Fisher

Guilherme Fernandes Castro de Oliveira

guifernandesco@hotmail.com

Introdução

Figura 1: Ronald Aylmer Fisher

Ronald Aylmer Fisher, FRS (Londres, 17 de fevereiro de 1890 — Adelaide, 29 de julho de 1962) foi um estatístico, biólogo evolutivo e geneticista inglês.

Foi descrito por Anders Hald como "um gênio que criou praticamente sozinho as fundações para a moderna ciência estatística"e Richard Dawkins, que o descreveu como "o maior dos sucessores de Darwin".

Complementando a medida de informação própria para medir incerteza sobre espaços desordenados — proposta com sucesso pela teoria da informação, e publicada em 1949 por Claude Shannon (1916-2001) e Warren Weaver (1894-1978) no livro Teoria matemática da comunicação (The Mathematical Theory of Communication) —, Fisher criou uma medida alternativa de informação apropriada para medir incerteza sobre espaços ordenados.

Carreira

Após se formar em matemática em 1912 pela Universidade de Cambridge, Fisher começou a trabalhar para uma empresa de seguros em Londres em 1913 e após se tornou professor no ensino médio e continuou sua pesquisa na area de estatística. Em 1919, Fisher se envolveu com pesquisa agrícola no centro de experimentos de Rothamsted Research em Harpenden na Inglaterra e desenvolveu novas metodologias e teoria no ramo de experimentos. Durante sua vida, Fisher escreveu 7 livros e publicou cerca de 400 artigos acadêmicos em estatística e genética e é considerado o biologista mais conhecido depois de Charles Darwin.

Premiações:

- Prêmio Memorial Weldon (1930)
- Medalha Real (1938)
- Medalha Guy de Ouro (1946)
- Medalha Copley (1955)
- Medalha Darwin-Wallace (1958)

Teste Exato de Fisher

O teste exato de Fisher é um teste de significância estatística utilizado na análise de tabelas de contingência. Embora na prática ele seja empregado quando os tamanhos das amostras são pequenos, é válido para todos os tamanhos de amostra. É nomeado em homenagem a seu inventor, Ronald Fisher, e é um de uma classe de testes exatos, assim chamados por conta da significância do desvio de uma hipótese nula (e.g., p-valor) que pode ser calculada exatamente, ao invés de depender de uma aproximação que se torna exata no limite conforme o tamanho da amostra cresce para o infinito, como em muitos testes estatísticos.

Fisher disse ter concebido o teste depois de um comentário da Dra. Muriel Bristol, que afirmou ser capaz de detectar se o chá ou o leite foi adicionado primeiro em sua xícara. Ele testou seu pedido no experimento "dama apreciadora de chá".

Propósito do Teste

O teste é útil para dados categóricos, que resultam de classificação de objetos em duas maneiras diferentes; ele é usado para examinar a significância da associação (contingência) entre os dois tipos de classificação. Assim, no exemplo original de Fisher, um critério de classificação poderia ser se o leite ou chá foi colocado na xícara primeiro; o outro poderia ser se a Dra. Bristol pensava que o leite ou chá, fora colocado em primeiro lugar. Queremos saber se essas duas classificações são associados - isto é, se a Dra. Bristol realmente poderia dizer se o leite ou o chá fora servido em primeiro lugar. A maioria dos usos do teste exato de Fisher envolvem, como neste exemplo, uma tabela de contingência 2×2. O p-valor do teste é calculado como se as margens da tabela fossem fixas, isto é, como se no exemplo da degustação de chá, a Dra. Bristol soubesse o número de xícaras com cada tratamento (leite ou chá primeiro) que havia e poderia, portanto, fornecer estimativas com o número correto em cada categoria. Como apontado por Fisher, isso leva, sob a hipótese nula de independência a uma distribuição hipergeométrica dos números nas células da tabela.

Com grandes amostras, um teste qui-quadrado pode ser utilizado nesta situação. No entanto, o valor de significância que ele oferece é apenas uma aproximação, pois a distribuição amostral da estatística de teste é calculada somente aproximadamente igual a teórica da distribuição qui-quadrado teórica. A aproximação é inadequada quando o tamanho das amostras é pequena, ou se os dados são muito desigualmente distribuídos entre as células da tabela, resultando na contagem de células previstas na hipótese nula (os "valores esperados") ser baixa. A regra de ouro usual para decidir se o teste de aproximação da qui-quadrado é bom o suficiente é que o teste qui-quadrado não é adequado quando os valores esperados nas células da tabela de contingência estão abaixo de 5, ou abaixo de 10 quando há apenas um grau de liberdade (esta regra é agora conhecida por ser excessivamente conservadora). Na verdade, para dados pequenos, esparsos, ou não balanceados, o p-valor exato e assintótico podem ser muito diferentes e podem levar a conclusões opostas sobre a hipótese de interesse. Em contraste, o teste exato de Fisher é, como seu nome indica, exato, conforme o procedimento experimental mantém os totais das linhas e colunas fixos, e pode, portanto, ser utilizado independentemente das características da amostra. Torna-se difícil calcular com amostras grandes ou tabelas bem equilibradas, mas felizmente esses são exatamente as condições em que o teste qui-quadrado é apropriado.

