JORGE KINGSTON

Eng. civil, Estatístico do Ministério da Agricultura,
Professor Catedrático Int. da Faculdade Nacional de Filosofia,
Membro do "Inter American Statistical Institute".

A TEORIA DA INDUÇÃO ESTATÍSTICA

Rio de Janeiro
Serviço Gráfico do Instituto Brasileiro de Geografia e Estatística
1945

Heloisiæ

uxori dilectissimæ

NOTA PRÉVIA

O grande desenvolvimento da estatística teórica nos derradeiros anos do século XIX, deve-se sobretudo a KARL PEARSON. *Pelo seu gênio criador, e pelo seu incansável devotamento de professor e orientador de uma escola de pesquisadores de difusão mundial, a êle devemos a estruturação fundamental, que deu à estatística um caráter verdadeiramente científico.*

Paralelamente a êsse surto na metodologia, ampliou-se vertiginosamente o campo de aplicação das novas técnicas criadas, mormente na elucidação dos problemas da hereditariedade e evolução. Essa constante associação da pesquisa teórica à sua aplicação a fenômenos naturais, que caracteriza a obra pearsoniana, deu-lhe um aspecto singular. Como o material biológico que servia de base a êsses estudos podia ser coletado em grandes amostras, foi nesse pressuposto que êle desenvolveu o seu sistema. Em muitos casos, as conclusões a que chegou foram assim mèramente aproximativas; e, embora essas aproximações fôssem adequadas aos problemas com que lidava, deixavam de o ser quando se teve de estender o seu emprêgo a problemas oriundos de outros setores científicos.

Deve-se a W. S. GOSSET, *que escreveu sob o pseudônimo de* "STUDENT", *o primeiro tratamento do problema das pequenas amostras, quando descobriu empìricamente, em 1908, a distribuição de médias obtidas de um universo normal. A dedução exata dessa distribuição só seria obtida em 1925, e desde então têm sido determinadas as distribuições de várias estatísticas, e estabelecidos métodos rigorosos de tratamento dos dados de observação, sobretudo sob a liderança de* R. A. FISHER, *o sucessor de* PEARSON *na cátedra da Universidade de Londres, e aos quais a Norte América tem trazido notável contribuição, com os trabalhos de* HOTELLING, WALD, WILKS *e outros.*

Note-se que o objetivo dêsses estudos não é tornar desnecessária a coleta de grandes amostras, mas a elucidação das verdadeiras bases do raciocínio estatístico. Não há uma separação estanque entre as duas teorias, mas sim continuidade. E. S. PEARSON *esclarece alhures a diferença essencial entre a antiga e nova concepção. Os métodos antigos eram apropriados para tratar com duas amostras de 100 elementos; os novos métodos abrangem, além dêsse, o caso de 100 amostras de dois elementos. A distinção é importante: em ambos os casos, temos o mesmo número de dados; mas, em certos problemas, a compreensão e contrôle adequado da situação só se podem obter secionando as observações em um grande número de pequenos grupos.*

O desenvolvimento moderno da estatística se tem processado, pois, no sentido de estabelecer as exatas distribuições por amostragem dos vários índices estatísticos. Êsse conhecimento é absolutamente necessário, em primeiro lugar, para podermos escolher as estatísticas mais eficientes, que permitirão, sôbre uma dada amostra, obter a melhor aproximação dos verdadeiros valores das características da população donde provém a mesma. Chega-se, assim, ao conceito de estimação,

associado ao princípio da máxima verossimilhança *de* FISHER, *e cuja fundamentação exige novos processos de lógica indutiva, além dos moldes da teoria clássica da probabilidade.*

Em segundo lugar, é sôbre êsse conhecimento que se apoiam os modernos testes de significância, *a que se ligam os nomes de* E. S. PEARSON *e* NEYMAN, *e cujo emprêgo é indispensável para podermos discriminar se as diferenças e relações aparentemente observadas são efetivamente reais ou devidas a causas fortuitas.*

Finalmente, outra característica da moderna conceituação diz respeito às interconexões entre a análise estatística e o planejamento dos experimentos, que, consoante R. A. FISHER, *se devem considerar apenas como dois aspectos de um único problema. Trata-se de questões de* eficiência, *de determinação da forma da experiência, de modo a dela poder extrair o máximo de informações. Pôs-se assim a lume a inanidade do princípio tradicional de que, numa experiência ideal, apenas um dos fatôres devia variar, todos os outros sendo mantidos constantes. Ao contrário, a combinação de vários fatôres numa experiência única, de arranjo adequado, onde êles variem concorrentemente segundo as suas possíveis combinações, fornece, quase sem aumento do trabalho ou custo correspondente à experimentação de um só, maior soma de informações do que no caso em que cada um dêles fôsse experimentado separadamente, por isso que esclarece também sôbre as interações entre os vários fatôres.*

São alguns dos aspectos fundamentais dessas novas teorias que afloramos no presente trabalho. Creio que já é tempo, entre nós, de se processar a renovação do conteúdo dos manuais e programas de ensino da estatística. A aplicação dos métodos modernos não é mais difícil que a dos antigos, muitos dos quais, obsoletos, grosseiramente aproximados, ou mesmo errôneos, continuam a ser expostos e ensinados rotineiramente.

Papel primacial na divulgação das novas técnicas deverá caber, sem dúvida, aos professôres de estatística. Urge, pois, que êles se compenetrem de que não se trata de uma ciência fossilizada, mas que a sua matéria se renova de dia a dia, com ímpeto cada vez maior. A formação adequada do professorado é uma questão primordial, para que não permaneçamos impermeáveis às novas conquistas.

O depoimento que, sôbre o recrutamento de professôres, nos presta um grande mestre, HAROLD HOTELLING, *é altamente pitoresco, mas verídico. A escolha faz-se quase segundo as leis do acaso, recaindo num qualquer jovem* JONES, *que então se inicia no estudo da disciplina. "He pursues "statistics" through the library card catalog and the encyclopedias... Perhaps he encounters probable errors. Eventually he learns that* KARL PEARSON *is the great man of statistics, and that* Biometrika *is the central source of information. Unfortunately most of the papers in* Biometrika *and of* PEARSON'S *writings, while not lacking in vigor, trail of into mathematical discourse of a kind with which young* JONES *feels ill at ease. What he wants is a textbook, couched in simple language and omitting all mathematics, to make the subject clear to a beginner...* JONES *decides that a simple book on statistics must be written, and that he will do it if he can ever succeed in mastering the subject".*[1] *Aos trambulhões, êle leva o curso a final, contentando-se entrementes com "such nonmathematical textbooks*

---

[1] HOTELLING, H., "The Teaching of Statistics", *Annals of Mathematical Statistics*, vol. XI (1940), pág. 460.

as may have been written by other young men who have earlier trod the same path". Daí a proliferação malsã de compêndios de forma estereotipada, onde as falhas e erros se repetem continuadamente. O mal, como se vê, não é apenas nosso...

