar com uma abordagem mais otimista, considerando ganhos, por exemplo. Para tal, é suficiente considerar o inverso aditivo da função de perda, dada pela função de utilidade  $U(d, \Theta)$ , isto é,  $U(d, \Theta) + L(d, \Theta) = 0$ .

A teoria da decisão também possui uma leitura frequentista, em que a amostra  $X$  é o único componente aleatório, apesar de  $X = x$ . Assim, o fato do parâmetro  $\Theta$  não ser encarado como variável aleatória impede que a esperança obtida em (2.2) seja avaliada no espaço paramétrico. Tal característica torna a abordagem mais restrita, uma vez que, para encontrar a ação ótima, é necessário limitar o espaço de decisões. Um caso bastante utilizado trata dos estimadores pontuais, em que a análise correspondente deve ser restrita à classe das decisões não-viesadas. Ou seja, considerando apenas o conjunto  $\Delta' = \{\delta(X) \in \Delta : \mathbb{E}[\delta(X)|\Theta = \theta] = \theta, \forall \theta \in \Omega\}$ . O estimador mais difundido procura a decisão tal que a perda máxima seja minimizada. Por tal motivo, esta ação é conhecida como decisão minimax (com raciocínio similar é possível trabalhar com a maximização do ganho mínimo, neste caso maximin).





## Capítulo 3

# Testes de Hipóteses

Os testes de hipóteses possuem o objetivo de indicar o cenário mais plausível dentre uma coleção de conjecturas. Entretanto, é usual trabalhar com apenas duas premissas, de modo que estas configurem uma partição do espaço paramétrico (ou de um subespaço de interesse). Tipicamente nomeadas como nula e alternativa, tais hipóteses seguirão a notação:

$$\begin{aligned}H_0 &: \Theta \in \Omega_0 \\H_1 &: \Theta \in \Omega_1,\end{aligned}$$

em que  $\Omega = \Omega_0 \cup \Omega_1$ ,  $\Omega_0 \cap \Omega_1 = \emptyset$ ,  $\Omega_0 \neq \emptyset$  e  $\Omega_1 \neq \emptyset$ .

Em termos teóricos, os procedimentos são especificados segundo uma função  $\varphi$ , definida na classe  $\psi = \{\varphi : \chi \rightarrow \{0, 1\}\}$ . De forma que decidimos por  $H_0$  se  $\varphi = 0$ , e  $H_1$  caso contrário, sendo o valor de  $\varphi$  determinado por meio de uma Região Crítica ( $RC$ ). Também denominada Região de Rejeição, tal subconjunto do espaço amostral é matematicamente dado por:  $\varphi^{-1}(\{1\}) = \{x \in \chi : \varphi(x) = 1\}$ , equivalentemente, a região de aceitação será:  $\varphi^{-1}(\{0\})$ .

No que tange a especificação das hipóteses, existem dois tipos de erros que podem ser cometidos. O erro do tipo I, que ocorre quando incorretamente assinalamos a hipótese alternativa como verdadeira e, analogamente, o erro do tipo II, que se refere à hipótese nula. Como propriedade, o decréscimo de um erro implica, necessariamente, no aumento do outro. Formalmente, os erros serão,

$$\begin{aligned}\mathbb{P}[\text{Erro do tipo I}] &= \alpha(\varphi) = \mathbb{P}[\varphi(X) = 1 | \Theta \in \Omega_0] \\ \mathbb{P}[\text{Erro do tipo II}] &= \beta(\varphi) = \mathbb{P}[\varphi(X) = 0 | \Theta \in \Omega_1].\end{aligned}$$

Para a análise das características probabilísticas, no contexto frequentista, trabalhamos com a Função Poder. Também denominado como Poder do Teste, tal quantidade associa a probabilidade de rejeitar a hipótese nula a cada valor de  $\Theta$ , ou seja,  $\gamma_\varphi(\theta) = \mathbb{P}[\varphi = 1 | \theta]$ . Posteriormente, definimos o tamanho do teste, dado pelo supremo da função poder, considerando apenas os valores de  $\Theta \in \Omega_0$ , ou seja,  $\alpha = \sup_{\Theta \in \Omega_0} \gamma_\varphi(\theta)$ . Por fim, chamamos o valor  $\alpha_0$  de nível de significância da família  $\psi^*$ , caso esse seja o limitante superior para os demais tamanhos de teste. Isto é,  $\psi^* = \{\varphi \in \psi : \alpha \leq \alpha_0\}$ , para  $0 < \alpha_0 < 1$ . Enquanto que, no contexto bayesiano, a abordagem é mais intuitiva, uma vez que trabalhamos diretamente com as probabilidades a posteriori das hipóteses.

Em relação à região de rejeição, teremos, no contexto frequentista, uma função denominada estatística de teste. Ao passo que, na leitura bayesiana, também dependerá do conhecimento a priori. Em termos gerais, tanto a especificação da região de rejeição, quanto o modo de interpretar os resultados, serão suficientes para distinguir as abordagens existentes na literatura. Segue a descrição das principais leituras para os testes de hipóteses.

### 3.1 Fisher & $p$ -valor

No contexto da inferência clássica, a leitura mais disseminada diz respeito à filosofia de Sir Karl Popper, difundida no meio estatístico por Sir Ronald Fisher. Segundo esta, uma teoria nunca pode ser comprovada mediante um estudo empírico sendo, entretanto, um contraexemplo suficiente

para a sua negação. Nos testes de hipóteses tal premissa implica em considerarmos um raciocínio indutivo, de modo que, independentemente da quantidade de evidências a favor da premissa em questão, esta não deverá ser aceita. Segundo o próprio Fisher (1935, pág.18):

“Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis”.

Apesar de não ser necessário indicar se estamos tratando da hipótese nula ou alternativa, uma vez que tal diferenciação não altera a referida teoria, é usual utilizarmos  $H_0$  como referência. Portanto, procuramos especificar tal hipótese de modo que o caso mais prejudicial, ou seja, aquele que atribuímos maior importância, seja o erro do tipo I.

Para o contexto discutido, o nível descritivo de significância observado (ou  $p$ -valor), introduzido por Pearson (1900), se apresenta como uma opção adequada. Esta escolha se deve ao fato do  $p$ -valor procurar, a partir das amostras não observadas, evidências em desacordo com a hipótese nula. Para tal, é necessário que o experimento relacionado esteja fixado, uma vez que trabalhamos com todas as possíveis amostras provenientes deste. Assim, por meio da ordenação do espaço amostral segundo  $H_0$ , examinamos a probabilidade de obter amostras tão extremas quanto a observada. Formalmente:

**Definição 4** *Seja  $T(X)$  uma estatística cujos valores pouco prováveis indicam evidências contra  $H_0$ . Se observarmos  $X = x$  com correspondente  $T(x) = t(x)$ , o nível descritivo de significância observado  $\alpha^*$ , será a probabilidade de obter valores tão ou mais extremos que o  $t(x)$  obtido, considerando-se que  $\theta$  pertença ao subespaço  $\Omega_0$ .*

Tal avaliação pode ser feita, por exemplo, por meio de cálculos do tipo  $\alpha^* = \sup_{\Theta \in \Omega_0} \mathbb{P}(T(X) \geq t(x) | \Theta \in \Omega_0)$ . Logo, a concordância entre o modelo especificado e os dados observados é proporcional à  $\alpha^*$ . Uma leitura usual trata da rejeição da hipótese se  $X \in \varphi^{-1}(\{1\}) = \{x \in \mathcal{X} : \alpha^* < \alpha^+\}$ , em que  $\alpha^+$  é a referência de quão extrema a amostra poderia ser. Idealmente, tal valor de referência deve ser discriminado pelo analista: “In choosing the grounds upon which a general hypothesis should be rejected, personal judgement may and should properly be exercised” Fisher (1956, pág.50). Contudo, o próprio autor indica 0,05 como um limite conveniente, dado que transcreve “desvios superiores a duas vezes o desvio padrão” (Fisher, 1925).

