REVISTA PROYECCIONES N° 9: 1-9
Julio 1985 - I.S.S.N. 0716-0917

ESTATÍSTICA E INFORMAÇÃO*

Dr. CARLOS ALBERTO DE BRAGANÇA PEREIRA**

O sucesso de um grupo de pesquisa depende mais da criatividade de seus membros do que da quantidade de conhecimento que estes já tenham adquirido. Para o pensamento criativo de um cientista¹ é crítico o momento em que este é obrigado a utilizar um método estatístico. Isto porque o método estatístico clássico, que por ser matematicamente elegante dominou por décadas a mente dos estatísticos, abandona por completo a experiência do cientista no momento da análise. Na maioria das vezes, este fica marginalizado das conclusões da análise e somente acredita nelas porque "confia" no seu colega estatístico. A complexidade dos métodos ou a ignorância sobre os mesmos o obriga a esta atitude.

¹ Cientista aqui pode ser qualquer elemento que recorre ao estatístico.

* Conferencia dictada en la XII Jornadas Nacionales de Estadística. Antofagasta 1985.

** Profesor Asistente del Instituto de Matemáticas y Estadística de la Universidad de Sao Paulo-Brasil.

Ao contrário do clássico, o estatístico moderno necessita da presença constante do cientista. Um dos métodos utilizados por este estatístico é o método Bayesiano, devido a Thomas Bayes, que é o objeto deste artigo. Neste método, a opinião ou a experiência do cientista é representada por um modelo probabilístico. Na verdade, o estatístico aqui pretende funcionar apenas como especialista em medir ou representar, por um modelo, opiniões frente a natureza. Assim, as conclusões de uma análise de resultados dependerá fortemente da opinião do cientista e não somente da aplicação de um método.

Basicamente, o método Bayesiano consiste em representar a opinião do cientista por uma distribuição de probabilidades (distribuição a priori) e após novas observações de evidências, esta distribuição se modifica acompanhando as trocas de opinião do cientista. A distribuição calibrada pelas novas evidências é denominada distribuição a posteriori. Assim, dois tipos de informação são consideradas no método Bayesiano: a informação do cientista e a contida nos dados ou resultados experimentais. Estes são resultados obtidos mediante a realização de um experimento² que tem como objetivo diminuir a incerteza. Combinando-se os dois tipos de informação chega-se a opinião a posteriori representada pela distribuição a posteriori a qual carrega a informação adicional que os dados ou resultados experimentais fornecem.

Assim como em outros trabalhos de estatística, o conceito de informação não é bem definido. Informação é o que ela provoca: a troca de opinião. A informação aqui considerada é sobre um estado da natureza, N , que é desconhecido. É claro que a incerteza sobre o verdadeiro valor de N depende da experiência do cientista envolvido e assume diferentes graus para diferentes cientistas. Na concepção Bayesiana da Estatística, os diferentes graus de incerteza são representados por modelos probabilísticos diferentes. Assim, um modelo probabilístico $P(N)$ representa a incerteza de um particular cientista sobre o verdadeiro valor de N .

² Experimento é dado mecanismo que transforma uma quantidade desconhecida em conhecida, o resultado experimental.

Ao representarmos por d os dados que consideramos relacionados a N , denotaremos por $v(N|d)$ a função de verossimilhança. Esta é uma função de N que para cada valor possível de N nos dá o valor da probabilidade de se obter d . Note que $v(N|d)$ não é uma função de probabilidade para N pois, fornecendo a probabilidade de d para cada possível N , não obtém-se necessariamente a unidade ao somar-se $v(N|d)$ para todo valor possível de N . A função de verossimilhança é o único elemento do processo que formaliza o relacionamento existente entre os dados, d , e o estado da natureza, N . Após a obtenção de $P(N)$ e $v(N|d)$, a etapa seguinte do processo é o cálculo da distribuição a posteriori representada por $P(N|d)$ e que deve ser entendida como a função de N que fornece a probabilidade de cada valor de N após observar-se o valor de d . Note que tanto $P(N)$ quanto $P(N|d)$ são funções de probabilidades para N . Ao utilizar-se o teorema de Bayes, tem-se:

