



DA VERDADE AUTO-PROCLAMADA À VEROSIMILHANÇA RECONHECIDA : UM PONTO CENTRAL NA FORMAÇÃO EM ESTATÍSTICA

Jean-Claude Régnier

Université Lumière Lyon II Bernard F69007 Lyon França
EA648 "Savoirs, Diversité et Professionalisation"

UNICAMP.0517 : Grupo de pesquisa "Psicologia da Educação matemática e educação estatística"
Faculdade de educação, Universidade Estadual de Campinas – SP Brasil

Jean-Claude.regnier@univ-lyon2.fr

Resumo

O presente artigo propõe-se a refletir sobre a *formação do espírito estatístico* baseado no *raciocínio estatístico* através das questões de reconhecimento e de controle dos riscos aos quais o pesquisador expõe-se quando toma uma decisão em situação incerta ou ainda do lugar e do papel que têm as noções de *representatividade* de uma amostra e de *significatividade* de um resultado no raciocínio mesmo dessa tomada de decisão. Às vezes pode-se reduzir a decisão ao escolher entre duas hipóteses alternativas, mas além dessa redução simplificativa que mutila bastante a complexidade da realidade estudada, o problema é que essa decisão baseia-se em informações parciais. Enfim será a hipótese mais *plausível* conforme os resultados obtidos através de uma amostra que o pesquisador escolherá e não necessariamente a *verdadeira* hipótese. A estatística constitui-se como uma fonte rica em ferramentas tanto conceituais quanto metodológicas para modelizar algumas situações de pesquisa em ciências humanas, particularmente em psicologia. De uma certa maneira nesse raciocínio explícito cabe também o raciocínio comum empregado espontaneamente quanto a tomada de decisão. Não obstante observe-se que além das ferramentas metodológicas o pesquisador permanece o ser humano que toma em última instância a decisão. Obviamente algumas problemáticas no campo da didática da estatística são colocadas para compreender-se as condições da aprendizagem do *raciocinar estatisticamente* e do desenvolvimento do *espírito estatístico* como G. Bachelard (1938) fala da formação do espírito científico.

Palavras Chaves : raciocínio estatístico, espírito estatístico, didática da estatística, *representatividade*, *significatividade*.

1. Introdução

Como reconhecer os riscos aos quais o pesquisador expõe-se a tomar a decisão de escolher uma hipótese entre duas conjecturas alternativas ? qual é o papel da noção de *representatividade* da amostra a partir da qual a noção de *significatividade*¹ do resultado obtido dessa amostra fundamentará a tomada de decisão ? Obviamente a resposta técnica já é muito conhecida pelos estatísticos. Assim R. A. Fisher (1935) ou J. Neyman e E. S. Pearson numa serie de artigos célebres publicados nos anos 1933 a 1938 contribuíram para a elaboração da teoria estatística dos testes. A nossa perspectiva é de contribuir ao desenvolvimento da problemática da formação em estatística dos pesquisadores em didática bem como da dos professores. Essas questões requerem a explicitação dos procedimentos auxiliares à tomada de decisão. Nesse raciocínio em lugar da idéia de *verdade* aparece a idéia de *plausibilidade*, de *verosimilhança*, pois face a duas hipóteses alternativas, o pesquisador escolherá aquela que considerará como a hipótese exata em vista das informações parciais que ele detém. Nós acreditamos que esse raciocínio *decisional* abranja varias formas de pensamento inclusive o pensamento comum implícito. Não obstante, o pesquisador pode amplificar o controle utilizando as ferramentas conceituais e metodológicas desenvolvidas no domínio da estatística na construção mesma do seu modelo, antecipando o tratamento dos dados recolhidos no tempo da sua investigação.

2. Partindo-se de um exemplo : avaliação de uma competência no contexto escolar.

Imaginemos um professor querendo avaliar o nível de competência atingido por cada aluno da turma ao fim do curso. Reduzindo-se ao máximo o modelo de avaliação nessa situação, esse professor é confrontado por cada aluno às duas hipóteses opostas :

- a primeira é de julgar o nível de competência como **insuficiente**
- a segunda, julgá-lo como **suficiente**

Para auxiliar a sua decisão simplificando a tarefa, o professor pode recorrer a um questionário constituído de uma ou mais questões independentes ou não, abertas ou fechadas. A dificuldade permanece em elaborar questões pertinentes para requerer a competência cujo desenvolvimento é esperado pelo professor.

Vale salientar que do seu lado o aluno em face de cada questão pode comportar-se conforme diversas atitudes :

- responder ou não responder
- se responder, fornecer uma resposta correta ou uma resposta errada

Considerando-se esse comportamento, o professor pode julgar cada questão conforme três modalidades seguintes :

¹ Nós preferimos esse neologismo *significatividade* à palavra mesma *significância* para designar a questão dessa significância.