O conhecimento da estatística não é assim fácil de se adquirir. Isso provém, não tanto do lato emprêgo que aí se faz das matemáticas superiores, mas ainda de que a bibliografia pertinente se acha espalhada por um sem número de revistas, correspondendo às suas mais variadas aplicações. E o pior é que "the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error". Não apenas. "He must also contend with the fact that a good deal that is important in statistics is still a matter of oral tradition, and some consists of laboratory techniques".[2]

Tivemos a sorte de poder aperfeiçoar os nossos conhecimentos de estatística — embora por um período que empecilhos de ordem administrativa, a nosso pezar, excessivamente encurtaram — no grande centro que é a COLUMBIA UNIVERSITY. *Cumprimos um dever ao consignar aqui a nossa gratidão a nossos mestres, os professôres* H. HOTELLING, F. C. MILLS *e* A. WALD, *e ao professor* E. J. GUMBEL, *da* NEW SCHOOL FOR SOCIAL RESEARCH, *pelo muito que lhes devemos de nossa iniciação nessas teorias. Especialmente nos penhorou o professor* HOTELLING, *assim pelo generoso apoio à concessão da bôlsa de estudos, como pela infatigável dedicação em orientá-los e em dirimir nossas dúvidas.*

*Mais do que poderíamos externá-lo, é imenso o nosso reconhecimento para com a "*JOHN SIMON GUGGENHEIM MEMORIAL FOUNDATION*", que, numa distinção desvanecedora, nos ensejou êsse período de estudos. Uma entre tantas florações do espírito liberal norte-americano, fixou-lhe as diretrizes o seu fundador, o Senador* GUGGENHEIM, *nos seguintes têrmos: "We strongly hope that this Foundation will advance human achievement by aiding students to push forward the boundaries of understanding, and will enrich human life by aiding them in the cultivation of beauty and taste". E, seguindo êsses ditames, vem a Fundação, com esplêndida generosidade, contribuindo para o avanço da cultura, no seu país e na América Latina. Restringindo-nos a exemplos no setor estatístico-econômico, foi sob seus auspícios que realizaram trabalhos notáveis cientistas e professôres do porte de* EZEKIEL, HANSEN, KNIGHT, LEONTIEF, ROOS, SCHULTZ, *e tantos outros, que, nas cátedras universitárias, na orientação dos negócios públicos ou na administração das indústrias, vêm cumprindo o lema da Fundação: "There is, moreover, a republic of learning and art which knows no boundary lines, and we desire only that scholars and artists of the American republics should meet and learn and teach what to them is Truth".*

Consideramos como inestimável privilégio o termos sido agraciados com uma bôlsa da FUNDAÇÃO GUGGENHEIM, *a que nomes tão ilustres têm emprestado o brilho de sua atuação. Recebemo-la como um incentivo para porfiar sempre nos trabalhos de investigação pura.*

Finalmente, agradecemos ao INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, *e ao seu eminente Secretário-Geral,* DR. M. A. TEIXEIRA DE FREITAS, *as facilidades concedidas para a impressão dêste trabalho.*

---

[2] HOTELLING, H., *Op. cit.*, págs. 462 e 463.

A TEORIA DA INDUÇÃO ESTATÍSTICA

Capítulo I

OS PROBLEMAS DA ESTATÍSTICA TEÓRICA

1.1 Estatística Descritiva e Indutiva.

Os problemas da estatística teórica têm sido correntemente classificados em dois grupos distintos.[1] No primeiro incluem-se aquêles que têm por objetivo descrever as séries de observações mediante reduzido número de elementos característicos. As peculiaridades mais importantes do fenômeno são assim sumariadas, e as constantes calculadas sintetizam o máximo, senão tôdas as informações contidas na série original. É a parte conhecida como Estatística Descritiva, porque apenas se propõe a descrição dos dados, independentemente de qualquer inferência indutiva.

A caracterização da série empírica faz-se, em geral, calculando certas funções dos dados observados, denominadas *estatísticas,* como as medidas de tendência central, assimetria, kurtosis, etc., ou ainda, de maneira mais lata, pesquisando a sua representação mediante uma função matemática. Karl Pearson estabeleceu um sistema de curvas contínuas, às quais se reconduz a maioria das distribuições obtidas na prática, ao passo que a escola escandinava de Thiele e Charlier chega aos mesmos resultados utilizando desenvolvimentos em séries. Contudo, êsses ajustamentos, pela deficiência de enquadramento num sistema lógico geral, podem-se considerar apenas como um tratamento preliminar dos dados experimentais, servindo de base às induções teóricas finais.[2]

O que importa, com efeito, é generalizar as conclusões, induzir as características de fenômenos não observados das constatadas nos que foram observados. Postula-se então que os dados com que lidamos constituem uma *amostra,* escolhida ao acaso, de um grupo mais geral ou infinito de dados, denominado *população* ou *universo,* e o problema que defrontamos é o de ajuizar sôbre os elementos característicos dessa população, denominados *parâmetros,* a partir da amostra observada. As inferências são aqui de natureza indutiva, e a esta parte denomina-se Estatística Indutiva. O seu escôpo é estimar a grandeza dos parâmetros no universo correspondente à série empírica, e as possíveis variações nas estatísticas oriundas de flutuações de amostragem.

1.2 Os problemas da Estatística Teórica.

A divisão dicotômica assim introduzida na estatística é, contudo, puramente formal. Na realidade, ambas as partes se entrosam intimamente.

Assinala R. A. Fisher[3] que a redução dos dados de observação a um certo número de características só se pode fazer construindo uma

---

[1] Rietz, H. L., *Mathematical Statistics* (Chicago, 1927), pág. 4; Kenney, J. F., *Mathematics of Statistics* (Nova Iorque, 1939), vol. 2, pág. 97.

[2] von Mises, R., *Probability, Statistics and Truth* (Londres, 1939), pág. 235.

[3] Fisher, R. A., "On the Mathematical Foundations of Theoretical Statistics", *Phil. Trans. Roy. Soc. of London,* Ser. A, vol. 222 (1922), pág. 312.

população hipotética infinita, da qual os dados se consideram como constituindo uma amostra ao acaso. A lei de distribuição dessa população contém apenas poucos parâmetros, que são suficientes para descrevê-la exaustivamente em relação aos atributos sob estudo.

A existência dessa população está implícita na concepção de uma distribuição de freqüência capaz de sintetizar os dados empíricos. Com efeito, nenhuma amostra finita pode ser descrita por uma curva de freqüência; ela só o pode ser por um histograma ou um polígono de freqüência. A construção da curva populacional implica a existência de um número infinito de elementos em cada classe, ao mesmo tempo que o número de classes em que se subdivide a população também se torna infinito.

Por conseguinte, o conceito de curva de freqüência corresponde à postulação de um universo hipotético infinito, distribuído segundo uma função matemática f da variável x e de m parâmetros. Isto é, a probabilidade estatística de que um valor observado de x caia no intervalo $x \pm \frac{1}{2}\, dx$ é dada por

$$dp = f(x, \theta_1, \theta_2, \ldots \theta_m)dx$$

em que os parâmetros θ são suficientes para caracterizar o universo. O problema fundamental é, pois, a descoberta ou inferência dessa lei a partir dos dados de observação. Uma série estatística de n elementos corresponde a uma amostra de tamanho n colhida nessa população, e o método estatístico tem por fim extrair dêsses dados os elementos relevantes de informação acêrca dos parâmetros θ, rejeitando os demais como irrelevantes.

Conceituação análoga emana da obra de VON MISES, quando estabelece que o problema geral da estatística é verificar se uma série de dados de observação goza das propriedades de um grupo finito derivado de uma seqüência infinita possuindo as características de um "coletivo", isto é, tal que as freqüências relativas dos atributos particulares dos elementos da seqüência tendam para limites fixos, independentemente da ordem de seleção.