$$P(N|d) = \frac{P(N) \cdot v(N|d)}{\sum P(N) \cdot v(N|d)}$$

onde $\sum$ indica soma sobre todos os valores possíveis de N e assim o denominador da expressão é uma constante. Então, ao representarse "proporcional a" por $\propto$, pode-se escrever

$$P(N|d) \propto P(N) \cdot v(N|d) \quad (*)$$

provando formalmente a operação Bayesiana que consiste em calibrar a opinião a priori, $P(N)$, pela informação sobre N contida em d , $v(N|d)$.

A receita Bayesiana descrita por (*) é bem simples e é aplicável a todas as situações práticas. Os exemplos descritos a seguir tentam esclarecer alguns pontos importantes do conceito de informação.

Primeiro Exemplo:

André, Bruno e Carlos são três prisioneiros da rainha Lúlian a qual decide libertar um dos prisioneiros e executar os outros dois na guilhotina. Os prisioneiros foram informados da situação mas, o nome daquele que seria libertado só iria ser divulgado no momento das execuções,

que se dariam no dia seguinte. André que se considera um especialista em teoria de probabilidades, teve o seguinte diálogo com Marcel, o carcereiro que nunca mente:

A - Eu sei que pelo menos um dos meus companheiros, B ou C, necessariamente vai morrer. Assim, você não estaria me dando nenhuma informação sobre a minha situação caso me conte o nome, excluindo o meu quando for o caso, de um dos que vai morrer.

M - Concordo com você e posso lhe dizer que Carlos vai morrer.

Assim que Marcel retirou-se, André descrevendo o diálogo para Bruno disse que com a informação recebida a sua chance de ser liberrado havia aumentado de $\frac{1}{3}$ para $\frac{1}{2}$. Bruno discordando de André disse:

B - Voce já sabia anteriormente que pelo menos um de nós dois, eu ou Carlos, iria necessariamente morrer. Como apenas o conhecimento de um nome, no caso Carlos, pode lhe trazer informação representada pela troca de probabilidades? Vou lhe provar que você está errado.

Bruno então deu a seguinte aula para André, que em nenhum momento se mostrou surpreso.

Os entes envolvidos têm a seguinte representação:

- (i) O estado da natureza, N , pode ser $A = \text{André será libertado}$ ou $\bar{A} = \text{André será executado}$.
- (ii) O resultado experimental é $d = \text{Marcel informa que Carlos será executado}$.

Os elementos do modelo, $P(N)$ e $v(N|d)$, devem ser definidos como:

iii) Ao admitir-se a escolha ao acaso, representando o desconhecimento total sobre quem foi selecionado, definiu-se a distribuição a priori como

$$P(N) = \begin{cases} 1/3 & \text{se } N = A \\ 2/3 & \text{se } N = \bar{A} \end{cases}$$

iv) Mostrando a não informatividade de d , por não variar, a função de verossimilhança é definida como

$$v(N|d) = \begin{cases} 1/2 & \text{se } N = A \\ 1/2 & \text{se } N = \bar{A} \end{cases}$$

isto é, $v(A|d) = v(\bar{A}|d) = 1/2$.

Para entender-se esta última função, lembremos que $v(A|d)$ é a probabilidade de Marcel informar que Carlos morrerá dado que Bruno e Carlos serão os executados. Assim como Marcel pode escolher qualquer dos dois, é razoável definir-se $v(A|d) = 1/2$. Por outro lado $v(\bar{A}|d)$ é a probabilidade de Marcel informar que Carlos morrerá quando André também será executado. Equivalentemente, como Marcel não pode citar o nome de próprio André, tem-se que $v(\bar{A}|d)$ é a probabilidade de André e Carlos morrerem dado que André irá morrer. Esta probabilidade condicional é a razão $(1/3) / (2/3)$ cujo numerador é a probabilidade conjunta de André e Carlos serem os condenados à morte e o denominador é a probabilidade de André ser executado. Assim, conclui-se que $v(N|d) = 1/2$ para $N = A$ e $N = \bar{A}$.

