Cálculo do Tamanho Amostral
Algumas Considerações Práticas

José Luiz Padilha da Silva
Paulo Cerqueira dos Santos Júnior
Rodrigo Citton Padilha dos Reis

Departamento de Estatística
Universidade Federal de Minas Gerais

Belo Horizonte, setembro de 2011

Sumário

Introdução

Estudos estatísticos são sempre melhores quando são cuidadosamente planejados.

- O problema deve ser cuidadosamente definido e operacionalizado.
- As unidades experimentais ou observacionais devem ser selecionadas da população apropriada.
- A aleatorização deve ser feita corretamente.
- Os procedimentos devem ser seguidos corretamente.
- Instrumentos confiáveis devem ser usados para obter as medidas.

Introdução

Finalmente o estudo deve ser de tamanho adequado.

- Deve ser "grande suficiente" para que um efeito de tal magnitude que seja de significância científica seja também de significância estatística.
- Importante, também, que não seja "tão grande" para que um efeito de pouca significância científica seja estatisticamente detectável.

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

- O tratamento de uma enfermidade muito comum, conhecida como *elbowpain*, era habitualmente feito com o medicamento Hopefull, que proporciona cura a aproximadadamente **metade dos casos**.
- Uma droga recém descoberta em 1957, temporariamente chamada de L3OV5E, foi produzida para ser testada num **ensaio clínico casualizado contra o Hopefull**. O pesquisador responsável procurou um potencial financiador do estudo, Mr. Dupont, que se interessou pelo projeto.
- O pesquisador explicou que trabalharia com amostra (**40 pessoas, 20 em cada grupo**), que iria sortear quais pessoas iriam receber Hopefull ou L3OV5E, e que utilizaria um teste estatístico (5% n.s.) para decidir a respeito da eficácia do novo remédio ... Mr. Dupont deu-se por satisfeito e financiou o estudo.

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

SIGNIFICAÇÃO CLÍNICA SEM SIGNIFICÂNCIA ESTATÍSTICA

Tabela: Resultados do ensaio clínico.

	Cura			
	Sim	Não	Total	Proporção de Curas
Hopefull	10	10	20	0,50
L3OV5E	15	5	20	0,75
Total	25	15	40	0,625

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

- Mr. Dupont ficou muito impressionado com os resultados! Afinal, o L3OV5E tinha proporcionado 75% de curas contra os 50% do Hopefull...
- O pesquisador fez um teste qui-quadrado, pegou sua tábua de distribuição de probabilidades do qui-quadrado e esta forneceu $p > 0,05$.
- "Infelizmente a diferença não é estatísticamente significante ... vamos ter de aceitar que essa diferença não existe na população da qual a amostra foi retirada."

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

SIGNIFICÂNCIA ESTATÍSTICA SEM SIGNIFICAÇÃO CLÍNICA

- Ah, quantas voltas o mundo dá... O tempo passou e Mr. Dupont fundou sua própria indústria farmacêutica, investindo pesado na descoberta de uma outra droga para o tratamento da *elbowpain* (mal que agora também sofria). Aquele ano de 1959 trouxe a esperança do H2AT4E.
- Ainda traumatizado pelo dissabor de 1957, Mr. Dupont resolveu financiar um **ensaio clínico casualizado para testar o H2AT4E contra o tradicional Hopefull**.
- Para não correr o risco de ter que rejeitar uma diferença clinicamente significativa, Mr. Dupont pensou bem grande: o estudo teria 1.000 pacientes em cada grupo!

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

Tabela: Resultados do ensaio clínico.

	Cura			
	Sim	Não	Total	Proporção de Curas
Hopefull	500	500	1.000	0,50
H2AT4E	550	450	1.000	0,55
Total	1.050	950	2.000	0,525

Introdução

Exemplo: Uma história tão inverossímil que poderia ser verdadeira...