No tratamento estatístico dos dados observacionais, distingue FISHER três tipos de problemas:

1) problemas de *especificação,* correspondendo à fixação da forma matemática da população;

2) problemas de *distribuição,* envolvendo a dedução das distribuições de estatísticas em amostras extraídas ao acaso de um universo de forma específica;

3) problemas de *estimação,* compreendendo a escolha dos métodos para calcular, a partir da amostra, as estatísticas capazes de fornecer uma boa avaliação dos parâmetros do universo, e a verificação de hipóteses formuladas sôbre os mesmos.

1.3 A especificação do universo.

O problema da especificação consiste essencialmente em estabelecer a forma funcional de f, dependente dos parâmetros θ desconhecidos, de modo a poder utilizá-la como base hipotética para a solução dos dois outros problemas mencionados.

A escolha do tipo de função é assunto pertinente ao estatístico prático, que se guia pela experiência para construir um modêlo teórico do fenômeno que observa. Evidentemente, devem tais modelos se restringir a funções de propriedades conhecidas, e cujo cálculo seja facilitado por tabelas existentes. Um dos mais antigos exemplos de especificação é a lei gaussiana dos erros. Desenvolvimentos subseqüentes levaram à introdução do sistema de curvas assimétricas de PEARSON e aos desenvolvimentos em séries de GRAM-CHARLIER.

Consideremos, por exemplo, um conjunto de observações dando os valores do custo da alimentação y correspondente a vários níveis de renda x. ENGEL mostrou que existe uma relação linear ligando essas variáveis

$$Y = \theta_1 + \theta_2 x.$$

É esta porém uma relação ideal, e, na prática, os valores observados y flutuam em tôrno de Y, de maneira que Y é apenas o valor médio correspondente a um dado x. Podemos admitir que y se distribua normalmente em tôrno de Y, isto é, segundo a lei

$$dp = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-Y}{\sigma}\right)^2} dy.$$

Completa-se assim o problema da especificação. Os dados da amostra permitem estimar os valores dos parâmetros desconhecidos θ_1, θ_2 e σ.

Muitas vêzes o estudo do fenômeno não levanta questões de especificação, porque a natureza qualitativa da população hipotética é conhecida; é o que se dá, por exemplo, nos estudos de hereditariedade mendeliana.

A especificação do universo deve ser tão realística quanto possível; entretanto, costuma-se freqüentemente introduzir hipóteses simplificadoras, de modo a ensejar o uso de métodos simples de estimação dos parâmetros e o emprêgo de conhecidos testes de significância. Daí a utilização generalizada da distribuição normal. Nesses casos, devemos porém ter em vista que tôdas as condições do problema, que foram abandonadas na simplificação da forma funcional, devem ser reconsideradas ao serem interpretados os resultados.

1.4 Problemas de distribuição.

A fim de verificar a hipótese de que um conjunto de valores de certa variável, obtidos presumidamente sob as mesmas condições, constitui uma amostra derivada fortuitamente de um universo hipotético, é necessário conhecer as possíveis variações oriundas do processo de amostragem, de maneira a poder compará-las com as realmente observadas. O mesmo critério se aplica quando se trata de funções das observações, ou seja, de *estatísticas*. Sòmente pelo estudo da distribuição por amostragem das diferentes estatísticas propostas, é que podemos nos guiar na escolha daquelas cujo emprêgo é o mais vantajoso.

Suponhamos que, de um universo contendo N elementos, extraimos amostras de tamanho n, estritamente ao acaso, isto é, de modo que a probabilidade de extração de cada elemento seja a mesma que a dos demais. Os resultados das extrações denominam-se *amostras*

aleatórias.[4] Se os elementos constituintes das amostras são anotados na ordem de extração e com reposição após cada tiragem, obtém-se assim N^n amostras. O processo denomina-se então de *amostragem simples* ou *irrestrita,* caracterizando-se por ser a probabilidade de extração de cada elemento igual e independente da dos demais. Êsse tipo corresponde também à amostragem de um universo infinito.

Outras condições de amostragem podem restringir o número de amostras de dado tamanho, deriváveis de um universo particular, mas quando tôdas as possíveis amostras, que dêle se podem extrair sob essas condições, são classificadas segundo o valor de uma certa estatística, a distribuição de freqüência que assim se obtém denomina-se a *distribuição por amostragem* dessa estatística.

O escopo da teoria da amostragem é exatamente a derivação e interpretação das características dessas distribuições. Por exemplo, a fórmula do êrro padrão da média caracteriza a dispersão da distribuição por amostragem da média aritmética em função do número de elementos na amostra e dos parâmetros do universo. Se êste é normal, êsse êrro padrão é uma das constantes da curva normal que descreve a distribuição de médias.

Nem sempre se conhece a forma funcional da distribuição, e por vêzes ela é de natureza tão complexa que há, em ambos os casos, vantagens em se utilizar funções de freqüência de propriedades conhecidas, como aproximações daquelas. Lembremos, a propósito, o extensivo uso que se faz da integral da curva normal na interpretação de erros padrões obtidos de grandes amostras. De grande alcance prático é também a possibilidade de reduzir as distribuições referentes a várias estatísticas a um pequeno número de distribuições fundamentais, pois assim se limita o número de tabelas a serem empregadas na verificação de hipóteses estatísticas.

A caracterização das distribuições por amostragem faz-se através das mesmas medidas de tendência central, dispersão, assimetria e kurtosis utilizadas para as comuns distribuições de freqüência; em particular, o desvio padrão denomina-se o *êrro padrão* da estatistica aferente.

Antigamente usava-se preferentemente o *êrro provável,* que tem sido abandonado por corresponder a duas definições divergentes: a) ao desvio quartílico da distribuição; b) a 0.6745 vêzes o êrro padrão. A identidade entre ambos os conceitos só se dá rigorosamente para as distribuições normais. Depois, a probabilidade de 50 por cento a que corresponde a definição é de nenhum valor na interpretação de testes de significância, que exigem probabilidades muito menores.

1.5 Estimação de parâmetros.

O problema da estimação tem por finalidade, dada uma amostra extraída ao acaso de um universo de forma funcional conhecida, determinar as estatísticas que melhor permitem avaliar as grandezas dos m parâmetros desconhecidos que caracterizam o universo segundo a equação

$$dp = f(x, \theta_1, \theta_2, \ldots \theta_m)\, dx\ .$$

---

[4] Na terminologia inglêsa "random samples". Note-se que a qualificação de aleatória refere-se mais pròpriamente à operação de escolha que à amostra mesma. Por isso, não nos parece adequada a tradução de "amostras equiprováveis", proposta pelo Dr. Octavio L. Martins. Com efeito, suponhamos que se joga com 10 moedas; a probabilidade de obter uma amostra de 10 caras não é a mesma de 5 caras e 5 coroas.

Se t_i representa a estatística correspondente ao parâmetro θ_i, obtemos assim a equação

$$dy = f(x, t_1, t_2, \ldots t_m)dx$$

como aproximação teórica da equação anterior, que representa a verdadeira distribuição de probabilidade.

Seja, por exemplo, uma distribuição normal, como a do § 1.3. A estimação do parâmetro σ pode ser feita de diferentes modos, entre os quais os que utilizam as fórmulas familiares

$$t_3 = \sqrt{\frac{\pi}{2}} \Sigma \frac{|y - Y|}{n} \quad \text{e} \quad t'_3 = \sqrt{\Sigma \frac{(y - Y)^2}{n}},$$

a somação estendendo-se a todos os n valores da amostra. A teoria da estimação estabelece critérios que permitem a escolha da forma mais vantajosa.