Apesar do  $p$ -valor, ou *tail area*, ser um instrumental coerente com a proposta da falsificação empírica, além de amplamente utilizado, tal métrica possui uma série de características indesejáveis. Dentre as mais citadas, temos:

- utiliza contra a hipótese nula não apenas a evidência de fato obtida, mas todos os valores mais extremos que  $t(x)$ , quando estes sequer são observados;
- o significado da rejeição da hipótese indica apenas que há algo de errado com o modelo utilizado, sem especificar qual é o problema. Tal limitação, pode ser referente à: distribuição assumida, hipótese de independência, desconsideração da hipótese alternativa, etc (Berger e Delampady, 1987a);
- a falta de coerência entre a rejeição de uma hipótese e seus subconjuntos, uma vez que não se trata de uma medida de avaliação do espaço paramétrico, mas sim do subespaço definido pela hipótese nula (Patriota, 2013);
- a sua magnitude ser dependente do tamanho da amostra, assim como sua interpretação decorre da verossimilhança adotada e regra de parada utilizada (Wagenmakers, 2007), coexistindo diferentes padrões para tal na literatura (Pereira e Wechsler, 1993);
- no contexto de duas hipóteses definidas segundo uma partição de  $\Omega$ , podemos conjecturar exemplos em que, devido a falta de exigência quanto a incorporação da hipótese alternativa na análise, obtem-se conclusões contraditórias (Pereira e Wechsler, 1993); e

- dificuldade de interpretação, dado que a definição condicional  $\mathbb{P}(x|\Omega_0)$  é sumariamente intuída como uma probabilidade condicional  $\mathbb{P}(\Omega_0|x)$  (van der Pas, 2010).

Ao procurar evidências contra  $H_0$ , ao invés de avaliar ambas as hipóteses, estamos tratando de uma classe específica de testes, denominada **testes de significância**. Por tal razão, os testes definidos segundo os princípios desta seção serão diferenciados ao longo do texto.

### 3.2 Neyman-Pearson & Razão de Verossimilhanças

A perspectiva defendida por Jerzy Neyman e Egon Pearson (N-P), complementa o cenário frequentista, no que tange os testes de hipóteses. Nesta leitura, o raciocínio é desenvolvido para conclusões complementares em relação às hipóteses. Utilizando para tal, um nível de significância pré-especificado e uma sequência de experimentos teórica.

Para a descrição desta leitura consideraremos, inicialmente, que o teste seja constituído por hipóteses simples, ou seja,  $H_0 : \Theta = \theta_0$  e  $H_1 : \Theta = \theta_1$ . Assim, a região crítica será dada em função da razão das verossimilhanças, avaliadas nos respectivos subespaços de  $\Omega$ , isto é,  $\lambda(X) = f(X|\theta_0)/f(X|\theta_1)$ . Contudo, dada a impossibilidade de controlarmos simultaneamente os dois erros envolvidos, a análise é limitada à família  $\psi^*$ , composta pelos testes de nível de significância  $\alpha_0$ . Formalmente, para  $k \geq 0$ ,

$$\varphi^*(x) = \begin{cases} 1 & \text{se } \lambda(x) < k \\ 0 & \text{se } \lambda(x) > k, \end{cases}$$

enquanto que para o caso  $\lambda(x) = k$ , qualquer um dos valores pode ser aceito. Adicionalmente teremos  $\gamma_{\varphi^*}(\theta_1) \geq \gamma_{\varphi}(\theta_1)$ ,  $\forall \varphi \in \psi^*$  e, pelo Lema de Neyman-Pearson,  $\varphi^*(x)$  será o teste mais poderoso de nível  $\alpha_0$ .

No caso de uma das hipóteses ser composta, digamos  $H_1$ , restringimos o domínio para os testes Uniformemente Mais Poderosos (*UMP*). Nesta família, é necessário garantir que o teste seja o mais poderoso para cada possível valor da hipótese composta, isto é, a região crítica não deverá depender de nenhum valor em particular de  $\Theta \in \Omega_1$ . O mesmo raciocínio é utilizado para o caso de hipóteses compostas unilaterais.

Em contextos mais gerais, como hipóteses bilaterais ou parâmetros desconhecidos, definimos o teste por meio da razão de verossimilhanças generalizada (TRVG), quantidade que trabalha com o supremo de cada uma das verossimilhanças. Contudo, neste caso, as propriedades garantidas são limitadas a uma família ainda mais restrita, em que apenas os procedimentos com o erro tipo I igual a  $\alpha_0$  são comparados, ou seja,  $\psi^{**} = \{\varphi \in \psi : \gamma_{\varphi^{**}}(\theta_1) = \gamma_{\varphi}(\theta_1)\}$ . Logo, um teste  $\varphi^{**}(x) \in \psi^{**}$ , com a região crítica dada por,

$$RC = \left\{ x : \frac{\sup_{\Theta \in \Omega_0} L(\theta|x)}{\sup_{\Theta \in \Omega_1} L(\theta|x)} \leq k, \right\}$$

em que  $k \geq 0$ , será um teste uniformemente mais poderoso (*UMP*). Em termos gerais, para ampliarmos as análises com algumas propriedades garantidas, será necessário seguir aplicando restrições no domínio dos testes. Como exemplos, podemos citar: testes não viesados uniformemente mais poderosos, testes invariantes e uniformemente mais poderosos, família exponencial, etc.

Um ponto importante para a caracterização desta abordagem diz respeito à especificação do teste ser realizada com ênfase no valor do erro tipo I. Tal prática está relacionada ao conceito de repetibilidade da amostra, sugerindo que cometeremos a inapropriada rejeição de  $H_0$  em  $\alpha\%$  das decisões. Adicionalmente, o resultado do procedimento não será uma inferência probabilística, mas sim um comportamento, no caso, o de aceitar ou rejeitar a hipótese de referência. Tal característica se deve à falta de mensurações associadas. Contudo, assim como na leitura anterior, existem pontos controversos para a análise de N-P, são eles:

- com incremento do tamanho da amostra, podemos ter um desequilíbrio na relação entre os erros I e II, e até mesmo a inversão da correspondência inicialmente especificada (Pericchi e Pereira

, 2013);

- dificuldade em especificar a distribuição amostral, sendo usual recorrer a distribuições assintóticas;
- para variáveis discretas, existe ainda a iminência de trabalhar com testes aleatorizados. Situação em que decidimos sobre os pontos vizinhos à região de rejeição de forma aleatória. Isto é, atribuímos para a aleatoriedade o cargo de agente decisor, em nome apenas de uma precisão matemática;
- após a obtenção dos resultados do experimento, a relevância do  $\alpha$  se torna questionável, dado que reflete amostras que não foram e não virão a ser observadas. Adicionalmente, a própria conclusão diz respeito a uma sequência de experimentos, quando o real interesse reside em tomar decisões pontuais; e
- adotar valores pré-especificados para o nível de significância, tipicamente  $\alpha = 0,05$  e  $\beta < 0,20$ , sem crítica quanto à natureza do evento ou ao modelo utilizado. Tendo como principal justificativa a incidência literária.

Esta teoria não se trata de um concorrente dos testes de significância, dado que aborda o problema segundo uma perspectiva diferente. Contudo, nas aplicações usuais, prevalecem os cálculos e as propriedades da metodologia de Neyman-Pearson, enquanto que as conclusões são dadas segundo o raciocínio das provas por contradição, defendido por Fisher. Além disso, as aplicações consideram a comparação do  $p$ -valor com um nível de significância pré-especificado. Tal combinação se mostra confusa, uma vez que um sistema considera o experimento em questão, enquanto o outro trabalha com uma série hipotética de repetições. Como consequência a esta falta de referência, podemos obter conclusões antagônicas (Christensen, 2005), chegando a apresentar um  $p$ -valor de 0.05 com uma probabilidade, frequentista, de 0.5 Berger (2003).

Note que dividimos os testes de hipóteses em duas classes. A primeira trata dos testes de significância, em que o foco é dado para as evidências contra ou a favor de uma determinada hipótese. Enquanto que a segunda analisa ambas as hipóteses, definindo a estatística de teste de modo “imparcial”. Tendo que esta última não possui denominação específica, será referenciada neste texto como: teste de hipótese. E, quando necessário contemplar a união destas, faremos a referência explicitamente.

Na abordagem frequentista como um todo, o fato de tratarmos apenas os dados como relevantes também configura uma limitação, conforme argumentado em Jeffreys (1961, pág.385) “An hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred”.

