

A DECADÊNCIA DO TESTE t

FREDERICO PIMENTEL GOMES

Escola Superior de Agricultura "Luiz de Queiroz"

Universidade de S. Paulo

1 — INTRODUÇÃO

O teste t , tão conhecido dos estudiosos da Estatística, foi introduzido em 1908 por Student, pseudônimo do matemático-estatístico inglês W. S. Gosset. Sua importância histórica é realmente extraordinária, pois foi o primeiro teste exato para pequenas amostras. O trabalho de Gosset foi posteriormente aperfeiçoado por R. A. Fisher e por outros e assumiu logo uma posição importantíssima na análise estatística moderna. No entanto, o declínio do teste t começou logo. De fato, com a introdução da análise da variância e do teste z , de Fisher, simplificado posteriormente por Snedecor, que criou o teste F , diminuiu a importância de t , que é apenas um caso particular desse teste. Entretanto, continuaram os estatísticos a usar o teste t para a comparação de médias, muito embora já se soubesse há muito que tal procedimento não é exato na maioria dos casos. Ultimamente, porém, com a introdução de métodos exatos para a comparação de médias, maior declínio sofreu o teste de Student, hoje quase sem função. E' de prever-se, pois, que, à medida que se divulguem e se aperfeiçoem os novos métodos, o teste t seja rapidamente deixado de lado como algo já superado por desenvolvimentos mais recentes.

2 — A COMPARAÇÃO DE MÉDIAS NA ANÁLISE DA VARIÂNCIA

A análise da variância, devida a R. A. Fisher, completada pelo teste F (ou seus equivalentes z ou η^2), é o método estatístico mais importante e de maior generalidade atualmente em uso. O teste F é exato e, se significativo, mostra que pelo menos um contraste entre as médias difere de zero. No caso de termos apenas duas médias, o teste F é exatamente equivalente ao t , de sorte que o uso posterior deste não se justifica. Quan-

do, porém, há mais de duas médias a comparar, o uso do teste t é indicado e rigoroso quando feito para avaliar um só contraste escolhido "a priori", ou contrastes ortogonais também previamente designados, cujo número não poderá, necessariamente, exceder o número de graus de liberdade para tratamentos. Isto, aliás, equivale a uma decomposição mais radical destes graus de liberdade, completada pelo teste F (ou z ou θ) aplicado a cada uma das componentes assim obtidas. Alternativamente, pode-se admitir o uso do teste t para avaliar *alguns* contrastes, apesar de não ortogonais, embora este já não seja um procedimento estritamente rigoroso.

O conhecimento dessas limitações do teste t é já relativamente antigo. O próprio R. A. Fisher na seção 24 de seu livro "The Design of Experiments", cuja primeira edição data de 1935, já salientava essas restrições e propunha um método empírico para contorná-las. Apesar disso, porém, experimentadores e estatísticos de todo o mundo usavam e abusavam das chamadas diferenças mínimas significativas, que equivalem ao uso indiscriminado e, portanto, condenado, do teste t . (Veja-se, por exemplo, o que dizem Cochran e Cox (9), seção 3.53, e Snedecor (10), seções 15.5 e 15.6). Para que se julgue bem como é deficiente o uso do teste t nesses casos basta esclarecer, como faz Duncan (3), que o limite de 5% de t quando usado no caso de três médias corresponde na verdade a um nível de 12,23%, e no caso de quatro médias a 20,34%, e assim por diante, até chegar a 91,83% quando temos vinte tratamentos.

Para fazer as comparações entre médias era preciso, pois, criar um novo teste que pudesse substituir com vantagem o teste t . Sugeriu-se, para isso, o uso da "amplitude total estudentizada" ("studentized range"). Tal sugestão parece ter sido feita por W. S. Gosset a E. S. Pearson, conforme afirma Duncan (3), e veio a lume num artigo de D. Newman em 1939. Mas o teste só muito lentamente se generalizou. Aparentemente na ignorância do mesmo, John W. Tukey propôs uma marcha detalhada de análise baseada num novo teste (aproximado) e numa combinação de métodos clássicos. Em 1952, porém, M. Keuls reviveu o uso da amplitude total estudentizada, quase simultaneamente