<i>modalidades</i>	resposta correta	resposta errada	ausência de resposta ou resposta incompleta não-interpretável
<i>código</i>	S (ucesso)	F (racasso)	NR = Não-Resposta

tabela 1 : avaliação das três modalidades

Voltando-se ao modelo da avaliação, a prática confirma que não é possível propor um questionário comportando um número muito elevado de questões. Portanto as respostas obtidas só podem dar **informações parciais** quanto à competência que não pode ser apreendida por um número tão restrito de questões. Para possibilitar a obtenção destas informações reconhecidas como parciais, o questionário com questões a *múltipla escolha* — QME — constitui-se como uma boa ferramenta, se as respostas propostas para a escolha forem construídas com muita prudência quanto à qualidade e à quantidade. Portanto, esse QME é uma serie de questões comportando cada uma um certo número de possibilidades. Para formalizar, nós diremos que este QME contem **q** questões cada uma com o mesmo número **m** respostas entre as quais o aluno deve escolher só uma ou não responder.

Depois de passar o questionário, o professor terá para cada aluno o numero **S** de sucessos, o número **F** de fracassos e o número **NR** de não-respostas. Observando-se que a soma $S + F + NR = q$. Uma perspectiva interessante seria de tomar a decisão a partir da tripla (S, F, NR) mas isso implicaria num tratamento tridimensional (Régnier 1991a 1991b) um pouco mais complexo que nós não exporemos aqui. Neste caso a **variável de decisão D** usual é a variável que a cada questionário faz corresponder o número S das respostas corretas. O espaço dos resultados possíveis é o conjunto dos números inteiros de 0 a q (número das questões). A partir dessa variável D o professor tomará decisão quanto ao nível de competência atingido por cada aluno. Naturalmente a sua posição extrema poderia ser escolher a hipótese de competência só quando tiver o valor máximo q. Mas essa exigência extrema parece excessiva por razões tanto teóricas quanto práticas. Na prática escolar o professor seria pressionado pelos alunos ou os demais parceiros. Na teoria conforme nós veremos abaixo, isso consistirá em reduzir um risco de um certo tipo de erro ao mesmo tempo que aumentar um outro risco.

Examinemos um modelo ainda mais simples para esclarecer a noção de risco controlado, subjacente a essa tomada de decisão. Em primeiro lugar consideremos que as questões são independentes duas a duas, quer dizer que nenhuma resposta influenciará as demais. Em segundo lugar suponhamos que a probabilidade² de responder corretamente é a mesma para todas as questões, nomeando o seu valor **p**. Com essas condições a variável de decisão D torna-se variável de tipo *binomial* de parâmetros q (número de questões) e p (probabilidade de dar uma resposta correta a uma questão). Esse modelo possibilita o calculo³ da probabilidade de obter um numero S dado entre 0 e q de sucessos ao questionário. Com essas

² aqui nós solicitamos os conhecimentos do leitor sobre a noção de probabilidade. Em qualquer caso, já o sentido comum pode dar uma compreensão dessa noção como uma maneira de medir a possibilidade de ocorrência de um evento.

³ Têm-se formulas matemáticas :

$$\text{Prob} \{ D = k \} = \frac{q!}{k!(q-k)!} p^k (q-k)^{q-k} \text{ probabilidade de obter apenas } k \text{ sucessos sobre } q$$

$$\text{questões. Prob} \{ D \geq k \} = \sum_{j=k}^q \text{Prob} \{ D = j \} \text{ probabilidade de obter } k \text{ ou mais sucessos sobre } q$$

$$\text{questões. Prob} \{ D < k \} = 1 - \text{Prob} \{ D \geq k \} \text{ probabilidade de obter menos de } k \text{ sucessos sobre } q \text{ questões (}$$

ferramentas nós podemos medir a probabilidade dos $q+1$ eventos determinados pelos $q+1$ resultados possíveis quanto ao número de sucessos.

Então o raciocínio será baseado mais na medida da probabilidade de realizar certos eventos definidos pelos resultados da variável de decisão do que nos valores mesmos desses resultados como nós o veremos agora.

3. Conjecturando-se duas propriedades alternativas.

Lembrando-se que na perspectiva do professor-avaliador a situação precedente suscita duas possibilidades de decisão em face de um resultado “número S de sucessos” caracterizadas pela aceitação ou pela recusa de considerar o indivíduo competente (ou incompetente) no domínio em questão. No nosso modelo essas possibilidades vão definir o conteúdo de dois enunciados hipotéticos cujo uso estatístico consagrou os nomes de hipótese nula H_0 e hipótese alternativa H_1 . Em si mesmo não existe uma regra absoluta para escolher qual de ambos os enunciados corresponde a uma ou outra das duas hipóteses estatísticas. Antes de tudo os enunciados hipotéticos devem ser exprimidos numa linguagem autorizando o procedimento dos testes estatísticos para tomar decisão, propriedade que não é comum a todos enunciados hipotéticos. Nesse caso nós tratamos com enunciados desse tipo.

Nós definimos um **teste estatístico** como um procedimento lógico-matemático sobre o qual apóia-se uma tomada de decisão para escolher entre duas hipóteses alternativas baseada nas informações parciais oriundas de uma amostra. Assim construir um teste estatístico de hipóteses consiste em definir explicitamente um conjunto de regras comunicáveis que fundamentam uma tomada de decisão. Nós retornaremos a construção desse procedimento posteriormente.