De posse de $P(N)$ e $v(N|d)$, pode-se calcular $P(N|d)$ que para $N = A$ tem-se

$$P(A|d) = \frac{P(A) \cdot v(A|d)}{P(A) \cdot v(A|d) + P(\bar{A}) \cdot v(\bar{A}|d)} =$$

$$= \frac{(1/3) (1/2)}{(1/3) (1/2) + (2/3) (1/2)} = 1/3$$

isto é, $P(A) = P(A|\hat{d})$ indicando que a opinião a priori é a mesma da posteriori o que corresponde a dizer que o resultado experimental $\hat{d}$ não traz nenhuma informação sobre o estado de natureza N .

Ao terminar a aula, Bruno teve o seguinte diálogo com André:

B - Espero ter provado que você estava equivocado no seu raciocínio.

A - Perdoe me Bruno, mas você decididamente não é bom observador. Nestes dias de convívio aqui no cárcere, percebi como Marcel gosta de você e despreza o convencido do Carlos. Assim, eu acredito que, no caso de você e Carlos terem sido os escolhidos para morte, com intuito de lhe poupar neste último dia de vida Marcel me diria, com probabilidade 1, o nome de Carlos. Esta informação que colectei é representada na minha verossimilhança a qual é definida por

$$v(N|\hat{d}) = \begin{cases} 1 & \text{se } N = A \\ 1/2 & \text{se } N = \bar{A} \end{cases}$$

Note que com isto temos

$$P(A|\hat{d}) = \frac{(1/3)1}{(1/3)1 + (2/3)(1/2)} = \frac{1}{2}$$

como eu havia considerado.

B - Caro André você tem toda razão. Você me mostrou evidências que passaram despercebidas aos meus olhos. Para reforçar o seu argumento devo informar-lhe que Marcel além de ser meu primo, é o meu melhor amigo desde a infância.

Deve-se aqui realçar o fato de que dois diferentes cientistas (A e B no exemplo), embora tenham observado o mesmo resultado experimental d (M informar que C irá ser executado), podem diferir em suas conclusões. A experiência e o conhecimento de um cientista devem influenciar decisivamente nas conclusões de uma análise estatística.

Espera-se que o leitor, ao analisar cuidadosamente o diálogo acima, tenha percebido os problemas que certamente poderão surgir em casos semelhantes. Objetivamente, para esta análise adicional, represente-se (fazendo analogia com A e $\bar{A}$) respectivamente por B, $\bar{B}$, C e $\bar{C}$ os eventos Bruno será libertado, Bruno será executado, Carlos será libertado e Carlos será executado. Dessa forma, o conjunto de alternativas para a rainha seria

$$\{(A \bar{B} \bar{C}), (\bar{A} B \bar{C}), (\bar{A} \bar{B} C)\}$$

A falta de informação total sobre a escolha da rainha implica em admitir-se igual probabilidade para as alternativas. Isto é, $P(A \bar{B} \bar{C}) = P(\bar{A} B \bar{C}) = P(\bar{A} \bar{B} C) = 1/3$. Assim, sabendo-se que $\bar{C}$ é verdadeiro então, a probabilidade de ocorrer A seria $1/2$. Formalmente isto é descrito por

$$P(A|\bar{C}) = P(A \bar{B} \bar{C}|\bar{C}) = \frac{P(A \bar{B} \bar{C})}{P(\bar{C})} = \frac{1/3}{2/3} = 1/2$$

pois, $P(\bar{C}) = P(A \bar{B} \bar{C}) + P(\bar{A} B \bar{C}) = 2/3$.