Um dêsses critérios é que a estima do parâmetro que se venha a obter se deve caracterizar pelo menor êrro ou maior precisão. No exemplo acima, a estima de t_3 exige uma amostra de 1.14 n para dar a mesma precisão que a estima de t'_3 baseada numa amostra de tamanho n, porque a razão do êrro de t_3 para t'_3 é $\sqrt{1.14}$. Ainda outros critérios devem ser atendidos.

Notemos que, tanto o processo de estimação, como o de amostragem, podem ser *tendenciosos* (biased), cumprindo discriminar os erros advindos num e noutro caso. Seja $\bar{t}_i$ a média da distribuição por amostragem da estatística t_i, adotada como estimativa do parâmetro θ_i; e seja $\tilde{t}_i$ a média numa distribuição de amostras obtidas rigorosamente ao acaso. Se a diferença $\tilde{t}_i - \bar{t}_i$ fôr nula, o processo de amostragem é *justo* (unbiased); caso contrário, o processo e os erros de amostragem se dizem tendenciosos. Em contraposição, quando $\tilde{t}_i$ iguala θ_i, o processo de estima é justo, e, sob condições de amostragem ao acaso, a estatística t_i chama-se uma *estimativa justa* do parâmetro θ_i. A diferença $\tilde{t}_i - \theta_i$, quando diferente de zero, dá-nos uma medida da tendenciosidade introduzida pelo processo de estimação.

Um critério muito empregado para a escolha do processo de estimação é o da *máxima verossimilhança* (maximum likelihood), que apresenta certas vantagens de precisão sôbre os demais. A probabilidade de que o conjunto parcial de valores constituindo a nossa amostra provenha de uma população, caracterizada por certa distribuição e certos parâmetros, é uma função dêsses parâmetros. A verossimilhança define-se como um múltiplo constante dessa probabilidade; e o critério de estima consiste em escolher os valores dos parâmetros que tornam máxima a verossimilhança. Os valores assim estimados denominam-se *estimativas ótimas* dos parâmetros. Por exemplo, as médias de amostras provenientes de universos normais são estimativas ótimas, ao passo que as variâncias têm de ser ajustadas segundo os graus de liberdade, a fim de satisfazer o critério de máxima verossimilhança.

1.6 Verificação de hipóteses estatísticas.

Estreitamente ligados à estimação, estão os processos de verificação de hipóteses estatísticas. Com efeito, uma das finalidades da análise estatística é estabelecer critérios para a aceitação ou rejeição de hipóteses que se possam formular sôbre os fatôres causais dos fenô-

menos observados, tal como influenciaram a amostra colhida. O critério ou critérios são calculados a partir da mesma, e confrontados com as suas distribuições teóricas, a fim de se obter uma medida da verossimilhança da hipótese. Se a probabilidade resultante é pequena, rejeita-se a hipótese; caso contrário, é ela aceita.

Uma hipótese que se submete à verificação, no pressuposto de que seja verdadeira, foi chamada por Fisher de *hipótese nula*. Isto é, postula-se a existência de uma população hipotética com certas características, e pesquisa-se até que ponto uma amostra como a obtida, poderia razoàvelmente dela provir. Devido às flutuações de amostragem não se pode, contudo, estabelecer um limite definido, a partir do qual seria impossível obter da população a amostra em causa; podemos apenas calcular a maior ou menor probabilidade de obtê-la. Conforme essa probabilidade seja pequena ou grande, a hipótese é rejeitada ou aceita, e, nesse último caso, atribuem-se os desvios entre os valores da amostra e da população a erros de amostragem. O que fixamos, pois, é o valor da probabilidade, a partir do qual rejeita-se a hipótese: é o que se chama *nível de significância* para a verificação da hipótese. Por exemplo, numa estatística com distribuição normal, a probabilidade de um desvio por amostragem maior de $\pm$ 1.96 vêzes o êrro padrão é de 5 por cento; se o valor da estatística, calculada para a amostra, corresponde a um desvio maior que o citado, diremos que a hipótese nula é rejeitada no nível de significância de 5 por cento. Qualquer desvio, correspondente a uma baixa probabilidade, que leve à rejeição da hipótese, considera-se como estatìsticamente significante.

Desde logo apresenta-se o problema: para que probabilidade devemos considerar um desvio como significante, isto é, que nível de significância devemos adotar? Os motivos que determinam a escolha dêsse nível serão ponderados mais tarde, mas antecipamos que uma regra prática que tem dado resultados satisfatórios é adotar 5 por cento como o nível crítico, o que corresponde aproximadamente, nas distribuições normais, a um desvio igual a 2 vêzes o êrro padrão ou 3 vêzes o êrro provável. Então, se a probabilidade P_δ de um desvio igual ou maior que $|\delta|$ fôr $P_\delta \geqslant .05$, δ não é significante; se tivermos $.05 > P_\delta > .01$, δ é significante; e se $P_\delta \leqslant .01$, δ é altamente significante. Outros autores exigem para a significância que $P_\delta \leqslant .01$, e para valores compreendidos entre .05 e .01, reputam necessárias informações adicionais para concluir pela significância do resultado.

Capítulo II

SIGNIFICÂNCIA DE MÉDIAS E OUTRAS ESTATÍSTICAS

2.1 *Valores médios e momentos.*

Observamos que nem sempre era possível, devido a dificuldades analíticas, determinar a forma funcional da distribuição de estatísticas em amostras aleatórias provindas de um universo especificado. Recorre-se então, para caracterizar a distribuição, ao cálculo de seus momentos.

Seja uma variável contínua x, distribuída segundo uma lei $f(x)$, sendo essa função monotônica e positiva, de modo que a freqüência relativa ou probabilidade de ocorrência de x no intervalo $a < x < b$ seja medida por $\int_a^b f(x)\,dx$.

Diz-se, então, que x é uma variável aleatória ou casual, e $f(x)$ a sua *distribuição* ou *função de probabilidade*. A diferencial $dp = f(x)\,dx$ dá a probabilidade de que x esteja contido no intervalo $x \pm \frac{1}{2}\,dx$.

Consideremos agora uma função arbitrária $\phi(x)$. Denomina-se *valor médio* ou *provável*, ou ainda *expectância* [1] dessa função, e representa-se mediante o operador E, a integral

$$E\left\{\Phi(x)\right\} = \int_{-\infty}^{\infty} \Phi(x) f(x)\,dx.$$

Dessa definição decorrem as seguintes propriedades da expectância, de verificação imediata:

1) a expectância de uma soma de variáveis é igual à soma de suas expectâncias,

$$E(x + y + \ldots) = E(x) + E(y) + \ldots,$$

2) a expectância do produto de uma variável por uma constante é igual ao produto da constante pela expectância da variável,

$$E(cx) = cE(x),$$

3) a expectância de um produto de variáveis independentes é igual ao produto de suas expectâncias,

$$E(x.y.z \ldots) = E(x).\,E(y).\,E(z)\ldots$$

Se, em particular, considerarmos uma função do tipo $\phi(x) = x^k$, sendo $k = 1, 2, \ldots$, teremos

$$E(x^k) = \int_{-\infty}^{\infty} x^k f(x)dx,$$

que é o momento μ'_k de ordem k da variável aleatória x em relação à origem. É o valor médio da potência k^{esima} da grandeza x.

---

[1] Preferimos essa denominação à usual de "esperança matemática" por ser têrmo castiço e afastar qualquer subjetivismo do conceito, além de corresponder ao inglês "expectation".