- Mr. Dupont calculou o qui-quadrado, consultou a tábua de probabilidades e, com a felicidade de uma criança, verificou que a probabilidade (de a diferença ter sido casual) era menor que 5%.
- Sim, Mr. Dupont obtivera uma diferença estatisticamente significante! O H2AT4E foi produzido em larga escala com o nome comercial de *Painless*.
- ...
- Poucos hoje recordam desse medicamento, retirado do mercado em 1960. O *Painless* era mais caro que o Hopefull e tinha que ser tomado em jejum, provocando náuseas em muitas pessoas. Essas desvantagens do *Painless* não compensavam a **pequena diferença na proporção de curas** que, convenhamos, **não era clinicamente significativa**.

Introdução

O tamanho amostral é importante por razões econômicas:

- Um estudo com tamanho amostral menor que o necessário pode ser um desperdício de recursos por não ter a capacidade de produzir resultados úteis.
- Um estudo com tamanho amostral exagerado usa mais recursos que o necessário.
- Num experimento envolvendo seres humanos ou animais, o tamanho amostral é questão de grande importância por razões éticas.
  - Com tamanho amostral exagerado, um número desnecessário de indivíduos é exposto a um tratamento potencialmente perigoso, ou não são beneficiadas de um potencialmente benéfico.

Introdução

Para tal importante problema há surpreendente pouca literatura.

Há vários enfoques para o tamanho amostral:

- Pode-se especificar o comprimento desejado do intervalo de confiança e determinar o tamanho amostral que atende tal meta;
- Um dos enfoques mais populares envolve estudar o poder de um teste de hipóteses.

Esse é o enfoque que será utilizado aqui.

Introdução

O enfoque baseado no poder envolve os seguintes elementos:

1. Especifique um teste de hipóteses sobre um parâmetro θ (junto com um modelo probabilístico para os dados).
2. Especifique um nível de significância α do teste.
3. Especifique um *tamanho de efeito* $\tilde{\theta}$ que seja de interesse científico.
4. Obtenha valores históricos ou estimativas dos outros parâmetros necessários para cálculo da função poder do teste.
5. Especifique um valor alvo $\tilde{\pi}$ do poder do teste quando $\theta = \tilde{\theta}$.

Introdução

O poder do teste é uma função $\pi(\theta, n, \alpha, \ldots)$, em que:

- n é o tamanho amostral, e
- e a parte "$\ldots$" refere-se aos parâmetros adicionais do passo 4.

O tamanho amostral necessário é o menor inteiro n tal que $\pi(\tilde{\theta}, n, \alpha, \ldots) \geq \tilde{\pi}$.

Exemplo

Suponha que planejamos conduzir um experimento de duas amostras para comparar um tratamento com um controle.

A variável resposta é pressão sanguínea sistólica, medida com um esfigmomanômetro padrão.

Espera-se que o tratamento reduza a pressão sanguínea;

Temos um teste unilateral: $H_0 : \mu_T = \mu_C$ versus $H_1 : \mu_T < \mu_C$, em que μ_T é a pressão média para o grupo tratamento e μ_C é a pressão média para o grupo controle.

O parâmetro $\theta = \mu_T - \mu_C$ é o efeito a ser testado; e escrevemos $H_0 : \theta = 0$ e $H_1 : \theta < 0$.

Exemplo

As metas do experimento especificam que queremos detectar uma situação em que a média do tratamento é 15 mm Hg menor que a do grupo controle; i.e., o tamanho do efeito é $\tilde{\theta} = -15$.

Tal efeito deve ser detectado com 80% de poder ($\tilde{\pi} = 0,80$) e nível de significância $\alpha = 0,05$.

Experiência passada com experimentos similares sugere que os dados sejam normalmente distribuídos com $\sigma = 20$ mm Hg.

Usaremos um teste t para duas amostras (com variância combinada) e n igual para cada grupo.

Fórmula para Cálculo

$$n_1 = kn_2$$

e

$$n_2 = \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2(1 + 1/k)}{\tilde{\theta}^2}$$

Tamanho Amostral e Poder

Com essas especificações é necessário um tamanho amostral de $n = 23$ por grupo. O poder real é de $0,8049$.