### 3.3 Bayes & Medidas Condicionais

Em contraste com as teorias frequentistas, que baseiam suas conclusões em amostras e eventos não observados, a inferência bayesiana apresenta conclusões derivadas diretamente do espaço paramétrico. Tal fato possibilita a análise explícita das probabilidades associadas às hipóteses, de acordo com as expectativas dos usuários. Quanto às conclusões, decidimos entre as hipóteses de acordo com a razão de probabilidades das hipóteses. Para avaliação desta, trabalhamos com constantes, que dependem da função de perda utilizada, isto é,

$$\frac{\mathbb{P}(\Theta \in \Omega_0 | x)}{\mathbb{P}(\Theta \in \Omega_1 | x)} = \frac{\mathbb{P}(X | \Theta \in \Omega_0) \mathbb{P}(\Theta \in \Omega_0)}{\mathbb{P}(X | \Theta \in \Omega_1) \mathbb{P}(\Theta \in \Omega_1)} \geq k(L(d, \Theta)).$$

Assim, no caso de  $H_0$  ser relativamente mais plausível que  $H_1$ , no contexto da amostra observada, teremos a referida razão maior que um. Este raciocínio é interessante, uma vez que contempla não apenas a aceitabilidade de uma hipótese isolada, mas também as circunstâncias do referido complementar, sem prioridades. Apesar de esta leitura permitir caracterizações com “regras de

bolso”, poderemos aceitar ou rejeitar a hipótese de interesse de modo subjetivo, desde que a razão de prioris seja incorporada. É possível também, em certos contextos, apresentar apenas a razão de verossimilhanças, então chamada de *Fator de Bayes* (Jeffreys, 1961), deixando a cargo do leitor a multiplicação pela razão de probabilidades a priori.

Uma crítica usual a esta abordagem, diz respeito ao conhecimento subjetivo ser agregado à análise. Contudo, é importante observar que isto ocorre inevitavelmente. A questão é apenas que grau de importância será dada à tal etapa. Na compreensão frequentista, por exemplo, o analista incorpora sua experiência através do desenho do experimento, modelo escolhido, ou até mesmo nas interpretações. Porém, apesar de ser possível trabalhar em conjunto com os especialistas nestes itens, a falta de formalidade dada a esta etapa, torna tal prática pouco usual. Em contraponto, no caso bayesiano, a priori configura um papel crucial na análise, de modo que a questão é obrigatoriamente abordada e, sempre que viável, tratada pelo próprio especialista - pessoa com o real conhecimento para tal.

Testes de hipóteses, em geral, possuem um importante caso especial, quando a hipótese de interesse possui medida de Lebesgue nula. Também conhecido como hipóteses precisas, tal tópico será detalhado no próximo capítulo.



## Capítulo 4

# Hipóteses Precisas

No caso do subespaço  $\Omega_0$  possuir uma dimensão inferior à de  $\Omega_1$ ,  $H_0$  passa a ser classificada como uma hipótese precisa. Em consequência, uma série de particularidades são impostas às análises. O exemplo mais conhecido é o caso em que o espaço paramétrico é definido na reta real, ou seja,  $H_0 : \Theta = \theta_0$  versus  $H_1 : \Theta \neq \theta_0$ , circunstância à qual daremos ênfase. Tal cenário é particularmente relevante, uma vez que permite o estudo de questões como o descarte de covariáveis em modelos estatísticos, ou a verificação da igualdade entre tratamentos, ambos cenários recorrentes no contexto experimental.

Contudo, a necessidade desta classe de hipóteses é constantemente posta em cheque, sendo classificada por muitos autores como utópica em termos práticos. O principal argumento pode ser resumido pelo seguinte parágrafo de [Tukey \(1991, pág.100\)](#):

“All we know about the world teaches us that the effects of A and B are always different - in some decimal place - for any A and B. Thus asking ‘Are the effects different?’ is foolish”.

De modo que, segundo esta visão, a questão deveria ser: “em qual direção o efeito de A difere do efeito de B”, tendo como opções: positivo, negativo ou incerto.

Apesar de a argumentação ser válida, entendemos, neste texto, que a escolha de qual tipo de hipóteses serão consideradas em uma análise trata-se de um problema de decisão. De modo que escolhemos entre, adotar uma hipótese falsa que constitua uma “boa” aproximação da realidade, ou uma hipótese mais, porém também mais complexa. Assim, após a especificação da referida função de perda, o teste indicará quando a aproximação é uma decisão ótima. Portanto, podemos tratar “testes de hipóteses precisas” como decisões ótimas de Bayes, desde que o analista tenha uma penalidade com tal tendência. Adicionalmente, é interessante notar que a escolha de modelos probabilísticos para análises diversas se tratam, igualmente, de aproximações.

Uma crítica adicional diz respeito a não ser possível aceitar uma hipótese precisa, no caso da priori sobre o espaço paramétrico ser contínua, dado que a reta real é não enumerável. Para esta questão, além do argumento já apresentado, temos os exemplos no contexto dos fenômenos físicos como ilustração válida.

### 4.1 Testes Frequentistas

Devido a ausência de probabilidade em  $\Omega$ , não há restrições matemáticas para hipóteses com medida de Lebesgue nula na leitura frequentista. Ainda assim, toda a discussão do preâmbulo deste capítulo se mantém, tendo diferentes reflexos para cada uma das leituras ([Cox, 1977](#)). No caso dos testes de significância, por exemplo, a restritividade do subespaço  $\Omega_0$  atribuí particular importância à estrutura “popperiana”, dado a limitação da hipótese em relação ao espaço paramétrico como um todo. Uma ilustração recorrente na literatura discute a afirmação “todos os cisnes são brancos”, em que, segundo Popper, independentemente da quantidade de cisnes brancos observados, não podemos

corroborar tal afirmação. Contudo, com apenas um cisne preto (ou cinza, tigrado, etc), a conjectura poderia ser classificada como falsa.

No caso das conclusões da leitura de Neyman-Pearson, temos os mesmos tipos de críticas apresentadas. Logo, temos o mesmo argumento que, independentemente da restritividade da hipótese, o analista precisa definir a ação a ser tomada.

## 4.2 Testes Bayesianos

No contexto bayesiano, caso a distribuição a priori sobre  $\Omega$  seja contínua, a probabilidade a posteriori do subconjunto  $\Omega_0$  será nula, impossibilitando a utilização das abordagens usuais. Logo, devido a tal dificuldade, outros critérios para análise desta classe de testes foram desenvolvidos. Segue a descrição das metodologias mais relevantes.

### 4.2.1 Jeffreys

Jeffreys (1961) contornou o problema da probabilidade nula, impondo a especificação de uma probabilidade positiva a priori para  $H_0$ . Desta forma, passamos a ter  $\pi_{\theta_0}(\theta, \zeta)$  definida segundo uma combinação das probabilidades:  $\pi(\theta)$ , para  $\Omega_1$ , e  $\zeta$  em  $\Omega_0$ . Ou seja,

$$\pi_{\theta_0}(\theta, \zeta) = \begin{cases} \zeta & \text{se } \Theta = \theta_0 \\ (1 - \zeta)\pi(\theta) & \text{se } \Theta \neq \theta_0, \end{cases}$$

em que  $0 < \zeta < 1$ . Assim, a probabilidade a posteriori será

$$\pi_{\theta_0}(\theta, \zeta|x) = \begin{cases} \zeta \frac{f(x|\theta_0)}{f(x)} & \text{se } \Theta = \theta_0 \\ (1 - \zeta) \frac{f(x|\theta)\pi(\theta)}{f(x)} & \text{se } \Theta \neq \theta_0, \end{cases}$$

para  $f(x) = \zeta f(x|\theta_0) + \int_{\theta \neq \theta_0} (1 - \zeta) f(x|\theta)\pi(\theta)d\theta$ . Logo, a escolha entre as hipóteses poderá ser feita seguindo a metodologia usual.

O teste de Jeffreys é difundido a ponto de, em alguns livros, ser nomeado como “teste de Bayes para hipóteses precisas”. Entretanto, ao atribuir uma massa de probabilidade positiva  $\zeta$  para a hipótese nula, estamos diminuindo a probabilidade dos demais pontos do espaço paramétrico na proporção de  $(1 - \zeta)$ . Tal reponderação não seria um problema, se esta fosse de fato a opinião do analista quanto ao comportamento do evento. Contudo, usualmente esse não é o caso, tornando esta metodologia uma abordagem incoerente.