Para $k = 1$, temos $E(x) = \mu'_1 = \tilde{x}$, sendo $\tilde{x}$ a média do universo. Para $k = 2$, temos o momento quadrático $E(x^2) = \mu'_2$. Chama-se variância o momento quadrático da grandeza aleatória $(x - \tilde{x})$, isto é,

$$\sigma_x^2 = E(x - \tilde{x})^2 = E(x^2) - (\tilde{x})^2 = \mu'_2 - (\mu'_1)^2$$

A raiz quadrada da variância é o desvio ou êrro padrão da distribuição.

Se tivermos duas variáveis aleatórias x e y com uma distribuição conjunta, $f(x,y)$, definiremos a expectância de $\varphi(x,y)$ como

$$E\left\{\varphi(x,y)\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x,y) f(x,y)\, dx\, dy$$

Não sendo x e y independentes, no sentido estatístico, então o coeficiente de correlação ρ do universo será por definição

$$\rho = \frac{E\left\{(x - \tilde{x})(y - \tilde{y})\right\}}{\sigma_x \sigma_y}.$$

2.2 Média e variância de uma função linear.

Consideremos uma variável w, que seja função linear de outras variáveis x_i $(i = 1,2 \ldots n)$, cada uma das quais distribuída arbitrária e independentemente; isto é,

$$w = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n,$$

onde os c_i são constantes quaisquer. Calculemos a média e a variância de w em função dos momentos das distribuições das variáveis x_i. Seja σ_i^2 a variância de x_i no universo a que pertence, e ρ_{ij} o coeficiente de correlação entre x_i e x_j. Então, a expectância de w será

$$E(w) = c_1 E(x_1) + c_2 E(x_2) + \ldots c_n E(x_n)$$

isto é

$$\tilde{w} = c_1 \tilde{x}_1 + c_2 \tilde{x}_2 + \ldots + c_n \tilde{x}_n. \qquad (2.1)$$

Temos ainda

$$E(w - \tilde{w})^2 = \sum_{i=1}^{n} c_i^2 E(x_i - \tilde{x}_i)^2 + \sum_{i \neq j} c_i c_j E\left\{(x_i - \tilde{x}_i)(x_j - \tilde{x}_j)\right\}$$

ou seja, introduzindo os coeficientes de correlação entre as variáveis,

$$\sigma_w^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2 + \sum_{i \neq j} c_i c_j \rho_{ij} \sigma_i \sigma_j$$

Se, porém os x_i são mùtuamente independentes, no sentido estatístico, de modo que $\rho_{ij} = 0$, a fórmula simplifica-se para

$$\sigma_w^2 = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + \ldots + c_n^2 \sigma_n^2. \qquad (2.2)$$

2.3 Êrro padrão da média.

Apliquemos essas fórmulas a distribuições de médias de amostras aleatórias de um universo arbitrário. Suponhamos os x_i provindo to-

dos de um mesmo universo, constituindo amostras de tamanho n. Se fizermos $c_1 = c_2 = \ldots = c_n = 1/n$, será w igual à média da amostra, $w = \bar{x}$, e, pelas condições de amostragem, $E(x_i) = \tilde{x}$ para cada um dos valores i. Portanto,

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \tilde{x}.$$

Isto é, a média de tôdas as possíveis amostras é igual à média do universo. Noutros têrmos, a média da amostra constitui uma estimativa justa de $\tilde{x}$.

Anàlogamente, pondo em (2.2) $w = \bar{x}$, e notando que $\sigma_i^2 = \sigma_x^2$ para todos os valores i, segue-se que

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_x^2 = \frac{\sigma_x^2}{n}$$

Isto é, a variância da distribuição de médias é igual à do universo dividida pelo número de elementos na amostra. Daí se conclui o êrro padrão da média como

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \qquad (2.3)$$

Em vista do uso corrente da distribuição normal como uma aproximação da distribuição por amostragem da média, convém verificar como esta distribuição se normaliza com o aumento do tamanho da amostra. A forma da distribuição depende dos momentos de ordem superior (assimetria e kurtosis). Poderíamos obtê-los pela repetição do processo acima; mas é preferível recorrer à noção de função característica.

2.4 Função característica.

A expectância da função e^{tx}, onde t é uma variável ordinária e x uma grandeza aleatória de distribuição $f(x)$, denomina-se *função característica* ou *função geratriz de momentos* [2] de $f(x)$. Representando-a por $\Psi(t)$, temos

$$\Psi(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx.$$

Essa única integral permite calcular todos os momentos de $f(x)$. Com efeito, supondo que e^{tx} se possa desenvolver em série de MACLAURIN, temos

$$\Psi(t) = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \ldots$$

O momento μ'_k de ordem k é, pois, o coeficiente de $\frac{t^k}{k!}$ no desenvolvimento da função característica. Podemos também definí-lo como a derivada de ordem k dessa função

---

[2] É de se estranhar que autores franceses, repetindo P. LÉVY (*Calcul de Probabilités*, Paris, 1925, pág. 161), atribuam a introdução da função característica a CAUCHY, em 1853, quando essa noção já aparece na obra de LAPLACE (*Théorie Analytique des Probabilités*, 3.ª ed., pág. 83), quando transforma a sua função geratriz mediante a substituição $t^x = e^{i\omega x}$.

$$\mu'_k = \left[\frac{d^k \Psi(t)}{dt^k}\right]_{t=0}$$

Sejam várias distribuições independentes $f(x_1), f(x_2), \ldots$ às quais correspondem as funções características $\Psi_1(t)$, $\Psi_2(t), \ldots$; a função característica $\Psi(t)$ aferente à soma das variáveis $x = x_1 + x_2 + \ldots$ não é senão o produto das funções $\Psi_1(t) . \Psi_2(t) \ldots$. Com efeito, a exponencial $e^{tx} = e^{t(x_1+x_2+\ldots)}$ é o produto das exponenciais $e^{tx_1} e^{tx_2} \ldots$; logo, a integral múltipla

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{tx_1} e^{tx_2} \ldots f(x_1) f(x_2) \ldots dx_1 dx_2 \ldots$$

é o produto de integrais simples $\int_{-\infty}^{\infty} e^{tx_i} f(x_i)\, dx_i$ isto é

$$\Psi(t) = \Psi_1(t) . \Psi_2(t) \ldots$$

Se pudermos desenvolver o logarítmo da função característica $\Psi(t)$ em série convergente segundo as potências de t, teremos

$$L(t) = \log_e \Psi(t) = \lambda_1 t + \lambda_2 \frac{t^2}{2!} + \lambda_3 \frac{t^3}{3!} + \ldots,$$

e os coeficientes λ_k serão os semi-invariantes da função $f(x)$, introduzidos por THIELE. $L(t)$ denomina-se a *função geratriz de semi-invariantes*.

Assim como as $\Psi_i(t)$ compõem-se segundo um produto, as $L_i(t)$ compõem-se segundo uma soma; e daí o teorema: quando uma distribuição resulta da adição de várias distribuições independentes, o seu semi-invariante λ_k é a soma dos semi-invariantes da mesma ordem das distribuições parciais. E' precisamente essa propriedade *aditiva* dos semi-invariantes que justifica o seu emprêgo.