Na definição do poder do teste, queremos ter uma chance razoável de detectar o tamanho do efeito estabelecido. Um valor de $0,80$ é muito comum – alguns autores sugerem poder maior, como $0,85$ ou $0,90$.

À medida que o poder aumenta, contudo, o tamanho amostral aumenta numa taxa crescente. No exemplo, um poder de $\tilde{\pi} = 0,95$ necessita de um tamanho amostral de $n = 40$ – cerca de 75% mais do que é necessário para um poder de $0,80$.

Tamanho Amostral e Poder

Algumas dificuldades...

- Quem nos disse que a meta era detectar uma diferença média de 15 mm Hg?
- Quem nos disse que $\sigma = 20$, se estamos apenas planejando o experimento e os dados nem foram coletados ainda?

Tais *inputs* do tamanho amostral são geralmente complicados...

...obter um tamanho de efeito de importância científica requer o conhecimento do pesquisador responsável pelo estudo. Por outro lado, há detalhes técnicos que requerem a habilidade de um estatístico.

Tamanho do Efeito

Importante no problema do tamanho amostral é definir um tamanho de efeito de interesse científico.

Essa é uma tarefa do pesquisador envolvido no estudo.

O problema é que o pesquisador nem sempre sabe o que está sendo perguntado, ou não reconhece como uma questão que seja de sua responsabilidade responder.

P: "Qual diferença seria importante detectar com 90% de poder usando um teste *t* Satterthwaite com $\alpha = 0,05$?"
R: "Você é o estatístico – o que recomenda?" ou "Qualquer diferença seria importante."

Questões concretas

P: “Quais resultados você espera ver?”

Podemos ter um limite *superior* para $\tilde{\theta}$.

Podemos estabelecer um limite *inferior* para o tamanho amostral.

P: “Um efeito de metade dessa magnitude seria de interesse científico?”

Mas metade de $\tilde{\theta}$ aproximadamente vai quadruplicar o tamanho amostral. . .

Vários cálculos de n para várias propostas pode ajudar

Podemos tentar uma seleção de tamanhos de efeito e seu poder correspondente, ex., "Com 25 observações, teremos 50% de chance de detectar uma diferença de $9,4$ mm Hg, e 90% de chance de detecar uma diferença de 16.8 mm Hg."

"Se você pode bancar mais 6 indivíduos por tratamento, poderá detectar uma diferença de 15 mm Hg com 90% de poder".

E o que *não* pode ser detectado:
P: "Qual a variação de irrelevância clínica?"

P: "Se você fosse o paciente, os benefícios de se reduzir a pressão em 15 mm Hg compensariam o custo, inconveniência, e potenciais efeitos colaterais desse tratamento?"

Tamanho do Efeito

A discussão da relação entre o tamanho da amostral e o tamanho do efeito requer tanto as habilidades técnicas do estatístico e o conhecimento científico do pesquisador.

As metas científicas e questões éticas devem ser ambas discutidas.

A discussão dos valores éticos envolve todos, incluindo pesquisadores, estatísticos, e técnicos de laboratório. . .

Escolha da Variância

Funções poder envolvem parâmetros não relacionados com as hipóteses. Geralmente envolvem uma ou mais variâncias.

No exemplo queremos saber a variância residual das medidas no experimento de duas amostras.

Opções são buscar a variância da experiência do pesquisador, ou usar dados históricos, ou conduzir um estudo piloto. Pode-se construir um histograma mostrando como se espera que os dados sejam distribuídos (95% dos dados correspondem a ± 2 desvios padrões, no caso normal).

"Qual a variação natural da pressão? Quais são os maiores e menores valores de pressão que você já viu?"

Escolha da Variância

Dados históricos ou de uma amostra piloto não precisam seguir o mesmo desenho do estudo planejado; mas deve-se tomar cuidado para que a variância correta seja estimada.