Esboçemos numa tabela a estrutura desse raciocínio :

Decisão do professor-avaliador	Status desconhecido de H_0 e H_1 na <i>realidade</i>	
	H_0 verdadeira (H_1 falsa)	H_1 verdadeira (H_0 falsa)
Conservar H_0 (rejeitar H_1)	Decisão CORRETA conservar H_0 quando H_0 é verdadeira conservar H_0 quando H_1 é falsa rejeitar H_1 quando H_0 é verdadeira rejeitar H_1 quando H_1 é falsa	Decisão INCORRETA (Erro de tipo 2) conservar H_0 quando H_0 é falsa conservar H_0 quando H_1 é verdadeira rejeitar H_1 quando H_0 é falsa rejeitar H_1 quando H_1 é verdadeira
Rejeitar H_0 (conservar H_1)	Decisão INCORRETA (Erro de tipo1) rejeitar H_0 quando H_0 é verdadeira rejeitar H_0 quando H_1 é falsa conservar H_1 quando H_0 é verdadeira conservar H_1 quando H_1 é falsa	Decisão CORRETA rejeitar H_0 quando H_0 é falsa rejeitar H_0 quando H_1 é verdadeira conservar H_1 quando H_0 é falsa conservar H_1 quando H_1 é verdadeira

tabela 2: Situações possíveis em termos de tipos de decisão tomada em função das hipóteses básicas consideradas, e tipos de erro associados.

Voltando-se a escolha dos enunciados respeitavelmente atribuídos a cada uma de ambas as hipóteses H_0 e H_1 . A dificuldade é de modelizar de tal maneira que todos os cálculos de probabilidades sejam possíveis,

o valor k excluído)

isso quer dizer que nós temos um modelo da realidade estudada bem especificado. Não obstante, essa especificação requer varias informações que motivam exatamente a pesquisa, assim encontramos-nos num círculo vicioso.

Na situação habitual dá-se a preferência ao controle sob condições ligadas à primeira hipótese chamada usualmente hipótese nula H_0 . Assim no nosso exemplo a hipótese nula enunciará que o indivíduo é *incompetente no domínio estudado*, pois a incompetência pode ser modelizada pelo comportamento de um “respondente aleatório” cuja probabilidade de fornecer uma resposta correta é o valor $p = 1/m$ (lembrando-se que m é o número de escolhas de cada questão). Entretanto sob a hipótese alternativa H_1 de *competência no domínio estudado* não é fácil de medir a probabilidade do evento “fornecer uma boa resposta a uma questão” ou ainda “escolher o bom rótulo”. Nós consideramos que a probabilidade de responder corretamente a uma questão é maior sob H_1 que sob H_0 , mas não temos critérios objetivos e absolutos para quantificar essa probabilidade. Para conhecer esse valor, temos que utilizar as ferramentas estatísticas desenvolvidas na área das teorias da estimação. Sem desenvolver mais, nesse caso um bom estimador é o que dá a freqüência de respostas exatas a uma questão a partir de uma amostra. Agora o problema é de controlar o risco que nós corremos quando escolhermos entre H_0 e H_1 .

4. Controlando-se o risco quando da escolha entre dois enunciados alternativos.

Especificando-se o modelo, nós podemos medir a possibilidade de errar como expõe a tabela seguinte:

Decisão do pesquisador	Status desconhecido de H_0 e H_1 na realidade	
	H_0 verdadeira (H_1 falsa)	H_1 verdadeira (H_0 falsa)
Conservar H_0 (rejeitar H_1)	Decisão CORRETA $1 - \alpha =$ Prob{conservar H_0 sob a condição H_0 verdadeira}	Decisão INCORRETA (Erro de tipo 2) $\beta =$ Prob{conservar H_0 sob a condição H_0 falsa} = Prob{conservar H_0 sob a condição H_1 verdadeira}
Rejeitar H_0 (conservar H_1)	Decisão INCORRETA (Erro de tipo1) $\alpha =$ Prob{rejeitar H_0 sob a condição H_0 verdadeira}	Decisão CORRETA $1 - \beta =$ Prob{rejeitar H_0 sob a condição H_0 falsa} = Prob{rejeitar H_0 sob a condição H_1 verdadeira } Potência do teste

tabela 3: Situações possíveis em termos de tipos de decisão tomada em função das hipóteses básicas e medidas das **Probabilidades** dos eventos associados.

Um bom teste de hipóteses será um teste planejado de modo que os erros de decisão sejam reduzidos ao mínimo. Isso também não é uma tarefa simples, pois para uma amostra de tamanho dado diminuir um risco de um tipo nos conduz a aumentar o risco do outro tipo. A redução simultânea de ambos os riscos é obtida pelo aumento do tamanho da amostra o que tem um custo, as vezes insuportável ou mesmo impossível.