Exemplo A. Calculemos a função característica da distribuição normal. Temos

$$\Psi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} e^{tx}\, dx = e^{\frac{\sigma^2 t^2}{2}} . \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\sigma^2 t)^2}\, dx = e^{\frac{\sigma^2 t^2}{2}}$$

Desenvolvendo $\varphi(t)$ em série, obtém-se

$$\Psi(t) = \sum_0^{\infty} \frac{1}{k!} \left[\left(\frac{t^2}{2}\right)^k \sigma^{2k}\right]$$

Todos os momentos ímpares se anulam, e o momento par de ordem $2k$ é dado por

$$\mu_{2k} = \frac{1}{k!} \left[\left(\frac{1}{2}\right)^k \sigma^{2k} (2k)!\right]$$

Em particular, $\mu_4 = 3\sigma^4$, de modo que o coeficiente de kurtosis é $\beta_2 = 3$.

Observemos que $\log_e \Psi(t) = \frac{1}{2}\sigma^2 t^2$. O único semi-invariante, além da média, é assim o de ordem 2, que tem por valor σ^2.

Exemplo B. Seja uma grandeza aleatória descontínua, que admite os valores $x = 0, 1, 2, \ldots$ com probabilidade $\frac{\mu^x}{x!} e^{-\mu}$ (função exponencial de Poisson). A sua função característica será

$$\Psi(t) = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} e^{tx} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{\mu(e^t - 1)}.$$

2.5 *Normalização da distribuição de médias.*

Consideremos a amostra $x_1, x_2 \ldots x_n$. Em vista da atividade dos semi-invariantes, temos que, provindo todos os x do mesmo universo, o semi-invariante de ordem k da soma dêsses n elementos será igual a n vêzes o semi-invariante do universo, $n\lambda_{k:x}$; e portanto, para a distribuição de médias de amostras de tamanho n, teremos

$$\lambda_{k:\bar{x}} = \frac{\lambda_{k:x}}{n^{k-1}}.$$

Denotando a razão do semi-invariante de ordem k para k^{esima} potência do desvio padrão por $\gamma_{k:x} = \frac{\lambda_{k:x}}{(\lambda_{2:x})^{k/2}}$, segue-se que

$$\gamma_{k:\bar{x}} = \frac{\gamma_{k:x}}{n^{\frac{k}{2}-1}}$$

Como casos particulares, temos os coeficientes de

assimetria $\beta_{1:\bar{x}} = \frac{\beta_{1:x}}{n}$,

kurtosis $\beta_{2:\bar{x}} = 3 + \frac{\beta_{2:x} - 3}{n}$.

As fórmulas acima mostram que, embora o universo se afaste do tipo normal, a distribuição de médias aproxima-se muito mais dêsse tipo, pois os índices característicos vêm divididos por n. Se o tamanho da amostra cresce indefinidamente, $n \to \infty$, também $\gamma_{k:\bar{x}} \to 0$ para todos os valores de k maiores que 2, contando que as razões correspondentes no universo $\gamma_{k:x}$ sejam finitas. Ora, essa propriedade caracteriza a curva normal (§ 2.4), donde se conclui que, se a amostra fôr suficientemente grande, a distribuição de médias tende para a normalidade, qualquer que seja a forma do universo donde provenham.

2.6 *Distribuições provenientes de universos normais.*

No caso de ser dada a especificação do universo, podemos determinar a forma analítica da distribuição por amostragem de médias. Um caso especial, de grande relevância, é o do universo normal.

Consideremos as variáveis independentes x_i $(i = 1, 2, \ldots n)$, obedecendo à lei normal com dispersão σ_i,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_i^2}{2\sigma_i^2}}$$

cuja função característica é (§ 2.4) $\Psi(t) = e^{-\frac{\sigma_i^2 t^2}{2}}$

Se tomarmos a soma $w = \sum_{i=1}^{n} c_i x_i$, a sua função característica será $e^{-\frac{c_i^2 \sigma_i^2 t^2}{2}}$, donde se conclui que w também segue a lei normal, com variância definida por $\sigma_w^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2$.

A essa conclusão poder-se-ia, aliás, chegar imediatamente, observando que, pela aditividade dos semi-invariantes, a distribuição de w tem também todos os semi-invariantes nulos, exceto os de ordem 1 e 2, resultando que também ela é normal.

É esta a chamada *propriedade reprodutiva* da lei normal: a soma de variáveis aleatórias independentes obedecendo a esta lei, também a obedece. Por isso, diz-se que a lei normal é *estável.*

Suponhamos agora que os x_i constituem uma amostra obtida de um universo de variância σ_x^2, isto é, $\sigma_i^2 = \sigma_x^2$ para cada valor i, e $c_i = \frac{1}{n}$. Então $w = \bar{x}$, e, de acôrdo com a propriedade acima $\sigma_{\bar{x}}^2 = \frac{\sigma_x}{n}$. Donde se conclui que a distribuição por amostragem de médias de um universo normal também é normal, com a mesma média do universo e variância igual à do universo dividida pelo tamanho da amostra.

2.7 *Amostragem em universos finitos.*

Nem sempre as condições de amostragem correspondem às que definimos como *simples* (§ 1.4); ao invés, na prática ela se realiza geralmente sem reposição dos elementos extraídos, e tem-se a amostragem em um *universo finito,* onde a extração de um elemento de dada categoria modifica a probabilidade da extração dos demais.

Seja N o número de elementos do universo e n o da amostra. A fórmula *(2.1)* ainda prevalece, e temos $E(\bar{x}) = \tilde{x}$. Para calcular a variância da distribuição, notemos que ela independe da origem de coordenadas, pelo que podemos supôr que esta coincide com a média do universo. Então a variância da soma $w = x_1 + x_2 + \ldots + x_n$ será

$$E(w^2) = \sum_{i=1}^{n} E\left(x_i^2\right) + \sum_{i \neq j} E(x_i x_j)$$

$$= n\sigma_x^2 + n(n-1)E(x_i x_j),$$

pois cada um dos x_i compõe-se com os restantes $n-1$ valores x_j.

O têrmo $E(x_i x_j)$ obtém-se calculando $E(w^2)$ para todo o universo, que é evidentemente nula. Segue-se que

$$E(w^2) = N\sigma_x^2 + N(N-1)E(x_i x_j) = 0,$$

donde $E(x_i x_j) = \frac{\sigma_x^2}{N-1}$.

Substituindo êsse valor na expressão acima, tem-se

$$\sigma_w^2 = n\sigma_x^2\left(1 - \frac{n-1}{N-1}\right) = n\sigma_x^2 \frac{N-n}{N-1}.$$

Para obter a variância da distribuição de médias, façamos a substituição de x_i para $\frac{x_i}{n}$ virá $\sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n}\frac{N-n}{N-1}$, ou seja

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}. \qquad (2.4)$$

De maneira análoga calcular-se-iam as demais características da distribuição [3]. As fórmulas obtidas são mais gerais e se reduzem às do § 2.5 para $N \to \infty$. As conclusões de CHURCH são que a distribuição por amostragem de médias de qualquer universo finito é pràticamente normal.

Torna-se interessante comparar as vantagens de um e outro processo de amostragem. Não havendo reposição, a variância dos erros de amostragem tende para zero, quando o número de elementos da amostra aproxima-se do tamanho do universo, o que só se dá no outro processo para $n \to \infty$. Também a variância é sempre ligeiramente menor não havendo reposição, mesmo quando a amostra constitui apenas uma diminuta parte do universo. Há, pois, certas vantagens no uso da amostragem sem reposição.

Contudo, as fórmulas correntes de erros padrões baseiam-se, por comodidade de derivação analítica, em amostragem num universo infinito. Invoca-se a favor de seu uso, além de sua simplicidade, o fato de que, quando o universo é muito grande relativamente à amostra, as diferenças entre os dois tipos de fórmulas são desprezíveis, e as interpretações baseadas sôbre as primeiras pecam a favor de uma maior segurança.