Exemplo: o fabricante do esfigmomanômetro pode ter publicado resultados de teste mostrando que o desvio padrão das leituras é de $2,5$ mm Hg.

Esse número não é apropriado para uso na determinação do tamanho amostral: reflete variações nas leituras feitas no mesmo indivíduo sob condições idênticas.

A variação residual no experimento da pressão inclui a variação entre indivíduos!

Escolha da Variância

A determinação e consideração das fontes de variação em estudos passados é bastante importante.

Incluem: atributos do paciente (sexo, idade, fatores de risco, demográficos, etc.), instrumentos, como, quando e quem administra os medicamentos e coleta os dados, e outros fatores.

Num estudo de um único fator, suponha que temos dados passados de um experimento de dois fatores em que homens e mulheres foram aleatorizados separadamente a grupos que receberam diferentes regimes de exercícios; e que a resposta é a pressão medida usando instrumentos idênticos àqueles que planejamos usar.

Escolha da Variância

Isso nos fornece dados úteis para planejar o novo estudo – mas temos que ter cuidado.

Exemplo: a variância residual do estudo anterior não inclui variações devido ao sexo.

Se o novo estudo inclui indivíduos de ambos os sexos, então a variação devido ao sexo deve ser incluída na variância do erro a ser usada no planejamento do tamanho amostral.

A mesma pessoa toma todas as medidas? Se isso é feito por várias pessoas – o treinamento delas é comparável?

Todos esses fatores podem afetar a variância do erro!

Escolha da Variância

Uma vez que os dados são coletados, é útil comparar as variâncias de fato observadas com aquelas usadas nos cálculos do tamanho amostral.

Isso não ajudará no desenho do estudo atual, mas é útil como parte do processo de aprendizagem para o desenho de estudos futuros.

Grandes discrepâncias devem ser estudadas para tentar identificar o que foi desconsiderado...

Quando não há escolha. . .

Frequentemente, um estudo tem orçamento limitado, e isso determina o tamanho amostral.

Outra situação comum é que um pesquisador sênior (ou toda a área de pesquisa) tem estabelecido alguma convenção sobre quanto é "suficiente".

É difícil argumentar com orçamentos, editores de jornal, e superiores. . .

O tamanho amostral é uma das características de um estudo estatístico; assim, se n é fixado, precisamos nos ater em outros aspectos da qualidade do estudo.

Quando não há escolha...

Por exemplo, dado o tamanho amostral que pode ser bancado (ou imposto), podemos encontrar o tamanho do efeito $\ddot{\theta}$ tal que $\pi(\ddot{\theta}, n, \alpha, \ldots) = \tilde{\pi}$.

Então o valor de $\ddot{\theta}$ pode ser discutido e avaliado com relação às metas científicas.

Se for muito grande, então o estudo tem baixo poder...

Talvez isso possa ser usado para pedir um orçamento maior. Talvez um melhor instrumento possa ser encontrado a fim de melhorar as medições.

Por fim, podem ser feitas melhoras no delineamento do estudo para reduzir a variância do estimador de θ, ex., usando estratificação ou blocagem...

Quando não há escolha...

Dizer que o estudo não possa ser feito é uma mensagem nunca bem vinda (se não totalmente inapropriada).

As melhores alternativas são que o escopo do estudo seja estreitado (ex., mais fatores fixados), ou que seja proposto como parte de uma sequência de estudos.

Não é porque o tamanho amostral é fixo que não haja outros aspectos que possam ser mudados no delineamento do estudo.

Quando não há escolha. . .

É mesmo possível que $\ddot{\theta}$ seja *menor* que o necessário – assim o estudo tem poder exagerado.

Pode-se também manter o tamanho amostral fixado, mas alargar o escopo do estudo (maior diversidade demográfica dos indivíduos, outros fornecedores de matéria-prima, etc.); isso fará os resultados mais amplamente aplicáveis. . .

Quando animais ou pessoas são envolvidos, um estudo com poder exagerado levanta um sério dilema ético.