aperfeiçoado e adotado por Tukey (8), o qual deixou de lado seu próprio método anterior. Mais ou menos ao mesmo tempo Henry Scheffé (5) introduziu um novo método, exato e muito geral, cuja exposição foi posteriormente melhorada e publicada na revista *Biometrika* (6). E em 1953 S. N. Roy e R. C. Bose (11) demonstraram rigorosamente o método da amplitude total estudentizada e provaram que, juntamente com o método de Scheffé, podia ser demonstrado a partir de um processo muito mais geral para a obtenção de intervalos de confiança simultâneos. Estes autores obtiveram também um novo teste para o caso da comparação de médias em experimentos fatoriais.

O magnífico artigo de Scheffé em *Biometrika* (6) atribui a Tukey o uso da amplitude total estudentizada (em forma aperfeiçoada), de sorte que este método ficou conhecido como "método de Tukey".

E' preciso salientar ainda que D. B. Duncan estudou detalhadamente o problema da comparação de médias, tendo publicado em 1947 sua tese sobre o assunto sob o título de "Significance Tests for Differences Between Ranked Variates Drawn from Normal Populations". E elaborou posteriormente dois folhetos notáveis sobre o assunto (3 e 4). Da leitura dos trabalhos de Scheffé, Tukey e Duncan fica claro que houve influência mútua entre os dois últimos e destes sobre o primeiro, de sorte que o método de Scheffé, embora independente da amplitude total, foi inspirado pelas pesquisas de Duncan e Tukey.

3 — CONCLUSÃO

Tanto o método primitivo de Tukey (7) como o de Duncan (3 e 4) parecem definitivamente superados pelos novos métodos de Tukey e Scheffé para a comparação de médias na análise da variância. Os dois últimos processos são, aliás, de aplicação mais simples. Para o caso de experimentos fatoriais o novo teste de Roy e Bose (11), também exato, é importante. Com isto fica o teste t quase sem função e terá de ser usado cada vez menos e só em casos muito especiais.

Os leitores interessados na aplicação dos novos testes de Tukey e de Scheffé poderão consultar, além dos trabalhos ci-

tados acima, um artigo do Autor (12), que é bem detalhado e acessível.

4 — BIBLIOGRAFIA CONSULTADA

- 1 — FISHER, R. A. — "The Design of Experiments", 6a. edição Hafner Publishing Company, Nova York, 1951.
- 2 — FISHER, R. A. — "Student". Annals of Eugenics, IX, parte I: 1-9, 1939.
- 3 — DUNCAN, D. B. — "A Significance Test for Differences Between Ranked Treatments in an Analysis of Variance". The Virginia Journal of Science, 2: 171-189, 1951.
- 4 — DUNCAN, D. B. — "On the Properties of the Multiple Comparisons Test". The Virginia Journal of Science, 3: 49-67, 1952.
- 5 — SCHEFFÉ, Henry — "A Method for Judging all Contrasts in the Analysis of Variance" (mimeografado). Columbia University, Nova York, 1952.
- 6 — SCHEFFÉ, Henry — "A Method for Judging all Contrasts in the Analysis of Variance". Biometrika, 40: 87-104, 1953.
- 7 — TUKEY, John W. — "Comparing Individual Means in the Analysis of Variance". Biometrics, 5: 99-114, 1949.
- 8 — TUKEY, John W. — "Reminder Sheets for Discussion of Paper on Multiple Comparisons" (mimeografado).
- 9 — COCHRAN, William G. e Gertrude M. Cox — "Experimental Designs". John Wiley & Sons, Nova York, 1950.
- 10 — SNEDECOR, George W. — "Statistical Methods" 3a. edição. The Iowa State Press, Ames, 1940.
- 11 — ROY, S. N. e R. C. Bose — "Simultaneous Confidence Interval Estimation" (mimeografado). Institute of Statistics, Raleigh, 1953.
- 12 — PIMENTEL GOMES, Frederico — "A Comparação entre Médias de Tratamentos na Análise da Variância". Anais da Esc. Sup. de Agr. "Luiz de Queiroz" (em publicação), 1954.