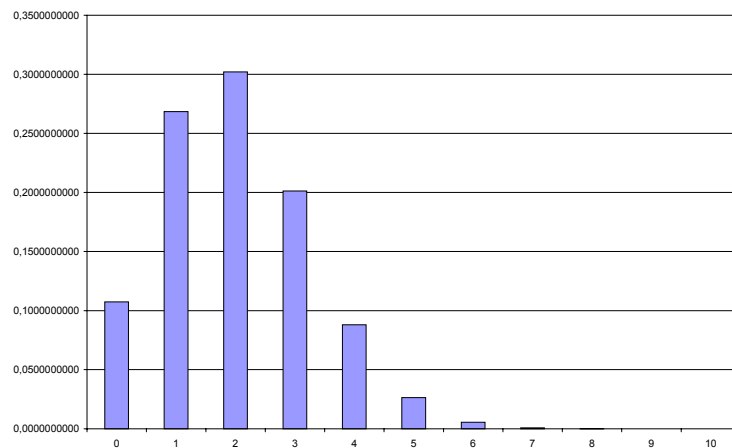
Voltando-se a nossa situação de avaliação, imaginemos que o questionário possua 10 questões, bem como

o espaço dos resultados da variável D de decisão seja composto dos valores inteiros de 0 até 10. Habitualmente o critério de decisão baseia-se no valor central 5 considerado como **valor crítico** separando o espaço de resultados possíveis em dois conjuntos respectivamente $A = \{0, 1, 2, 3, 4\}$ e $R = \{5, 6, 7, 8, 9, 10\}$. Então, quando um indivíduo obtiver um escore referido a A o professor-avaliador prefere optar para o julgamento de *incompetência*, o que quer dizer : conservar a hipótese H_0 como a hipótese válida, mas quando ele obtiver um escore referido a R, o professor-avaliador prefere optar para o julgamento de *competência*, o que quer dizer : rejeitar a hipótese H_0 como a hipótese válida, portanto conservar a hipótese H_1 .

Tentemos agora medir os riscos no nosso modelo no qual a variável D é uma variável binomial. Imaginando-se que cada questão seja composta por 5 modalidades de resposta onde uma única é a resposta exata esperada, portanto $p = 1/5 = 0,20$ é a probabilidade⁴ de acertar uma questão.

Distribuição da variável binomial $n=10$ e $p=0,2$

0	0,1073741824
1	0,2684354560
2	0,3019898880
3	0,2013265920
4	0,0880803840
5	0,0264241152
6	0,0055050240
7	0,0007864320
8	0,0000737280
9	0,0000040960
10	0,0000001024



O cálculo das probabilidades nos conduz aos valores seguintes

$$\alpha = \text{Prob}\{D \geq 5 \text{ sob } H_0 \text{ verdadeira}\} = \text{Prob}\{\text{"escore obtido pertence a R"}\} \approx 0,0063$$

No caso onde H_0 for errada, poderíamos considerar que a probabilidade⁵ de acertar uma questão vale $p = 0,80$ e calcular sob a hipótese H_1 :

$$1 - \beta = \text{Prob}\{D \geq 5 \text{ sob } H_1 \text{ verd.}\} = \text{Prob}\{\text{"escore obtido pertence a R"}\} = 0,9672$$

ou ainda

$$\beta = \text{Prob}\{D < 5 \text{ sob } H_1 \text{ verd.}\} = \text{Prob}\{\text{"escore obtido pertence a A"}\} = 0,0327$$

Então nós podemos observar que um aluno *incompetente* respondendo aleatoriamente ao questionário de *múltipla escolha* terá somente 6 chances sobre 1000 de ultrapassar o ponto crítico de 5, contudo, do ponto

⁴ Em outros termos um indivíduo que responde de maneira aleatória tem 20% de chance de fornecer uma resposta correta a uma questão.

⁵ em outros termos do ponto de vista do professor-avaliador, isso quer dizer que o evento "um indivíduo fornece uma resposta correta a uma questão" é um evento que tem 80% de chance de acontecer. Aqui nós encontramos uma dificuldade na modelização, na medida em que modelizar o comportamento do indivíduo *incompetente* por o comportamento de "respondente estocástico" parece bem adequado, enquanto que modelizar o comportamento *competente* dessa maneira parece não traduzir o fato que ele sabe, portanto, ele não vai responder aleatoriamente. Não obstante, do ponto de vista do observador exterior, o comportamento de um indivíduo possui em parte um caráter aleatório.

de vista do professor-avaliador um aluno *competente* terá 967 chances sobre 1000. Daí o professor-avaliador que deve tomar a decisão pode utilizar essa informação probabilística para fundamentar a sua escolha. A probabilidade do evento que um aluno obtiver qualquer escore superior ou igual a 5 vale somente $\alpha = 0,63\%$ quando este aluno é *incompetente no domínio estudado*, porém, ela sobe até $1-\beta = 96,7\%$ quando ele é *competente no domínio estudado*. O aluno que obtiver a nota 6, por exemplo, será considerado como *competente* rejeitando-se a hipótese H_0 . Esta decisão pode ser errada, pois o procedimento possibilita essa eventualidade de obter 6 até no caso da *incompetência*, e por esta razão de ser sujeito a correr o risco de erro do primeiro tipo, nós diremos que a decisão de rejeitar a hipótese H_0 de *incompetência do aluno no domínio considerado* é tomada com um **nível de significância** menor que $\alpha = 0,63\%$, que fornece uma **medida** de risco de erro de primeiro tipo.