Nem todos os problemas são iguais

Nem todos os problemas de tamanho amostral são os mesmos, nem o tamanho amostral tem a mesma importância em todos os estudos.

Exemplo: questões éticas numa pesquisa de opinião são muito diferentes daquelas num experimento médico, e as consequências de um estudo com muito ou pouco poder também são diferentes.

Na indústria, leva-se apenas minutos para se obter os dados, havendo poucas consequências se o tamanho amostral é muito pequeno.

Um estudo clínico pode ter relativamente curta duração e envolver risco potencial aos pacientes. Pode ser desejável proceder uma sequência de pequenos experimentos, com análises entre eles.

Nem todos os problemas são iguais

Questões de tamanho amostral são geralmente mais importantes quando tomam muito tempo para coleta dos dados.

Um experimento agronômico pode requerer toda uma estação, ou mesmo uma década, para ser concluído.

Se o tamanho amostral não é adequado, há graves consequências.

Assim, torna-se mais importante planejar cuidadosamente, e colocar grande ênfase na possibilidade de subestimar a variância, já que causaria subestimar o tamanho amostral.

Nem todos os problemas são iguais

Parte da discussão sobre o planejamento do tamanho amostral deve girar em torno das consequências do que pode dar errado:

- E se quisermos um estudo de acompanhamento?
- Quanto isso nos custará?
- Podemos bancar?
- Quais são as questões éticas num estudo muito grande ou muito pequeno?

As respostas não ajudarão a decidir quão liberais ou conservadores precisamos ser no cálculo do tamanho amostral.

Nem todos os problemas são iguais

Os problemas também variam em sua complexidade...

Se esperamos dados normalmente distribuídos... podemos usar tabelas ou algum software.

Se a análise é uma análise de variância multifatorial de efeitos mistos, há muitos testes e componentes de variância a serem considerados...

Se há não resposta substancial, censura... o único recurso pode ser simulação de vários cenários plausíveis para se ter uma ideia de quão bom seja o estudo proposto.

Nem todos os problemas são iguais

Existem complicações adicionais para dados de atributo, devido a falhas de testes assintóticos, incapacidade de alcançar o tamanho estabelecido devido ao dados serem discretos, ou situações incomuns tais como inferências sobre atributos raros.

Métodos de simulação são novamente úteis na discussão desses problemas.

Finalmente, não há de fato apenas um problema de tamanho amostral.

Nem todos os problemas são iguais

O tamanho amostral é apenas um aspecto do desenho do estudo.

Muitas questões devem ser perguntadas e respondidas antes de chegar ao n:

- Quais são exatamente as metas?
- Qual é a variável resposta, como são planejadas as medições, e há instrumentos alternativos?
- O que pode sair errado?
- Qual sua taxa estimada de não resposta?
- Quais são as fontes importantes de variação?
- Qual é o tempo disponível?
- Quais são as outras restrições práticas?

Tamanhos de Efeito Pré-definidos

Uma das práticas a serem evitadas é o comum mau uso das medidas de tamanho de efeito descritas em Cohen (1988).

Para um teste t com variâncias combinadas, Cohen define o tamanho do efeito d como a diferença das médias dividido pelo desvio padrão do erro (i.e., $d = \tilde{\theta}/\sigma$).

Esse tamanho do efeito d é *padronizado* porque não depende de escala, comparado com um tamanho de efeito absoluto como $\tilde{\theta}$ que carrega unidades (tais como mm Hg).

Tamanhos de Efeito Pré-definidos

Cohen sugere classificações para d: é "pequeno," "médio," ou "grande" se d é $0,20$, $0,50$, ou $0,80$ respectivamente.

Essas avaliações foram baseadas numa extensiva pesquisa de estatísticas na literatura das ciências sociais.

Muitos pesquisadores têm se guiado mal ao usá-las como meta; ex., encontrar o tamanho amostral necessário para detectar um efeito "médio" com poder de 80%.