Evidentemente, neste último cálculo não se leva em conta a fonte que informa sobre a ocorrência de $\bar{C}$. No diálogo entre André e Bruno, é fundamental o fato de Marcel poder escolher entre Bruno e Carlos caso a escolha da rainha tenha sido $(A \bar{B} \bar{C})$. Note que no desenvolvimento acima nada foi dito sobre como foi obtida a informação da ocorrência de $\bar{C}$. Chega-se então a um paradoxo, conhecido como paradoxo de Borel. Isto porque o valor de $P(A|\bar{C})$ varia também com o raciocínio utilizado.

O exemplo descrito a seguir, apresenta uma situação mais simples onde o paradoxo fica mais aparente.

Segundo Exemplo:

O senhor Edu, um amante de recreação, estava formando um time de futebol com os meninos de sua rua. Evidentemente necessitava de pelo menos 11 meninos para formar o time que representaria a rua no campeonato do bairro. Infelizmente, perdeu dois de seus jogadores, filhos da sua ex-vizinha que acabara de mudar-se para outra cidade. Por coincidência Mitz, a nova vizinha que chegaria na manhã do dia da inscrição do time, era mãe de duas crianças. Seu irmão, Zico, que alugou a casa para Mitz, informou-lhe por telefone que pelo menos uma das crianças era menino. Representando menino e menina por M e $\bar{M}$, respectivamente. Edu considerou o conjunto de casos possíveis, $\{MM, M\bar{M}, \bar{M}M, \bar{M}\bar{M}\}$ cujos pontos, antes de telefonema de Zico, eram igualmente prováveis, ou seja, $P(MM) = P(M\bar{M}) = P(\bar{M}M) = P(\bar{M}\bar{M}) = \frac{1}{4}$. Após a informação (o telefonema) de que o evento $C = \{MM, M\bar{M}, \bar{M}M\}$ era certo, Edu calculou a probabilidade das duas crianças de Mitz serem meninos e obteve $1/3$ ou seja,

$$P(MM|C) = \frac{P(MM)}{P(C)} = \frac{1/4}{3/4} = 1/3$$

No dia anterior ao da inscrição, Zico veio a casa de Edu trazendo o menino Carlos. Assim que soube que Carlos era filho de Mitz, Edu ficou feliz pois a probabilidade da outra criança ser menino era de $1/2$. Dessa forma, a probabilidade de Mitz ter dois filhos homens aumentou de $1/3$ para $1/2$.

Note que a informação de Edu, pelos menos uma das crianças de Mitz é menino, não mudou. A diferença aqui é apenas a fonte de informação. Inicialmente, havia o telefonema de Zico e posteriormente ele viu Carlos.

O paradoxo de Borel, introduzido pelos dois exemplos, pode ser explicitado pelos modelos probabilísticos se for obedecida a seguinte recomendação: na construção dos modelos estatísticos, não basta indicar-se a informação obtida, é necessário considerarse também de que forma ou de que fonte obteve-se a informação.

Com este trabalho o autor pretende mostrar que uma técnica estatística nunca deve ser indiscriminadamente utilizada. Todo problema científico possui características próprias e como consequência exige cuidados especiais no momento de se optar por uma ou outra metodologia estatística. Uma conclusão científica não pode ficar a mercê da magia de uma técnica ou de um número, desconhecidos do cientistas. Caso contrario, institucionaliza-se a anticiência.

Sugestões para Leitura.

LINDLEY, D.V. - Lectures on Bayesian Statistics - Publicações do IMEUSP, 1984.

PEREIRA, C.A. de B. e VIANA, M.A.G. - Elementos de Inferência Bayesiana
5º SINAPE - IMEUSP, 1982.

DAWID, A.P. & DICKEY, J.M. - Likelihood and Bayesian Inference from
Selectively Reported Data - Journal of the American Statistical Association, 1977, vol. 72, N° 360, p.845-850.