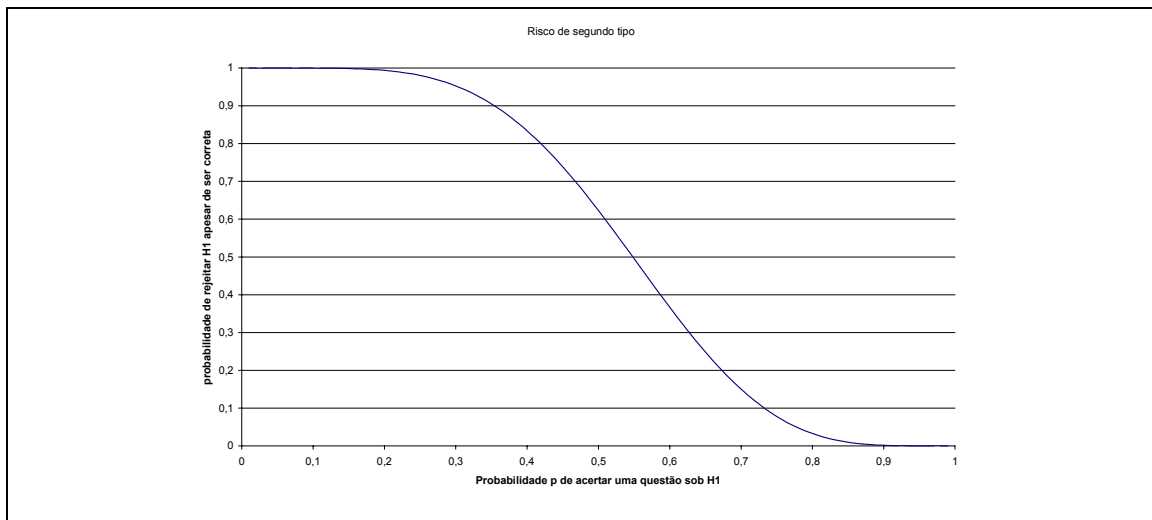
Agora que podemos dizer quanto a situação do aluno que obtiver a nota 3 ?

A probabilidade do evento que um aluno obtiver qualquer escore inferior a 5 vale somente $1-\alpha = 99,37\%$ quando este aluno é *incompetente no domínio estudado*, porém, ela desce até $\beta = 0,328\%$ quando ele é *competente no domínio estudado*. O aluno que obtive a nota 3, por exemplo, será considerado como *incompetente* conservando a hipótese H_0 . Todavia uma tal decisão pode ainda ser errada, pois o procedimento possibilita a eventualidade de obter 3 mesmo por um aluno *competente no domínio estudado*, e por esta razão nós estamos sujeitos a correr o risco de erro do segundo tipo tomando a decisão de conservar a hipótese H_0 de *incompetência do aluno no domínio considerado* apesar de ser *competente*. Nesse caso particular, o nível do risco de erro de segundo tipo vale menos de $\beta = 0,328\%$. Notemos que o valor $1-\beta$ permite de medir a **potência**⁶ do teste de hipóteses considerada como uma medida da capacidade de um teste estatístico de tomar a decisão correta de conservar a hipótese H_0 quando ela for justa.

Partindo-se de todas essas considerações nós constatamos que o *raciocínio estatístico* não possibilita a estabelecer com certeza a verdade nem a falsidade de um enunciado hipotético. Cada escolha está sujeita a um risco de erro entre ambos os tipos descritos acima. Por conta dessa observação nós privilegiamos o estatuto de **verossimilhança** ou de **plausibilidade** de um enunciado em lugar do de **verdade** ou para o enunciado alternativo, do de **falsidade**. A situação mais favorável aparece quando for possível de acompanhar de um nível de risco de erro, a decisão tomada quanto a uma ou a outra de ambas as hipóteses opostas.

A maior parte do tempo, infelizmente como nós dissemos acima, a modelização não permite de especificar precisamente o comportamento sob a hipótese H_1 de *competência* no sentido que não podemos precisar o valor da probabilidade de acertar uma questão além de conjecturar que este valor for superior ao valor fixado pelo modelo sob a hipótese H_0 de *incompetência*. No nosso exemplo, sob H_0 fixamos o valor $p = 0,20$, o gráfico seguinte dá os valores de $\beta = \text{Prob}\{D < 5 \text{ sob } H_1\}$:

⁶ Para um valor dado do nível de significância, dentro de dois testes o melhor é o teste que tem a maior potência.



Para o valor $p = 0,50$ sob a hipótese H_1 de *competência*, o valor β ainda fica alto pois $\beta \approx 62,3\%$. Acima do valor $p=0,75$ o valor é reduzido à $\beta \approx 7,8\%$. Acima de $p = 0,80$, seu valor cai a $3,2\%$.

Para o professor-avaliador essa característica do raciocínio estatístico que fundamenta uma tomada de decisão nas informações parciais, traduz-se por uma modificação do seu comportamento de avaliador no sentido de considerar uma parte de incerteza na sua decisão. A sua apreciação será orientada mais pela noção de plausibilidade, de verossimilhança que pela noção de verdade. Dizendo-se que quando um aluno obtiver a nota 6, pareceria que a hipótese de *competência* fosse mais *verossimilhante* que a de *incompetência*. Contudo quando só ele atingisse a nota 3, pareceria melhor conservar a hipótese de *incompetência* como a mais *verossimilhante*.