Estamos apelando para convenções e evitando falar sobre $\tilde{\theta}$ ou σ – o que não é uma boa ideia!

Exemplo

Num experimento industrial as medidas podem ser feitas usando uma máquina (com precisão de alguns microns), um paquímetro (com precisão de décimos de milímetro), ou uma régua escolar (com precisão de um milímetro).

Independente de qual usar, sempre teremos o mesmo tamanho amostral para um efeito "médio" com poder de 80%.

Obviamente, a escolha do instrumento tem grande impacto nos resultados, e deve afetar os cálculos do tamanho amostral.

Outro Exemplo

Numa regressão linear simples de uma variável y sobre outra x, a correlação (ou correlação ao quadrado) entre x e y pode servir como uma medida de tamanho de efeito padronizado.

Ela carrega três quantidades: inclinação da linha, variância do erro, e a variância dos valores de x.

São, respectivamente, tamanho do efeito absoluto, variância, e delineamento experimental – três aspectos do estudo que enfatizamos.

É preciso que essas três quantidades sejam consideradas separadamente, ao invés de serem confundidas numa única medida R^2.

Evite Planejamento Retrospectivo

Outra forma de tentar contornar a determinação de tamanho do efeito e variâncias é adiar essas questões até que o estudo seja concluído.

Então teremos estimativas de todos os tamanhos de efeitos e variâncias que precisamos, e podemos fazer uma espécie de *cálculo retrospectivo* de tamanho amostral ou análise de poder.

Se o teste se mostra "não significativo" o pesquisador pode querer um estudo de acompanhamento com dados suficientes para que um efeito do tamanho daquele observado seja detectado.

Evite Planejamento Retrospectivo

Em outras palavras, se especifica $\tilde{\theta} = \hat{\theta}$ no desenho do estudo de acompanhamento.

Isso é uma forma de especificação *post hoc* ou retrospectiva do tamanho do efeito. Muito diferente da especificação baseada em metas científicas.

A meta agora é realmente coletar dados suficientes para se obter significância estatística, ignorando o significado científico.

Evite Planejamento Retrospectivo

Outra estratégia popular, igualmente ruim, ao lidar com resultados não significativos: tentar fazer inferências com base no poder com um tamanho de efeito observado:

$$\pi_{obs} = \pi(\hat{\theta}, n, \alpha, \ldots)$$

em que $\hat{\theta}$ é a estimativa de θ.

Essa quantidade é o "poder observado".

Apesar de sua popularidade, o poder observado confunde as coisas; não fornece qualquer informação adicional para além dos resultados de um teste estatístico!

Evite Planejamento Retrospectivo

O principal ponto técnico é que pode ser mostrado que π_{obs} é uma função decrescente do valor p do teste; já sabemos interpretar valores p, assim não precisamos do poder observado.

Uma argumentação comum dos proponentes do poder observado é: se o teste é não significativo mas o poder é alto, então há forte evidência estatística sustentando que H_0 seja verdadeiro.

Contudo, já que o poder observado aumenta à medida que o valor p diminui, alto poder observado constitui evidência *contra* a hipótese nula – o oposto da argumentação dos proponentes.

Evite Planejamento Retrospectivo

O poder observado pode ser usado também de forma que distorce ou exagera os resultados da análise estatística: "Não apenas é significativo, mas o teste é realmente poderoso!" ou: "Os resultados não foram significativos, mas isso porque o teste não é muito poderoso."

A relação entre π_{obs} e valores p mostra que se o teste é não significativo, o poder deve ser alto, e quando não é significativo, o poder deve ser baixo.

(No caso de um teste t, ou outra estatística aproximadamente simétrica, o limiar em que $p = \alpha$ corresponde a $\pi_{obs} \approx 50\%$.)