2.8 Universos não-normais. Verificações experimentais.

As distribuições por amostragem de médias provenientes de universos não-normais têm sido determinadas analìticamente para vários tipos de especificação.[4] Assim, as derivadas de um universo assimétrico do tipo III de PEARSON foram estudadas separadamente por CHURCH, IRWIN e CRAIG. O resultado é outra distribuição do mesmo tipo, que se normaliza mesmo para $n < 50$. IRWIN ocupou-se de distribuições oriundas de universos do tipo II de PEARSON, entre cujos casos particulares se inclui o universo retangular, também tratado por RIETZ e HALL. Outros tipos de universo foram estudados por BAKER, CRAIG e outros autores.

As conclusões são uniformes em evidenciar que, para os mais variados tipos de universo, as distribuições de médias aproximam-se do tipo normal, mesmo para baixos valores de n.

Essas conclusões também têm sido confirmadas por várias verificações experimentais. Assim, SHEWHART [5] fez 4 000 tiragens com reposição de dois universos, um retangular e outro triangular, e registrou as distribuições de médias em amostras de 4 elementos. Os resultados experimentais situam-se pràticamente sôbre uma curva normal teórica, calculada mediante as características conhecidas do universo.

---

[3] Cf. CHURCH, A. E. R., "On the means and squared standard-deviations of small samples from any population", *Biometrika*, vol. 18 (1926), pág. 321.

[4] Cf. RIDER, P. R., "A Survey of the Theory of Small Samples", *Annals of Mathematics*, vol. 31, 1930; RIETZ, H. L., "Some Topics in Sampling Theory", *Bull. Amer. Mathem. Soc.*, vol. 43, 1937.

[5] SHEWHART, W. A., *Economic Control of the Quality of Manufactured Product* (Nova Iorque, 1931), pág. 181.

CARVER [6] apresenta o resultado da distribuição de médias de 1 000 amostras de 25 elementos, extraídas sem recolocação de um universo arbitrário, extremamente assimétrico, constatando-se a normalidade da distribuição. Dêsse e de outros resultados experimentais, CARVER conclui que, se a amostra se compõe de 50 ou mais elementos, e o universo é no mínimo 10 vêzes maior que a amostra, a forma do universo tem relativamente pouca influência sôbre a forma da distribuição de médias.

2.9 A significância da média. O teste u.

Tôdas as conclusões relatadas são no sentido de que, mesmo quando a população se afasta do tipo normal, a distribuição por amostragem de médias é aproximadamente dêsse tipo, com média $\bar{x}$ e variância $\frac{\sigma_x}{\sqrt{n}}$. podemos, pois, utilizar a distribuição normal para determinar a significância da média, isto é, para verificar a hipótese de que ela difira significantemente de certo valor prefixado para a média da população.

Fazendo a transformação $u = \frac{\bar{x} - \tilde{x}}{\sigma_{\bar{x}}}$, u distribuir-se-á normalmente (ou aproximadamente) em tôrno de zero e com desvio padrão unitário, e podemos utilizar a tabela das áreas da curva normal

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

para calcular a probabilidade de se obter uma amostra aleatória, na qual $\bar{x}$ difira de $\tilde{x}$ tanto ou mais que o desvio reduzido $|\delta|$ calculado sôbre a amostra. Essa probabilidade será, pela simetria da curva normal,

$$P_{\delta} = 2\int_{\delta}^{\infty} \phi(u)du = 1 - 2\int_{0}^{\delta} \phi(u)\,du.$$

É êste o chamado *teste u*, a ser empregado quando conhecemos a variância do universo, ou quando lidamos com grandes amostras, de modo que os erros de amostragem na estimativa da variância sejam desprezíveis. Em geral, as tabelas existentes nos manuais dão apenas a segunda integral; as de FISHER e YATES [7] fornecem diretamente P

Exemplo A. A densidade do ouro é 19.3 gr/cm³, e admite-se que o processo de sua determinação está sujeito a um êrro padrão de 0.12. Uma amostra de metal, que se presume seja ouro, é examinada mediante 4 determinações de densidade, obtendo-se os resultados 19.5, 19.2, 18.8 e 18.7, cuja média é 19.05. Pode-se concluir que se trata realmente de ouro?

O êrro padrão de uma amostra de 4 elementos é $\sigma_{\bar{x}} = \frac{0.12}{\sqrt{4}} = 0.06$; e como a diferença entre o valor determinado e o teórico é 0.25, segue-se que

---

[6] CARVER, H. C., "Fundamentals of the Theory of Sampling", *Annals of Mathematical Statistics*, vol. 1 (1930), pág. 111.

[7] FISHER, R. A., e YATES, F., *Statistical Tables for Biological, Agricultural and Medical Research* (Edinburgo, 1938).

$\delta = \frac{0.25}{0.06} = 4.17$. A tabela de áreas da curva normal mostra que a probabilidade de um desvio numèricamente maior que êsse é $P\delta = 0.00005$, donde se conclui que a amostra não é de ouro puro.

Exemplo B. A idade média, ao morrer, de uma população, é de 41.5 anos, com desvio padrão $\sigma = 10$. Uma companhia de seguros quer determinar essa característica num grupo de assegurados com um êrro máximo de 2% e segurança de 95%. De que tamanho deve ser a amostra?

O êrro admitido é nesse caso $0.02\,\bar{x} = 0.83$ anos; portanto

$$\delta = \frac{0.83}{\sigma_{\bar{x}}} = 0.083\sqrt{n}.$$

De outro lado, a probabilidade de 95% corresponde a

$$\int_0^{\delta} \phi(u)\,du = \frac{0.95}{2} = 0.475,$$

e a interpolação inversa na tabela da integral da curva normal fornece $|\delta| = 1.96$. Igualando os dois valores, temos

$$n = \left(\frac{1.96}{0.083}\right)^2 = 557.$$

Uma solução rápida poder-se-ia obter utilizando nomogramas.[8] Calculado o coeficiente de variação $CV = \frac{10}{41.5} = 0.241$ tem-se imediatamente, ligando êsse ponto a $P = 0.02$, que $n = 560$ aproximadamente.

2.10 *Limites fiduciais da média do universo.*

Freqüentemente não nos é possível ajuizar *a priori* nenhum valor para a média da população, e o que queremos é precisamente inferir êsse valor a partir da amostra que possuimos. Trata-se, então, de estabelecer limites entre os quais podemos, com certa confiança, presumir que se enquadre essa característica da população.

Suponhamos que se deseja alcançar uma precisão de 95%. Se considerarmos a probabilidade P_δ de um desvio maior que $|\delta|$ e fizermo-la igual a 0.05, uma interpolação inversa na tabela das áreas da curva normal nos dará o valor do desvio correspondente $|u| = \delta$, ou seja $\frac{|\bar{x} - \tilde{x}|}{\sigma_{\bar{x}}} = \delta$. Daí $\tilde{x} = \bar{x} \pm \delta\sigma_{\bar{x}}$.