### 4.2.2 FBST

Visando desenvolver um teste de significância bayesiano que mantivesse as suposições de coerência, Pereira e Stern (1999) introduziram o teste de significância genuinamente bayesiano. Conhecido também como FBST (*Full Bayesian Significance Test*), tal métrica é desenvolvida com base no princípio da menor surpresa<sup>1</sup>, de modo que procuremos por evidências a favor da hipótese nula. Apesar deste teste ser viável para aplicações em diferentes espaços, desde que exista uma preocupação diferenciada em relação a uma das decisões, sua contribuição é mais expressiva no contexto das hipóteses precisas.

Dada a ordenação do espaço paramétrico segundo a probabilidade a posteriori, procuramos o  $\theta^*$  pertencente a região de  $H_0$ , tal que sua densidade seja máxima. E, a partir deste, formatamos o conjunto tangente à hipótese nula, ou conjunto de credibilidade  $T(x)$ , que é configurado por todos os pontos com densidade inferior ao  $\theta^*$  obtido. Assim, definimos a quantidade  $EV(\Omega_0, x)$ , ou *e-value*, como a probabilidade a posteriori do conjunto  $T(x)^C$ . Formalmente,

<sup>1</sup>Construção de um subconjunto do espaço paramétrico, cujos pontos são mais inesperados do que o observado.



**Definição 5** Para o conjunto tangente  $T(x) = \{\theta : \pi(\theta|x) > \sup_{\Omega_0} \pi(\theta|x)\}$ , a medida de evidência do FBST, a favor de  $H_0$ , será

$$EV(\Omega_0, x) = 1 - \mathbb{P}(\Theta \in T(x)|x) = \int_{T^c(x)} \pi(\theta|x) d\theta.$$

Para valores altos de  $EV(\Omega_0, x)$ ,  $\theta_0$  estará entre os pontos mais prováveis a posteriori e, portanto, favorecerá a hipótese nula. Como principais vantagens deste procedimento podemos citar,

- lógica intuitiva com interpretação geométrica, além de se tratar de uma probabilidade no espaço paramétrico de interesse;
- aplicabilidade mesmo sem uma distribuição a priori disponível. Bastando para tal substituir a posteriori pela verossimilhança normalizada, caso essa exista (Pereira *et al.*, 2008);
- consistência e invariância sob transformações um-a-um do parâmetro;
- conformidade com o Princípio da Verossimilhança (Madruga *et al.*, 2003);
- possibilidade de considerar qualquer dimensão, tanto do parâmetro quando do espaço amostral, uma vez que trabalha com otimização numérica e técnicas de integração; e
- procedimento exato, sem a necessidade de aproximações assintóticas, eliminação de parâmetros ou atribuição *ad hoc* de probabilidade positiva para conjuntos com medida de Lebesgue nula.

Adicionalmente, o FBST configura um teste de Bayes, ou seja, é o resultado da minimização de uma função de perda esperada, conforme será discutido no próximo capítulo.



# Capítulo 5

## Funções de Perda

A variedade de abordagens existentes, tanto para os testes de hipóteses quanto para os testes de significância, bem como a relevância destes no contexto aplicado, justificam a ampla literatura comparando e relendo as principais metodologias (Aitkin *et al.*, 2005; Berger e Sellke, 1987). Porém, o fato de cada abordagem ser construída segundo estruturas e objetivos distintos, variando inclusive os espaços em que são avaliados (amostral ou paramétrico), torna muitos pontos discutíveis, limitando as conclusões aos cenários especificados.

Como alternativa, iremos abordar os testes segundo uma linguagem única: a teoria da decisão. A partir desta perspectiva, poderemos realizar comparações mais justas, incorporando a descrição do raciocínio condizente com cada metodologia. Para tal, utilizaremos o espaço de decisões  $\mathcal{D}$ , dado por  $\{d_0, d_1\}$ , em que  $d_i$  denota a ação de aceitar a hipótese  $H_i : \Theta \in \Omega_i$ , com  $i \in \{0, 1\}$ , e  $\Omega$  contemplando os estados da natureza. Assim, a função de perda será especificada para cada possível consequência  $\mathcal{C}$  ( $\mathcal{D} \times \Omega$ ), de modo que reflita o entendimento pretendido.

Neste trabalho, seguiremos a teoria apresentada no Capítulo 2, sempre atribuindo aleatoriedade ao conjunto  $\Omega$ . Note que tal especificação não invalida a generalização das análises, dado que o interesse é replicar a filosofia de cada abordagem e não o sistema em si. Inicialmente, consideraremos o caso mais simples, em que a função de perda só incorpora os erros, dados segundo números reais. Deste modo, utilizando a notação de Parmigiani e Inoue (2009), teremos a perda segunda a Tabela 5.1.

	$d_0$	$d_1$
$\Theta \in \Omega_0$	0	$L_1$
$\Theta \in \Omega_1$	$L_0$	0

**Tabela 5.1:** Função de perda genérica.

Logo, para uma amostra  $x$ , a regra de decisão  $\delta$  será definida no contexto binário. E a função de risco será dada por,

$$R_\pi(\delta) = \begin{cases} L_1 \mathbb{P}(\delta(X) = d_1 | \Theta) & \text{se } \Theta \in \Omega_0 \\ L_0 \mathbb{P}(\delta(X) = d_0 | \Theta) & \text{se } \Theta \in \Omega_1. \end{cases}$$

Portanto, a perda esperada a posteriori será,

$$\begin{aligned} r_{\pi(\cdot|x)}(\delta) &= \mathbb{E}_{\pi(\cdot|x)}[L(\delta(X), \Theta)] \\ &= \begin{cases} L_0(1 - \pi(\theta \in \Omega_0|x)) & \text{se } \delta = d_0 \\ L_1\pi(\theta \in \Omega_0|x) & \text{se } \delta = d_1. \end{cases} \end{aligned}$$

Deste modo, teremos a regra,

$$\varphi_{\pi}(x) = \begin{cases} d_0 & \text{se } \frac{\pi(\Theta \in \Omega_0|x)}{1-\pi(\Theta \in \Omega_0|x)} > \frac{L_0}{L_1} \\ d_1 & \text{se } \frac{\pi(\Theta \in \Omega_0|x)}{1-\pi(\Theta \in \Omega_0|x)} < \frac{L_0}{L_1}. \end{cases} \quad (5.1)$$

Assim, a conclusão será dada segundo o já conhecido Fator de Bayes. Contudo, podemos caracterizar a função de perda segundo outros critérios, resultando em outra regra de decisão, coerente com as peculiaridades especificadas.

É interessante ressaltar que, por incorporarem meios não experimentais às análises, a teoria da decisão e a inferência bayesiana apresentam uma forte relação. De modo que a combinação destas permite uma maior abrangência nas análises, tanto do ponto de vista matemático quanto subjetivo. A exceção diz respeito ao caso de tratar  $\Omega$  segundo uma leitura não aleatória, apesar desta não se apresentar como uma escolha sustentável, conforme discutido anteriormente. Por fim, a minimização da perda esperada garante que a referida decisão seja compatível com os axiomas de coerência. Por tais razões, iremos identificar as funções de perda que caracterizam os testes como decisões de Bayes, indicando as características e raciocínios condizentes.

## 5.1 Testes de Hipóteses

No casos dos testes de hipóteses, a função de perda apresentada na Tabela 5.1, definida através de constantes, é suficiente para a obtenção dos testes. Ainda assim, observamos peculiaridades em outras vertentes.