Para um pesquisador o **raciocínio estatístico** apesar de ser uma ferramenta extraordinária para distinguir o conhecimento científico do conhecimento de *sensu comum*, deve ser empregado com uma consciência das propriedades e dos limites. Nessa perspectiva nós vamos examinar as conseqüências das características do raciocínio estatístico sobre um procedimento bem conhecido : o procedimento experimental

5. Aceitando-se os riscos no procedimento experimental

Lembre-se que no procedimento experimental encontram-se duas fases importantes referidas nos propostos acima. Na fase inicial tem-se que construir dois grupos respectivamente chamados grupo C “controle” e grupo E “experimental”. Antes de experimentar, ambos os grupos devem ser **idênticos** do ponto de vista das variáveis relacionadas à pesquisa. Aqui aparece a necessidade de **testar essa identidade**. Para facilitar o esclarecimento, imaginemos que a problemática da pesquisa seja centrada sobre a aprendizagem de um conjunto de conceitos em matemática e o efeito da prática de auto-avaliação. Enfim imaginemos que o critério de comparação seja reduzido à média de uma certa variável utilizada pelo pesquisador, por exemplo a variável D evocada anteriormente. Por motivos óbvios nós consideramos que os dados resultam de duas amostras aleatórias independentes constituídas por ambos os grupos E e C oriundos da população dos alunos. Aqui vale salientar que nós solicitamos a noção de **representividade** das informações que traduz o modo operatório de obtenção de ambas as amostras : **sorteando**. Após a

amostragem, deve-se controlar a identidade de ambas as amostras quanto à média, isso requer o emprego de um teste estatístico de hipóteses. Nesse caso utilizar-se-ia o teste de igualdade de duas médias que conduz-nos à formular da maneira seguinte ambas as hipóteses como H_0 e H_1

H_0 “a média da variável D sobre a população da qual é oriunda o grupo amostral C é **igual** à média da variável D sobre a população da qual é oriunda o grupo amostral E”

H_1 “essas duas médias **são diferentes**”

Não detalharemos todas as operações, nem todas as condições requeridas, principalmente aquelas ligadas às variâncias, para aplicar esse teste. Todavia, nós podemos observar que o alvo do pesquisador é de obter dois grupos idênticos, portanto, a um nível de significância dado *a priori*, ele quer conservar a hipótese nula H_0 para continuar as operações experimentais planejadas. **Isso quer dizer que desde o início esse procedimento não pode evitar de estar sujeito a um risco de erro de tipo II** sem falar das dificuldades para inibir os efeitos das múltiplas variáveis da situação.

Na fase final deve-se decidir entre as hipóteses H_0 e H_1 qual é a mais *verossimilhante*. Mas essa decisão ainda está sujeita a correr o risco de erro. Isso quer dizer que ao final a conclusão relativa ao efeito duma prática pedagógica, voltando ao nosso exemplo, poderá conduzir a :

formulação 1

- “ Há um efeito sobre a aprendizagem” quando a hipótese H_0 for rejeitada com um nível⁷ α de risco de erro de tipo I

formulação 2

- “ Não há um efeito sobre a aprendizagem” quando ao nível α de significância a hipótese H_0 não puder ser rejeitada. Portanto, a hipótese H_0 será conservada ou que a hipótese H_1 será rejeitada com um risco de erro de tipo II de nível β .

Isso quer dizer que até o final, esse procedimento não pode evitar de sujeitar-se a correr um risco de erro de tipo I ou de tipo II

Para concluir, nós poderíamos afirmar que o pesquisador deve se conscientizar dos limites da validade dos conhecimentos científicos produzidos quando da utilização do procedimento experimental. Particularmente, ele deve estar atento aos riscos que ele se expõe desde o início do procedimento com o risco de erro de tipo II até o fim principalmente com o risco de erro de tipo I quando seu objetivo é de confirmar o efeito estudado.

Além disso, nós gostaríamos de chamar a atenção sobre um tipo de erro pouco evocado e referente à escolha das várias ferramentas tanto para construir o modelo teórico no quadro do qual a pesquisa é conduzida, quanto para realizar os diversos tratamentos. Esse risco que nós nomeamos aqui como **risco de erro de terceiro tipo**, inclui vários erros tais como : erro de escolha de teste estatístico, erro referente à natureza duma variável, erro de transcrição de dados, etc.. O seu controle passa principalmente pelos intercâmbios entre pesquisadores, mas nunca através de α e de β .

⁷ muitas vezes o valor de α é fixado pelo pesquisador a 0,05 ou 0,01.

6. Concordando-se com as noções de *representatividade* e de *significatividade*.

A perspectiva oferecida pela abordagem estatística apresentada não deve ser considerada como a de um *quantitativismo* redutor. Ela tem como mero objetivo de reduzir a subjetividade produzindo regras de decisão o mais possível explícitas. Essa explicitação tem a vantagem de poder dar conta dessas regras e de discuti-las num espaço de confrontações científicas. A quantidade enorme de ferramentas estáticas já construídas permite ao pesquisador em ciências humanas e sociais, em didática da matemática de escolher as que são adequadas a seu modelo teórico e que podem enriquecê-lo. Nesse sentido ainda, a vontade de precisar os termos utilizados constitui um objetivo particular, tais como *significativo* e *representativo* cujo emprego usual ofusca o sentido.