Para os cálculos à mão, o teste só é viável no caso de uma tabela de contingências 2×2. No entanto, o princípio do teste pode ser estendido para o caso geral de uma tabela m×n, e alguns pacotes estatísticos fornecem um cálculo (às vezes usando um método de Monte Carlo para obter uma aproximação) para o caso mais geral.

Exemplo do Teste

Por exemplo, uma amostra de adolescentes pode ser dividida em masculino e feminino por um lado, e aqueles que estão e não estão atualmente a estudar para um exame de estatística no outro. Vamos supor, por exemplo, que a proporção de pessoas estudando é maior entre mulheres do que entre os homens, e queremos testar se a diferença de proporções que observamos é significativa. Os dados podem parecer como a tabela abaixo:

	Homens	Mulheres	Total da linha
Estudiosos	1	9	10
Não estudiosos	11	3	14
Total da coluna	12	12	24

A pergunta que se faz sobre esses dados é: sabendo que 10 destes 24 adolescentes são estudiosos, e que 12 dos 24 são do sexo feminino, e supondo que a hipótese nula de que homens e mulheres têm a mesma probabilidade de estudar, qual é a probabilidade de que esses 10 estudiosos seria tão desigualmente distribuídos entre as mulheres e os homens? Se tivéssemos que escolher 10 dos adolescentes ao acaso, qual a probabilidade de que 9 ou mais deles estarem entre as 12 mulheres, e apenas 1 ou menos estarem entre os 12 homens?

Antes de prosseguir com o teste de Fisher, primeiro apresentamos alguns apontamentos. Nós representamos as células pelas letras a, b, c e d, chame os totais das linhas e colunas de totais marginais, e represente o total por n. Assim, a tabela agora tem esse aspecto:

	Homens	Mulheres	Total da linha
Estudiosos	a	b	a + b
Não estudiosos	c	d	c + d
Total da coluna	a+c	b + d	a + b + c + d = n

Fisher mostrou que a probabilidade de obtenção de tais valores é dada pela distribuição hipergeométrica:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,n!}$$

Com os dados acima, isso nos dá:

$$p = \frac{\binom{10}{1}\binom{14}{11}}{\binom{24}{12}} = \frac{10!\,14!\,12!\,12!}{1!\,9!\,11!\,3!\,24!} \approx 0.001346076$$

Para calcular a significância do dado observado, isto é, a probabilidade total de observar os dados como extremos ou mais extremos de a hipótese nula é verdadeira, temos que calcular os valores de p para essas duas tabelas e então somá-los. Isto nos dá um teste unicaudal, com $p \approx 0.001346076 + 0.000033652 = 0.001379728$. Por exemplo, no programa R, este valor pode ser obtido por

```
fisher.test(rbind(c(1,9),c(11,3)), alternative="less")p.value
```

Este valor pode ser interpretado como a soma das evidências providas pelos dados observados - ou qualquer tabela extrema - para a hipótese nula (de que não há diferença nas proporções de estudiosos entre homens e mulheres). Quanto menor o valor de p, maior a evidência para rejeitar a hipótese nula; Então aqui a evidência é forte de que homens e mulheres não são igualmente prováveis de serem estudiosos.

Para testes bicaudais devemos considerar também tabelas que são igualmente extremas, mas na direção oposta. Infelizmente, classificação de tabelas de acordo como se são ou não "tão extremas quanto"é problemática. Uma abordagem usada pela funçãofisher.test em R é computar o p valor assumindo as probabilidades para todas as tabelas menores ou iguais aquela observada. No exemplo, o p-valor bilateral é duas vezes o p-valor unilateral - mas no geral pode diferir substancialmente para tabelas com quantidades pequenas, exceto para o caso de testes estatísticos que tem uma distribuição amostral simétrica.

Controvérsias

Apesar do fato de que o teste de Fisher dá os p-valores exatos, alguns autores têm argumentado que o teste é conservador, isto é, que a verdadeira taxa de rejeição é menor que o nível de significância. A aparente contradição decorre da combinação de estatística discreta com níveis de significância fixados. Para ser mais preciso, considere a seguinte proposta para o teste de significância no nível de 5%: rejeite a hipótese nula para cada tabela para as quais o teste de Fisher atribui um p-valor menor ou igual a a 5%. Porque o conjunto de todas as tabelas é discreto, pode não haver uma tabela para as quais igualdade é alcançada. Se α_e é o maior p-valor menor que 5% que pode realmente ocorrer para alguma tabela, então o teste proposto efetivamente testa no nível α_e. Para amostras de tamanho pequenos, α_e pode ser significantemente menor que 5%. Enquanto este efeito ocorre para qualquer estatística discreta (não somente em tabelas de contingência ou para testes de Fisher), tem sido argumentado que o problema é composto pelas condições de teste de Fisher nas marginais. Para evitar o problema, muitos autores desencorajam o uso de níveis de significância fixos quando lidar com problemas discretos.

Referências

https://pt.wikipedia.org/wiki/Teste_exato_de_Fisher
https://pt.wikipedia.org/wiki/Ronald_Fisher#Biografias_de_Fisher