### 5.1.1 Razão de Verossimilhanças

O teste de N-P é aplicado de modo que os erros tipo I e II sejam especificados separadamente e, usualmente, atribuindo relevâncias distintas para estes. Entretanto, dadas as limitações já discutidas, DeGroot (1970) releu a questão segundo uma perspectiva mais abrangente. Nesta, ao invés de determinar o valor do erro tipo I e minimizar o erro tipo II, trabalhamos com a minimização da combinação linear dos erros. Contudo, este desenvolvimento é delimitado às hipóteses simples. Neste contexto, Pericchi e Pereira (2013) generalizaram a ideia, mediante a ponderação das verossimilhanças. Assim, para as funções  $\pi_0(\theta)$  e  $\pi_1(\theta)$ , definimos,

$$\begin{aligned} \varpi_0(x) &= \int_{\Omega_0} f(x|\theta)\pi_0(\theta)d\theta \quad \text{e} \\ \varpi_1(x) &= \int_{\Omega_1} f(x|\theta)\pi_1(\theta)d\theta. \end{aligned}$$

Similarmente ao caso inicial, para  $L_0$  e  $L_1$  positivas<sup>1</sup>, minimizamos a função:

$$\text{SERRORS}(\varphi) = L_1 \cdot \alpha(\varphi) + L_0 \cdot \beta(\varphi).$$

De modo que obtemos a regra,

$$\varphi_{\pi}^{RV}(x) = \begin{cases} d_0 & \text{se } \frac{\varpi_0(x)}{\varpi_1(x)} > \frac{L_0}{L_1} \\ d_1 & \text{se } \frac{\varpi_0(x)}{\varpi_1(x)} < \frac{L_0}{L_1}, \end{cases}$$

e, no caso de  $L_1 \cdot \varpi_0(x) = L_0 \cdot \varpi_1(x)$ , qualquer decisão é viável. Ou seja, obtemos um teste globalmente ótimo, além de um balanceamento entre os erros especificados e o tamanho da amostra.

Um ponto particularmente relevante diz respeito à descaracterização do Paradoxo de Lindley como uma divergência entre métodos frequentistas e bayesianos (Lindley, 1957). Isto porque os teste de N-P e DeGroot apresentam o mesmo paradoxo, mesmo no caso de não contemplarmos prioris (Pericchi e Pereira, 2013). Tal conclusão deve-se ao fato dos testes frequentistas clássicos

<sup>1</sup>O que seguramente ocorrerá uma vez que trabalhamos com funções de perda *regret*

manterem o erro tipo I fixo, independente do tamanho da amostra. Enquanto que, para a minimização da soma dos erros, ou para os testes bayesianos, os erros são inversamente proporcionais ao tamanho da amostra e, portanto, convergem para zero.

### 5.1.2 Jeffreys

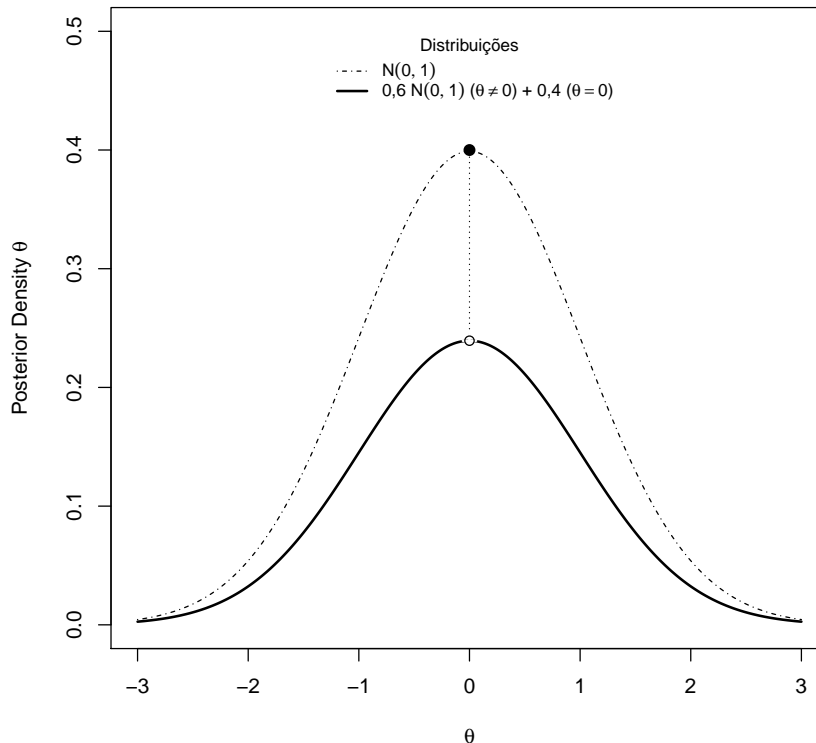
A solução introduzida por Jeffreys é um subcaso da regra de Bayes, apresentada na introdução deste capítulo. Como consequência, temos apenas a substituição da probabilidade do subespaço  $\Omega_0$ , conforme segue:

$$r_{\pi(\cdot|x)}(\delta) = \mathbb{E}_{\pi(\cdot|x)}[L(\delta(X), \Theta)] = \begin{cases} L_0\zeta & \text{se } \delta = d_0, \\ L_1(1 - \zeta)\pi(\theta \notin \Omega_0|x) & \text{se } \delta = d_1. \end{cases}$$

Assim,  $\varphi_{\pi}^J(x) = d_0$  se,

$$\frac{\zeta}{(1 - \zeta)\pi(\theta \notin \Omega_0|x)} > \frac{L_0}{L_1}.$$

Ao atribuímos probabilidade positiva a um ponto com densidade originalmente nula, temos como motivação apenas um paliativo prático para uma limitação matemática, sem qualquer justificativa filosófica. Deste modo, a natureza contínua do parâmetro é distorcida, ignorando o espaço paramétrico contínuo como sistema de ordenação genuíno, isto é, com probabilidade nula para cada ponto. Por consequência, temos a diminuição de todos os pontos pertencentes à  $\Omega_1$ , conforme apresentado no exemplo da Figura 5.1, para o caso de  $\pi(\theta) \sim N(0, 1)$  versus  $\pi_{\{\theta=0\}}(\theta, 0.4) \sim N(0, 1)$ .



**Figura 5.1:** Efeito da Priori de Jeffreys

Ao imputar tal probabilidade, sem uma real motivação, passamos a violar os princípios de coerência. Deste modo, no contexto do *Dutch Book Argument*, um *bookmaker* pode configurar um jogo em que teríamos prejuízo, independentemente da ação.

Adicionalmente, no contexto dos testes de hipótese, é necessário uma maior quantidade de evidências contra  $H_0$  para possibilitar sua rejeição, apesar de não estarmos tratando de testes de significância. Uma questão adicional, é que ao assumir uma massa de probabilidade positiva para a hipótese nula, a função de perda do FBST é descaracterizada como decisão de Bayes (Madruga *et al.*, 2001). Além de perdemos a separabilidade, já que a decisão passa a depender da densidade  $\pi$  para  $\zeta > 0$  (no caso de  $\zeta = 0$  temos o FBST usual).

## 5.2 Testes de Significância

Para os testes de significância, considerando que existe uma postura diferenciada em relação à hipótese nula, as decisões são apresentadas em relação à  $H_0$ . Isto é,  $d_1$  passa a ser lida como a rejeição de  $H_0$ . Adicionalmente, tal comportamento conservador é incorporado à análise por meio da função de perda.

### 5.2.1 FBST

O FBST faz uso apenas da informação contida na densidade a posteriori, e por tal razão é classificado como *Full Bayesian*. Porém, foi apenas no trabalho de Madrugá *et al.* (2001) que esta medida foi analisada segundo a teoria da decisão, sendo obtida através da minimização de uma função de perda:

**Definição 6** A função de perda correspondente ao e-valor,  $L^E : \mathcal{D} \times \Omega \times \chi \rightarrow \mathbb{R}^+$ , para  $a, b, c > 0$ , é dada por:

	Aceito $H_0$	Rejeito $H_0$
$\theta \notin T(x)$	$b$	$a$
$\theta \in T(x)$	$b + c [\mathbb{I}\{\theta \in T(x)\}]$	$0$ .