A citação abaixo ilustra a utilização superficial desses termos (Meirieu 1993 p.290) “Além disso o desenvolvimento de instituições escolares sobre o modelo ocidental nos países do terceiro mundo produziu, **evidentemente**, efeitos **frequentemente negativos**, descreditaando os saberes tradicionais, as vezes muito úteis, sem conseguir transmitir saberes novos de forma **significativa**...”.

Esta idéia do caráter significativo das informações constitui também o objeto de uma tentativa de explicitação nessa perspectiva. Em que sentido ele é utilizado ? No de considerar o que pode significar um resultado empiricamente obtido em comparação a um outro que se obteria teoricamente sob certas condições (em particular sob H_0 o sob H_1). A idéia seria assim de fazer de tal maneira que o risco corrido quando da tomada de decisão dependa somente da variação da amostra e não de erros de observação, de medida ou transcrição dos resultados. Seria interessante de refletir neste momento aos efeitos das aproximações numéricas decorrentes de arredondamento. Na perspectiva desenvolvida neste artigo, a idéia de significatividade está relacionada a de probabilidade de obtenção de um resultado pelo mero acaso. Quanto mais o valor da probabilidade deste evento é grande, menos o resultado é considerado como significativo, isto é que traga uma significação ao fenômeno estudado e a hipótese formulada. Assim, no nosso exemplo a hipótese H_0 de *incompetência* é submetida a prova dos fatos.

Com relação à noção de representatividade da amostra, utilizada para obter uma realização das variáveis estudadas e para calcular o valor tomado pela variável de decisão, nós forneceremos alguns exemplos. Nós encontramos frequentemente em dissertações acadêmicas ou monografias reservas com relação a representatividade da amostra utilizada, sem por isso impedir o uso das informações assim coletadas, na argumentação, inclusive se apoiando sobre elas no que diz respeito a significatividade.

Parece aqui se tratar muito mais de precaução formal que de uma observação oriunda de uma análise do método de obtenção da amostra. Isto é confirmado pela maneira como as informações da amostra são retomadas na argumentação. Abordadas dessa forma, tanto a representatividade quanto a não representatividade, assim como os vieses por elas induzidos, são estabelecidos por uma argumentação baseada meramente em opiniões, se consideramos apenas ao relatório escrito. Entretanto os erros possíveis relacionados às observações ou às medidas, nunca são evocados.

Durante a orientação metodológica de pesquisa, a questão da representatividade é colocada sistematicamente, mas inicialmente e às vezes exclusivamente a propósito do tamanho da amostra.

Em que sentido pode-se empregar o qualificativo *representativo* para uma amostra oriunda de uma população ?

Em primeiro lugar, no sentido de um fragmento desta população permitindo de coletar informações as mais próximas possíveis, das quais obteríamos mediante **recenseamento**⁸, do tipo “modelo reduzido” pelo meio de uma transformação homotética.

Em segundo lugar, pode-se encarar essa noção através da probabilidade de um sujeito⁹ pertencer à amostra. Geralmente uma amostra é dita representativa quando ela é obtida por meio do **sorteio** sob a condição de equiprobabilidade. Neste caso, chama-se o método de constituição aleatória de amostra, de método de *amostragem casual*, designando *amostragem não-casual* ou *amostragem acidental* quando do uso dos demais métodos.

A idéia meramente básica sobre as propriedades da amostragem é de fazer de tal maneira a obter-se uma amostra autorizando-se inferências da *parte* (amostra) para a *totalidade* (população). Por razões óbvias, ao lado do procedimento de **sondagem** utilizado, o **tamanho n da amostra** constitui-se mais numa medida da quantidade de informações fornecidas que do caráter *representativo* da amostra. O tipo usual de amostra é o da *amostra casual (aleatória) simples* constituída por extração estocástica¹⁰ de n unidades com ou sem repetição. Desse ponto de vista, uma amostra casual de tamanho $n = 1$ apesar de parecer não-representativa, é uma amostra representativa que pode trazer uma informação interessante. Por exemplo, imaginemos uma caixa de 10000 parafusos idênticos, basta medir uma única unidade para conhecer todas as características dimensionais de todos os elementos da população. Por enquanto, uma amostra aleatória de tamanho $n = 1000$ extraída de uma população (uma caixa de 10000 parafusos de vários comprimentos) com repetição que não suscita nenhuma dúvida quanto a sua representatividade, poderia conter pouca informação se o sorteio fizesse sair 1000 vezes a mesma unidade (evento que não é impossível, somente a sua probabilidade é infinitésima igual a $1/10^{12}$).

Constata-se que o conhecimento do significado particular dos termos em uso parece muito importante para entender textos enunciando resultados baseados em abordagens estatísticas. Assim o trecho seguinte de artigo (Laveault & Fournier 1990 p.54-57) a partir de uma pesquisa sobre *auto-avaliação e metacognição* , ilustra esta observação :

“A amostra é constituída por 30 alunos canadenses (Quebec) do 11º ano (classe), 13 meninas e 17 meninos (idade média : 16 anos). ... Os sujeitos do grupo dito *forte* (grupo do terceiro quartil) compartilham mais a mesma opinião relacionada as questões que serão escolhidas para as provas, que os sujeitos do grupo dito *fraco* (grupo do primeiro quartil). Pois o W de concordância de Kendall para os oito estudantes-avaliadores do grupo *forte* é de 0,2362 (significativo a 0,01), enquanto, esse valor caiu à 0,1208 ($p = 0,348$) para o grupo *fraco*. Isso indica que os sujeitos que erraram mais ao exame, são os que discordam mais quanto às questões que serão selecionadas pelo professor. Os sujeitos do grupo *forte* concordam também mais quanto à preparação. Pois este grupo obteve um $W = 0,1946$ (significativo a

⁸ o que supõe um arrolamento completo de todas as unidades da população, procedimento custoso e de vez em quando impossível.