Evite Planejamento Retrospectivo

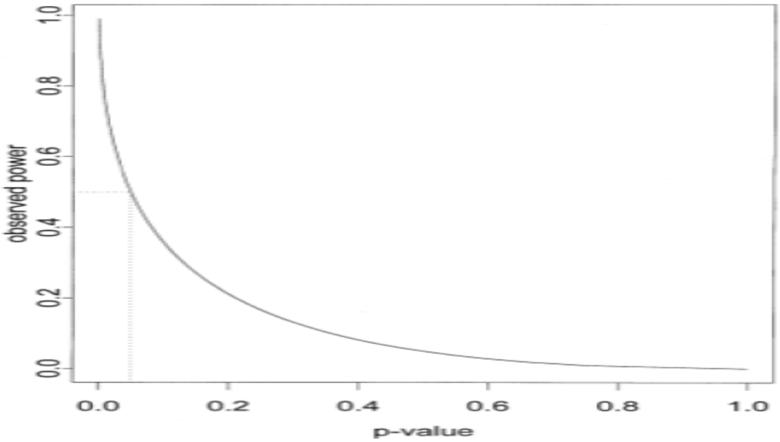

Figura: "Poder observado" como função do valor p para um teste no qual $\alpha = 0,05$. Quando $p = 0,05$ o poder estimado é de 50%.

Evite Planejamento Retrospectivo

Há outro tipo de poder retrospectivo digno de ser mencionado.

Suponha que examinamos o valor de $\pi(\tilde{\theta}, n, \alpha, \ldots)$ depois de coletar os dados (usando os dados para estimar parâmetros auxiliares como o SD do erro).

É diferente do poder observado que usa o tamanho de efeito especificado $\tilde{\theta}$ de significado científico, ao invés do efeito observado $\hat{\theta}$.

Se esse poder retrospectivo é alto num caso em que a hipótese nula não é rejeitada, argumenta-se que se pode estabelecer uma certeza razoável que o tamanho do efeito não seja maior que $\tilde{\theta}$.

Evite Planejamento Retrospectivo

Obviamente, usar o poder retrospectivo para fazer inferência é um caminho complicado para seguir.

A principal fonte de confusão é que ele tende a ser usado para adicionar interpretação a um teste estatístico não significativo; começa-se a contemplar a possibilidade de que $|\theta|$ seja de fato pequeno, e se quer provar isso.

Mas isso implica num teste estatístico diferente! A forma correta de proceder não é olhar no poder do teste original – para o qual as hipóteses foram formuladas inadequadamente – mas fazer um teste formal de equivalência.

Hipóteses $H_0 : |\theta| \geq \tilde{\theta}$ versus $H_1 : |\theta| < \tilde{\theta}$ em que, como antes, $\tilde{\theta}$ é um tamanho de efeito considerado de importância científica.

Conclusões

Planejar o tamanho amostral é frequentemente importante, e quase sempre difícil.

Requer cuidado ao definir os objetivos científicos e ao obter informação quantitativa adequada antes do estudo.

A resolução bem sucedida do problema requer colaboração do estatístico e do especialista da área.

Não se pode evitar de discutir questões de definição do tamanho do efeito (em termos absolutos) e estimar a variância do erro, por mais difícil que possa parecer.

Conclusões

Efeitos padronizados não se traduzem em declarações honestas sobre as metas do estudo.

O poder observado acrescenta pouca informação para a análise, e a determinação retrospectiva do tamanho do efeito muda a atenção par obter resultados significativos independente do significado científico.

Nesses métodos retrospectivos é usado um tamanho de efeito estimado no lugar daquele determinado por questões científicas.

O erro em confundir essas é exatamente o erro feito quando se confunde significância estatística com significância científica.

Conclusões

Uma realidade prática é que nem sempre o tamanho amostral é determinado pelas metas científicas.

É importante, então, avaliar o estudo proposto para ver se atenderá os padrões científicos.

Vários tipos de mudança ao estudo podem ser recomendadas se ele se mostra com muito ou pouco poder.

Problemas de tamanho amostral são dependentes do contexto.

Referência

Lenth, V. Russel, **Some Practical Guidelines for Effective Sample-Size Determination**, *Department of Statistics, Iowa*, 2001.