Note que, diferentemente da teoria clássica em que a perda é definida em termos do espaço de decisões e do espaço paramétrico, aqui consideramos uma classe mais abrangente. Nesta, a amostra observada também é incorporada à função de perda. Assim, dado que o espaço tangente é definido a partir da densidade a posteriori, ou seja, a partir da amostra, teremos  $L(d, \theta, x)$ . Logo, verificando que o FBST é resultado da minimização da função de perda apresentada na Definição 6, o risco a posteriori será:

$$\begin{aligned} r_{\pi(\cdot|x)}(d_0) &= \mathbb{E}_{\pi(\cdot|x)}[L(d_0, \Theta)] = b + c(1 - EV(\Omega_0, x)), \\ r_{\pi(\cdot|x)}(d_1) &= \mathbb{E}_{\pi(\cdot|x)}[L(d_1, \Theta)] = a EV(\Omega_0, x). \end{aligned}$$

E a aceitação da hipótese  $H_0$  ( $\varphi_\pi^E(x) = d_0$ ) se dará se, e somente se,

$$EV(\Omega_0, x) > \frac{b+c}{a+c}.$$

O FBST é construído de modo que quanto maior a probabilidade a posteriori do conjunto tangente, menos provável será a região a qual  $H_0$  pertence. Em termos da função de perda, avaliamos o conjunto tangente, que nada mais é do que o reflexo do nosso conhecimento e das informações trazidas pela amostra. Ou seja, devemos ter erros associados, dado que tais fontes de conhecimento não são absolutas. Contudo, por se tratar de uma medida de significância, a penalidade associada à rejeição da hipótese nula quando  $\{\theta \in T(x)\}$  é menor. E, por meio de uma adaptação da função *Regret*, tal quantidade será nula. Refletindo o fato de darmos menor importância para o erro associado à rejeição, conforme Definição 6.

Para um melhor entendimento, iremos apresentar os riscos para os casos de evidências: máxima, mínima e “intermediária”. Assim, seguindo a Tabela 5.2, para o caso da nossa opinião e amostra indicarem a rejeição de  $H_0$ ,  $EV(\Omega_0, x) = 0$ , logo  $d_1$  obterá risco zero. Enquanto que, para  $EV(\Omega_0, x) = 1$ , ou seja, quando temos todos os indicativos possíveis de que a hipótese nula deve ser aceita, ainda assim, teremos a penalidade  $b$  associada. Adicionalmente, a penalização do nosso

“acerto” quanto a escolha de  $H_0$ , pode ser superior ao erro sobre  $H_1$ , sendo necessário para tal apenas que  $b > a$ . Por fim, no caso da evidência intermediária, teremos a mesma penalização para  $a$  e  $c$ , nossos quase-erros tipo I e II (o quase se deve ao fato de avaliarmos  $T(x)$  e não  $\Omega$ ) e, portanto, a decisão é dada através de  $b$ .

	$r_{\pi(\cdot x)}(d_0)$	$r_{\pi(\cdot x)}(d_1)$
$EV(\Omega_0, x) = 0$	$b + c$	$0$
$EV(\Omega_0, x) = 1$	$b$	$a$
$EV(\Omega_0, x) = 0.5$	$b + 0,5c$	$0,5a$

**Tabela 5.2:** Função de Perda esperada - casos descritivos.

Ao especificarmos a função de perda segundo os “reais” valores do espaço paramétrico, sem nos preocuparmos com o nosso conhecimento ou com os resultados da amostra, estamos trabalhando com a penalidade relacionada ao fenômeno, e não ao estudo. Isto é, por tal cenário se tratar de uma simplificação, deixamos de incorporar nossas preferências quanto aos reflexos do estudo, como questões de caráter psicológico, financeiro e até mesmo social, vinculados aos respectivos resultados. Enquanto que, ao tratarmos das perdas segundo o que vamos de fato obter, e não com um cenário utópico de conhecimento absoluto, temos uma leitura muito mais realista, além de uma compreensão plena da situação analisada.

Thulin (2014) apresenta a função de perda do FBST como um metodologia que necessita de “a good deal more subjectivity”, afirmação esta que além de estar correta, é, neste trabalho, vista como uma característica desejável. Por fim, temos que a penalidade associada ao FBST é apenas um exemplo desta classe de funções, bem como Bernardo e Smith (1994), por exemplo, que formalizaram uma perda para a função de perda total.





## Capítulo 6

# Outra Opção Genuinamente Bayesiana para Testes de Significância

O Professor Carlos Pereira foi desafiado pelo Professor O. Kempthorne a “apresentar uma alternativa bayesiana ao *p-value*” (Pereira, em comunicação pessoal). A resposta a tal desafio é a gênese do FBST por Pereira e Stern (1999), e da consequente formalização por Madruga *et al.* (2001). Recentemente, Rice (2010) introduziu um novo teste bayesiano de significância para hipóteses precisas, tendo como base o princípio de Popper. Este capítulo descreve o teste de Rice, apresentando comparações entre este e o FBST.

### 6.1 Popper & Rice

Visando especificar uma função de perda que refletisse a filosofia popperiana, Rice (2010) considerou uma adaptação para a interpretação do espaço de decisões. Nesta,  $d_0$  representa a decisão “Nada a declarar”, enquanto que, para  $d_1$ , fornecemos resultados que indicam a rejeição de  $H_0$ . Ao optar por fazer tal declaração, as conclusões são apresentadas através de uma estimativa  $\hat{\theta}$ , e a referida perda é avaliada por meio da perda quadrática usual. Por outro lado, ao omitir os resultados, a perda se dá em termos dos desdobramentos que deixaram de ser conhecidos, e de quão informativos estes seriam. Este prejuízo é representado de modo proporcional à distância entre  $\theta$  e  $\theta_0$ , de modo que ambas as funções estarão na mesma escala e, portanto, podem ser tratadas diretamente. Além disso, o agente decisor irá especificar a sua opinião quanto à relação entre estas consequências, por meio de um fator  $\gamma$ . Assim teremos,

**Definição 7** A função de perda desenvolvida por Rice,  $L^R : \mathcal{D} \times \Omega \rightarrow \mathbb{R}^+$ , com  $\Omega \subset \mathbb{R}$  e  $d \in \Omega$ , é dada por:

$$\frac{\Theta \in \Omega}{\left| \begin{array}{cc} \text{Nada a declarar} & \text{Rejeito } H_0 \\ \gamma^{1/2} (\theta_0 - \theta)^2 & \gamma^{-1/2} (d - \theta)^2 \end{array} \right.}$$

Logo, segundo a regra de Bayes, reportaremos resultados ( $\varphi_{\pi}^R(x) = d_1$ ) se,

$$R(\Omega_0, x) = \frac{\mathbb{E}_{\pi(\cdot|x)}^2[\Theta - \theta_0]}{\text{Var}_{\pi(\cdot|x)}[\Theta]} \geq \frac{1-\gamma}{\gamma}.$$

Em termos práticos, rejeitaremos  $H_0$  caso a estimativa esteja “distante” do valor testado, enquanto que, para valores próximos, não teremos conclusões. O receio relacionado às declarações, característico desta leitura, é representado pelo valor  $\gamma$ , de modo que quanto menor for esta quantidade, mais cética será a análise. Além da penalidade por reportar resultados nunca ser inferior à alternativa. Notando ainda que o produto dos pesos é fixo, segue na Tabela 6.1 o efeito para alguns valores  $\gamma$ .

Quanto à função de perda, teremos uma correspondência entre as unidades de imprecisão da estimativa  $d$  e a perda inerente à falta de resultados. Sendo esta última mensurada segundo a

$\gamma$	Penalidade por Reportar	Penalidade por Não reportar	Inverso da Variabilidade
0,01	10	0,100	99
0,1	3,162	0,316	9
0,3	1,826	0,548	2,333
0,5	1,414	0,707	1
0,7	1,196	0,837	0,429
0,99	1,005	0,994	0,010

**Tabela 6.1:** *Efeitos para diferentes  $\gamma$ .*

sua distância em relação à  $\theta_0$ . De modo que o agente deverá procurar o ponto de equilíbrio entre estas entidades. Cabe ainda notar que a conclusão é tomada com base no inverso de uma medida de variabilidade, similar ao quadrado do coeficiente de variação, porém com foco no espaço paramétrico a posteriori. Assim, quando tal medida de variabilidade for expressiva, teremos indicativos sobre uma falta de acurácia e, conseqüentemente, os resultados não serão reportados.

Na leitura feita por Rice, identificamos vários ganhos relevantes em relação ao teste de Fisher. Como exemplo: falta de suposições quanto a repetibilidade e regras de parada do experimento, incorporação obrigatória da hipótese alternativa, coerência entre a rejeição da hipótese e seus subconjuntos e resultados informativos mesmo para grandes amostras. Quanto ao último ponto, a relação entre as conclusões e o tamanho da amostra é um ponto particularmente problemático no caso do  $p$ -valor e, na perspectiva de Rice, tal característica será observada apenas em cenários específicos. No caso, quando obtivermos  $\theta$  e  $\theta_0$  muito próximos, a evidência contra a hipótese nula crescerá proporcionalmente ao tamanho da amostra. Em contraponto, especificando valores altos para  $\gamma$ , estaremos pouco inclinados a concluir nada e, portanto, o ceticismo devido ao tamanho da amostra será cancelado pelas evidências a favor da hipótese. Por fim, tal abordagem permite incorporar o conhecimento subjetivo prévio, além de trabalhar com conclusões provenientes do espaço paramétrico, circunstâncias estas muito mais próximas à realidade.