⁹ unidade estatística como elemento da população estudada.

¹⁰ mediante uma tábua de números aleatórios ou duma função aleatória integrada a um programa tal como Excel5 a função *ÁLEA()* .

0,05), enquanto que o grupo fraco obteve um $W = 0,1381$ ($p = 0,190$) Pode-se interpretar este resultado como significando que o grupo *forte* tem uma maior tendência a concentrar os esforços do estudo sobre os mesmos itens que o grupo *fraco*.”

Obviamente a compreensão deste texto solicita uma formação mínima em estatística. Dar um sentido mais específico e mais circunscrito a noções frequentemente utilizadas na língua comum ou correspondendo às práticas concretas, como as de risco, de significatividade ou de representatividade que intervêm nas tomadas de decisão, substituindo-os nos domínios da estatística matemática e da estatística aplicada, nos parecem ser um dos objetivos prioritários das ações de formação em estatística.

7. Conclusão

Ao se abordar alguns aspectos técnicos do raciocínio estatístico, surgem claramente questões relativas à formação em estatística. Fazemos referência às questões de ensino-aprendizagem que precisam de uma problematização na área da didática da estatística. Nesse sentido nós tentamos abordar esse assunto nas nossas pesquisas (Régnier 2002a) (Régnier & Falcão 2002b) (Régnier 2003). Um trabalho importante já foi realizado também por C. Batanero. Em particular a respeito desse raciocínio ela escreveu (Batanero 2001 p.106) “ La investigación sobre la comprensión de los métodos de inferencia muestra la existencia de concepciones erróneas ampliamente extendidas, tanto entre los estudiantes universitarios, como entre los científicos que usan la inferencia estadística en su trabajo diario. En particular los estudiantes tienen dificultades en los aspectos siguientes:

- La determinación de la hipótesis nula H_0 y la hipótesis alternativa H_1 ;
- La distinción entre los errores Tipo I y Tipo II;
- La comprensión del propósito y uso de las curvas características operativas o curvas de potencia;
- La comprensión de la terminología empleada al establecer la decisión.”

Entre os erros comuns o mais freqüente consiste na interpretação errada a respeito ao conceito de nível de significância α e do p-valor. Este erro consiste na troca de ambos os termos da probabilidade condicional, isso quer dizer, em interpretar o nível de significância como a probabilidade de que a hipótese nula seja certa se tenhamos tomado a decisão de recusá-la.

A etapa seguinte seria de integrar essas observações de pesquisa em didática da estatística nas construções das situações problemas para possibilitar a aprendizagem do raciocínio estatístico mobilizado pelas ferramentas dos testes de hipóteses.

8. Bibliografia

- Bachelard G. (1938) La formation de l'esprit scientifique. Contribution à une psychanalyse de la connaissance objective. Paris Librairie Philosophique J. Vrin (14^{ème} ed. 1989)
- Batanero C. (2001) *Didáctica de la Estadística* Granada GEEUG.
- Fisher R.A. (1935) *The Design of Experiments*. Edinburgh, Oliver & Boyd
- Laveault, D., Fournier, Ch. (1990) Evaluation par objectifs : une approche métacognitive, *Revue Mesure et Evaluation*, vol 13,1 pp 54-57
- Meirieu, Ph, (1993) Objectifs, obstacles et situations d'apprentissage, in Houssaye,J., *La pédagogie : une encyclopédie pour aujourd'hui*, Paris, ESF-France

- Régnier, J-C, & al (1991a) *Autonomie et travail personnel dans l'enseignement des disciplines scientifiques au lycée : Des outils pour faciliter l'approche d'une pédagogie de l'autonomie*, MEN-CRDP Dijon- França
- Régnier, J-C, (1991b) *Autonomie et travail personnel dans l'enseignement des disciplines scientifiques au lycée: Témoignage d'un travail conduit sur une année scolaire*, MEN-CRDP Dijon- França
- Régnier, J-C, (2002a) A propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique . *Questions éducatives Revue du centre de recherche en éducation*, Saint-Étienne : Université Jean Monnet. pp.157-201
- Régnier, JC, Da Rocha Falcão, J.T.,(2002b) Sobre os métodos quantitativos na pesquisa em ciências humanas: riscos e benefícios para o pesquisador, *Revista brasileira de Estudos Pedagógicos*, Vol. 81 n°198 pp. 229-243
- Régnier, J-C, (2003) Statistical education and e-learning *Proceedings of Satellite Statistics & the Internet, Berlin, Germany* .IASE
URL=<http://www.stat.auckland.ac.nz/~iase/publications.php?show=6>