No contexto inferencial, em que o objetivo é aprender sobre o parâmetro, não reportar resultados soa pouco apropriado. Contudo, o comportamento cauteloso, nada mais é do que uma característica da filosofia da qual faz parte. Perspectiva que é perfeitamente válida para determinados cenários. Como ilustração, considere a leitura de uma ultrassonografia fetal, visando a especificação do sexo da criança. Neste caso, se for possível identificar o membro masculino, rejeitamos a hipótese do bebê ser do sexo feminino. Caso contrário, talvez não seja possível fazer qualquer alegação quanto a amostra observada. Contudo, se a priori for especificada com grande crença em  $H_0$ , e o valor de  $\gamma$  não for demasiadamente pequeno, teríamos a divulgação dos resultados. Um caso mais complexo, tratando de marcadores bialélicos, pode ser consultado em Rice (2010).

Um ponto curioso sobre a modelagem da filosofia “fisheriana”, segundo a perspectiva da teoria da decisão, trata-se da ironia científica envolvida: Sir Ronald Fisher dava pouco crédito à análise de decisões, argumentando ser uma técnica inútil para cientistas e aproveitável apenas para comerciantes - conjectura pouco próxima à realidade (Souza, 2002), além de ser uma opinião matematicamente e epistemologicamente absurda.

## 6.2 FBST $\times$ Rice

Apesar destas metodologias compartilharem o mesmo princípio, as abordagens se dão de formas diferentes. Enquanto que o teste de Rice tem como preocupação central reportar ou não a rejeição da hipótese, o FBST mensura a consistência desta. Assim, no primeiro caso temos o enfoque em termos de distância, enquanto que no segundo, priorizamos a análise da “surpresa”, dado em termos do conjunto tangente.

Em ambas análises mantemos o Princípio da Verossimilhança, além de não ser necessário atribuir massa de probabilidade positiva à hipótese nula. Contudo, enquanto o FBST pode ser utilizado

para qualquer dimensão, o teste de Rice só é aplicável para o caso do parâmetro estar na reta real. Em contraponto, a função de perda do Rice não depende de  $x$  explicitamente, diferentemente do FBST. Apesar deste último ponto não ser exatamente uma vantagem, dado que não há limitações teóricas/práticas para tal característica, é fato que existe uma falta de destreza com tal classe.

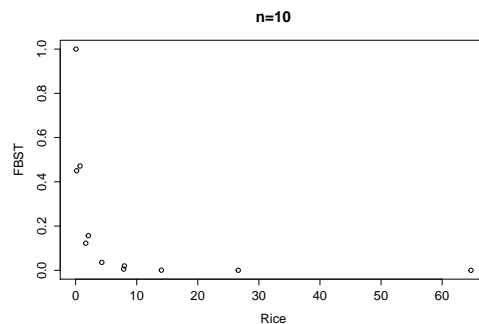
Em termos práticos, o teste de Rice não apresenta uma medida de evidência a favor (nem contra) à hipótese, enquanto que o FBST não utiliza estimativas pontuais  $\mathbb{E}_{\pi(\cdot|x)}[\Theta]$  - mas poderia. No entanto, quando  $\Omega \subset \mathbb{R}$ , os testes são essencialmente equivalentes, no sentido frequentista de existir uma relação uma a um entre as estatísticas de teste:

**Exemplo:** Considere uma amostra aleatória c.i.i.d.<sup>1</sup> Bernoulli's, com parâmetro  $\theta$ . Supondo o interesse em testar as hipóteses  $H_0 : \theta = 0.3$  versus  $H_0 : \theta \neq 0.3$ , para uma priori Beta(1, 1), e uma amostra de tamanho 10, toda a informação das  $2^{10}$  amostras possíveis pode ser examinada por meio do número de sucessos obtidos. Assim, na Tabela 6.2, temos os valores obtidos para as estatísticas de Rice e FBST, respectivamente.

# Sucessos	Rice	FBST
	$\varphi_{\pi}^R(x) = d_1$	$\varphi_{\pi}^E(x) = d_0$
0	7,989	0,020
1	1,664	0,123
2	0,173	0,450
3	0,065	1,000
4	0,728	0,471
5	2,080	0,156
6	4,294	0,036
7	7,865	0,006
8	14,040	0,001
9	26,624	0,000
10	64,716	0,000

**Tabela 6.2:** Exemplo:  $R(\Omega_0, x) \times EV(\Omega_0, x)$

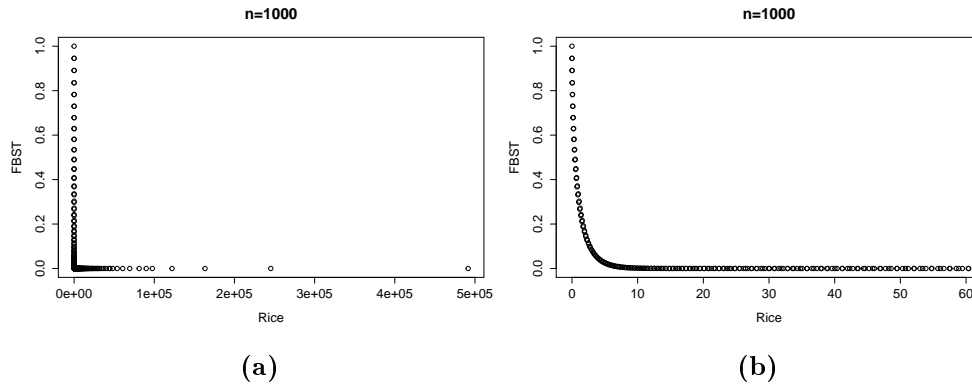
Apesar da diferença de magnitudes, podemos observar uma coerência entre ambos os resultados e  $H_0$ . Por meio do Gráfico 6.1, temos também que tais quantidades são negativamente associadas, conforme esperado, dado que estamos comparando decisões opostas. Em termos de referência, optaríamos por reportar os resultados, independente da amostra, no caso de um  $\gamma$  inferior a 0,015.



**Figura 6.1:** Exemplo ( $n = 10$ ):  $R(\Omega_0, x) \times EV(\Omega_0, x)$

Visando identificar melhor a relação neste caso, faremos a mesma análise para uma amostra de tamanho  $n = 1.000$  (Gráfico 6.2a). Tendo, entretanto, que existe uma diferença considerável entre as escalas, no Gráfico 6.2b seguem os resultados para o teste de Rice limitado a 60. O exame dos gráfico revela imediatamente a equivalência entre os dois testes.

<sup>1</sup>Condicionamente independente e identicamente distribuída, dado  $\Theta = \theta$ .



**Figura 6.2:** *Exemplo* ( $n = 1000$ ):  $R(\Omega_0, x) \times EV(\Omega_0, x)$

Por fim, esta classe de testes tem como característica comum o princípio jurídico *Onus Probandi*, em que o réu deve ser presumido como inocente (Stern, 2003). Tal compreensão é interessante para o caso do analista procurar por análises conservadoras, considerando, por exemplo, que a contradição da hipótese testada não é absoluta, a amostra se apresenta como uma ferramenta demasiadamente limitada, etc. Além de que, independente do cenário em questão, o analista pode, simplesmente, se identificar com uma postura mais cautelosa.

## Capítulo 7

# Conclusão

Neste trabalho discutimos as principais abordagens dos testes de hipóteses e testes de significância, particularmente no que diz respeito às hipóteses precisas, identificando as ideias elementares de cada alternativa. Com o levantamento da estrutura de cada teste, revisamos a leitura destes segundo a teoria da decisão estatística, considerando as respectivas funções de perda associadas. Assim, por meio de uma abordagem normativa, em que os objetivos do estudo são descritos pelas funções de perda (subjetivas) do pesquisador, passamos a contemplar diferentes objetivos, cenários e princípios, além das conclusões focarem nas necessidades da pesquisa e não da metodologia em si. De modo que a significância estatística e a significância prática possam representar uma única unidade.

Quanto ao entendimento de probabilidade, necessário para a teoria da decisão, consideramos a perspectiva bayesiana, uma vez que a abordagem frequentista: dispõe de conclusões que não atendem as reais necessidades da ciência, trabalha com suposições pouco próximas à realidade, e permite cenários incoerentes.

Para as hipóteses precisas, os testes de significância passam a possuir duas opções satisfatórias do ponto de vista de coerência: FBST e Rice. Adicionalmente, tanto o conhecido FBST quanto a recente proposta do Rice - que engloba a perspectiva de Popper sobre não aceitar hipóteses - passaram a ser suportadas pelo fator de Bayes. Em relação ao FBST, concluímos que a função de perda pode ser especificada segundo as mesmas estruturas dos erros tipo I e II. Considerando, entretanto, um elemento adicional, responsável por incorporar a posição conservadora do analista quanto a rejeição de  $H_0$ . Por fim, ao compararmos ambas leituras, obtemos resultados harmoniosos com as respectivas propostas.

Para outros trabalhos, a grande proposta é estender o teste de Rice para espaços paramétricos de dimensões maiores que um, além de, por meio de exemplos, corroborar a relevância das funções de perda que contemplam a amostra. Adicionalmente, entender melhor a proposta de outras escolas, sob a ótica da teoria da decisão, pode se mostrar útil para entender a relação entre as leituras probabilísticas atuais.

“We **feel** that using the Bayesian paradigm will give misleading answers less often than use of alternative paradigms.” Berger e Delampady (1987b).



# Referências Bibliográficas

- Aitkin et al. (2005)** Murray Aitkin, Richard J. Boys e Tom Chadwick. Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing*, 15(3):217–230. Citado na pág. 17
- Berger (2003)** James O. Berger. Could Fisher, Jeffreys and Neyman have agreed on testing? *Statistical Science*, 18(1):1–12. Citado na pág. 10
- Berger e Berry (1988)** James O. Berger e Donald A. Berry. Statistical analysis and the illusion of objectivity. *American Scientist*, 76:159–165. Citado na pág. 1
- Berger e Delampady (1987a)** James O. Berger e Mohan Delampady. Testing precise hypotheses. *Statistical Science*, 2(3):317–352. doi: 10.1214/ss/1177013238. Citado na pág. 8
- Berger e Delampady (1987b)** James O. Berger e Mohan Delampady. [testing precise hypotheses]: Rejoinder. *Statistical Science*, 2(3):348–352. doi: 10.1214/ss/1177013245. Citado na pág. 27
- Berger e Sellke (1987)** James O. Berger e Thomas Sellke. Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, 82(397):112–122. Citado na pág. 17
- Bernardo e Smith (1994)** José M. Bernardo e Adrian F. M. Smith. *Bayesian Theory*. John Wiley & Sons, New York. ISBN 0-471-92416-4. Citado na pág. 21
- Berry (1997)** Donald A. Berry. Teaching elementary bayesian statistics with real applications in science. *The American Statistician*, 51(3):241–246. Citado na pág. 2
- Birnbaum (1962)** Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–326. Citado na pág. 1
- Christensen (2005)** Ronald Christensen. Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59:121–126. URL <http://EconPapers.repec.org/RePEc:bes:amstat:v:59:y:2005:m:may:p:121-126>. Citado na pág. 10
- Cox (1977)** D. R. Cox. The role of significance tests. *Scandinavian Journal of Statistics*, 4(2):49–70. URL <http://www.jstor.org/stable/4615652>. Citado na pág. 13
- DeGroot (1970)** Morris Herman DeGroot. *Optimal Statistical Decisions*. McGraw-Hill. ISBN 0-471-68029-X. Citado na pág. 18
- Fisher (1925)** Ronald Aylmer Fisher. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd. Citado na pág. 8
- Fisher (1935)** Ronald Aylmer Fisher. *The design of experiments*. Edinburgh: Oliver and Boyd. Citado na pág. 8
- Fisher (1956)** Ronald Aylmer Fisher. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd. Citado na pág. 8

- Jeffreys (1961)** Harold Jeffreys. *The theory of probability*. Oxford University Press., 3 ed. Citado na pág. 10, 11, 14
- Kadane (2011)** Joseph B. Kadane. *Principles of Uncertainty*. Chapman & Hall. Citado na pág. 2
- Lindley (1957)** Dennis V. Lindley. A statistical paradox. *Biometrika*, 44(1-2):187–192. Citado na pág. 18
- Loschi e Wechsler (2002)** Rosangela H. Loschi e Sergio Wechsler. Coherence, bayes's theorem and posterior distributions. *Brazilian Journal of Probability and Statistics*, 16:169–185. Citado na pág. 2
- Madruga et al. (2001)** Maria Regina Madruga, Luis Gustavo Esteves e Sergio Wechsler. On the bayesianity of pereira-stern tests. *Sociedad de Estadística e Investigación Operativa*, 10:291–299. Citado na pág. 20, 23
- Madruga et al. (2003)** Maria Regina Madruga, Carlos Alberto de Bragança Pereira e Julio Michael Stern. Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference*, 117:185–198. Citado na pág. 15
- Mosteller e Youtz (1990)** Frederick Mosteller e Cleo Youtz. Quantifying probabilistic expressions. *Statistical Science*, 5(1):2–12. doi: 10.1214/ss/1177012242. Citado na pág. 2
- Parmigiani e Inoue (2009)** Giovanni Parmigiani e Lurdes Y. T. Inoue. *Decision Theory: Principles and Approaches*. John Wiley & Sons. Citado na pág. 17
- Patriota (2013)** A.G. Patriota. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*. URL <http://dx.doi.org/10.1016/j.fss.2013.03.007>. Citado na pág. 8
- Pearson (1900)** Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175. doi: 10.1080/14786440009463897. Citado na pág. 8
- Pereira e Stern (1999)** Carlos Alberto de Bragança Pereira e Julio Michael Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110. ISSN 1099-4300. doi: 10.3390/e1040099. URL <http://www.mdpi.com/1099-4300/1/4/99>. Citado na pág. 14, 23
- Pereira e Wechsler (1993)** Carlos Alberto de Bragança Pereira e Sergio Wechsler. On the concept of p-value. *Revista Brasileira de Probabilidade e Estatística*, 7:159–177. Citado na pág. 8
- Pereira et al. (2008)** Carlos Alberto de Bragança Pereira, Julio Michael Stern e Sergio Wechsler. Can a significance test be genuinely bayesian. *Bayesian Analysis*, 3(3):79–100. Citado na pág. 15
- Pericchi e Pereira (2013)** Luis Raul Pericchi e Carlos Alberto de Bragança Pereira. Changing the paradigm of fixed significance levels: Testing hypothesis by minimizing sum of errors type i and type ii. *Brazilian J Probability & Statistics*. Citado na pág. 9, 18
- Rice (2010)** Kenneth Rice. A decision-theoretic formulation of fisher's approach to testing. *The American Statistician*, 64(4):345–349. Citado na pág. 23, 24
- Souza (2002)** Fernando Menezes Campello de Souza. *Decisões racionais em situações de incerteza*. : Universitária da UFPE, Recife. Citado na pág. 3, 24
- Stern (2003)** Julio Michael Stern. Significance tests, belief calculi, and burden of proof in legal and scientific discourse. *In Frontiers in Artificial Intelligence and Applications*, 101:139–147. Citado na pág. 26



- Thulin (2014)** Måns Thulin. Decision-theoretic justification for bayesian hypothesis testing using credible sets. *Journal of Statistical Planning and Inference*, (146):133–138. Citado na pág. 21
- Tukey (1991)** John W. Tukey. The philosophy of mutiple comparisons. *Statistical Science*, 6(1): 100–116. Citado na pág. 13
- van der Pas (2010)** Stéphanie L. van der Pas. Much ado about the p-value. Bachelor thesis, *Mathematisch Instituut, Universiteit Leiden*. Citado na pág. 9
- Wagenmakers (2007)** Eric-Jan Wagenmakers. A practical solution to the pervasive problems of *p values*. *Psychonomic Bulletin & Review*, 14(5):779–804. Citado na pág. 8