Os dois valores $\tilde{x}_1 = \bar{x} - \delta\sigma_{\bar{x}}$ e $\tilde{x}_2 = \bar{x} + \delta\sigma_{\bar{x}}$ foram denominados por FISHER[9] *limites fiduciais* correspondentes à probabilidade P_δ, e a proporção de inferências corretas *probabilidade fiducial* (no caso vertente 95%) correspondente à amostra observada. NEYMAN[10] usa as denominações correspondentes de *limites* e *coeficiente de confidência.*

Temos, com efeito, que

$$\bar{x} - \delta\sigma_{\bar{x}} \leq \tilde{x} \leq \bar{x} + \delta\sigma_{\bar{x}}$$

equivale a

$$\tilde{x} - \delta\sigma_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \delta\sigma_{\bar{x}}.$$

---

[8] KINGSTON, J., "Dimensionamento de amostras", *Revista Brasileira de Estatística*, num. 19 (1944), pág. 303.

[9] FISHER, R. A., "Inverse Probability", *Proc. Cambridge Phil. Soc.*, vol. 26 (1930), pág. 528.

[10] NEYMAN, J., "On two different aspects of the representative method", *Jour. Roy. Stat. Soc.*, vol. 97 (1934), pág. 558.

Ora, a probabilidade de que $\bar{x}$ satisfaça essa última desigualdade é 0.95, de modo que, em média, acertaremos em nosso julgamento relativo ao valor dêsse parâmetro em 95% dos casos.

Exemplo. Em face das determinações físicas constantes do Ex. 2.9, entre que limites se deve presumir, com uma segurança de 95%, que se enquadre a verdadeira densidade do metal examinado?

A probabilidade de 0.95 corresponde na curva normal a um desvio de $|\delta| = 1.96$; portanto $\delta\,\sigma_{\bar{x}} = 1.96 \times 0.06 = 0.118$. Os limites fiduciais da média são pois $\tilde{x}_1 = 19.05 - 0.118 = 18.92$ e $\tilde{x}_2 = 19.05 + 0.118 = 19.17$.

Òbviamente, podemos determinar limites correspondentes a outras probabilidades fiduciais, por exemplo para nível de 99% e para outras constantes estatísticas, que possuam distribuições por amostragem contínuas e conhecidas. De um modo geral, definiremos os limites fiduciais do seguinte modo: se usamos uma estatística t para estimar uma grandeza desconhecida θ, denominamos t_1 e t_2 $(t_1 < \theta < t_2)$ os limites fiduciais de θ relativamente às observações dadas e à probabilidade P se, na hipótese de que $\theta = t_1$, a probabilidade de que a estatística calculada do mesmo modo que t e sôbre uma amostra semelhante exceda t é $\frac{1}{2}P$, e se a probabilidade, na hipótese de $\theta = t_2$, de que esta estatística seja menor que t é também $\frac{1}{2}P$.

O conceito de probabilidade fiducial não se deve confundir com conclusões baseadas no teorema de BAYES, e que resultam em afirmações do tipo: "dado o valor $\bar{x}$ da amostra, a probabilidade do valor $\tilde{x}$ do universo estar compreendido entre t_1 e t_2 é 0.95". Isso implica a concepção de uma super-população de valores correspondentes ao valor $\bar{x}$, e dos quais 95% estariam compreendidos entre os limites. Diversas dificuldades adviriam dessa situação, entre as quais se destaca a falta de elementos em que basear as características da distribuição dessa super-população. Ao contrário, fundamentados na probabilidade fiducial, o que afirmamos é que, para todos os possíveis valores $\bar{x}$ de amostras extraídas de quaisquer populações, 95% dos valores $\tilde{x}$ dessas populações enquadram-se entre os limites fiduciais, o que não envolve nenhuma hipótese relativa à distribuição dos parâmetros $\tilde{x}$.

2.11 *Diferenças e funções lineares de estatísticas.*

Na dedução do § 2.2, consideremos os x_i como *estatísticas* t_i, cujas variâncias σ_{t_i} são conhecidas; então, a variância da distribuição por amostragem de uma função linear qualquer das mesmas será dada por

$$\sigma_w^2 = \sum_{i=1}^{n} c_i^2 \sigma_{t_i}^2 + \sum_{i \neq j} c_i c_j \rho_{ij} \sigma_{t_i} \sigma_{t_j}. \qquad (2.5)$$

Se as estatísticas provêm de amostras independentes, $\rho_{ij} = 0$, e a fórmula simplifica-se para

$$\sigma_w^2 = \sum_{i=1}^{n} c_i^2 \sigma_{t_i}^2. \qquad (2.6)$$

Um caso especial, de grande utilidade prática, é o atinente à variância da diferença entre médias de duas amostras $\bar{x}_1$ e $\bar{x}_2$. Fazendo $c_1 = 1$, $c_2 = -1$, temos

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \frac{\sigma^2_{x_1}}{n_1} + \frac{\sigma^2_{x_2}}{n_2}, \qquad (2.7)$$

sendo n_1 e n_2 os tamanhos respectivos das amostras. Se, em particular, as amostras provêm do mesmo universo,

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_x\left(\frac{1}{n_1} + \frac{1}{n_2}\right). \qquad (2.8)$$

Essas expressões permitem verificar a hipótese de que duas médias diferem apenas por flutuações de amostragem. Suposto que as médias se originem de universos com a mesma média, a diferença entre elas terá uma distribuição, cuja média é $E(\bar{x}_1 - \bar{x}_2) = 0$, e cuja dispersão é dada pelas fórmulas acima. Como a distribuição de médias é aproximadamente normal, a distribuição de diferenças entre elas também será (§ 2.6) dêsse tipo, e, portanto, a probabilidade de que a diferença $\frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x}_1-\bar{x}_2}} = \delta$ seja excedida numèricamente é dada por

$$P_\delta = 2\int_\delta^\infty \phi(u)\,du.$$

Suponhamos agora ambas as amostras do mesmo tamanho. Adotando o nível de significância correspondente a 2σ, deduz-se a regra prática de que a diferença entre médias é significante se fôr maior que 3 vêzes o êrro padrão de uma das médias, pois então

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sqrt{2}\,\sigma^2_{\bar{x}} \text{ e } 2\sqrt{2}$$

é aproximadamente igual a 3.

Exemplo. Suponhamos que em uma segunda amostra, que se presume ser do mesmo metal referido no Ex. 2.9 A, fizeram-se 6 determinações de densidade, obtendo-se os resultados 19.1, 19.1, 19.0, 18.8, 18.7 e 18.6, cuja média é 18.88. Confirma-se a expectativa de que se trata realmente do mesmo metal?

Temos

$$\sigma_{\bar{x}_1-\bar{x}_2} = 0.12\sqrt{\frac{1}{4} + \frac{1}{6}} = 0.12 \times 0.646 = 0.078,$$

$$\delta = \frac{19.05 - 18.88}{0.078} = 2.18,$$

a que corresponde a probabilidade $P\delta = 0.029$. A diferença é, pois, significante, e mostra que, no caso, se trata de metais diferentes.

Se a média da segunda amostra, 18.85, tivesse resultado também de 4 determinações, poderíamos empregar a regra prática mencionada. Teríamos que a diferença 0.17 é menor que 3×0.06, e portanto não significante, confirmando-se a hipótese formulada.

2.12 *Erros padrões de outras estatísticas.*

Determinemos os erros padrões de outras estatísticas de uso freqüente. Temos assim:

A) Separatrizes.[11]

[11] Para um estudo geral da distribuição por amostragem das separatrizes, vêr GUMBEL, E. J., "Les valeurs extrêmes des distributions statistiques", *Annales de l'Institut Henri Poincaré*, vol. 4 (Paris, 1935), pág. 115.

Consideremos a separatriz S, de grau p, de uma amostra de tamanho n da variável x, que obedece a uma lei de distribuição contínua $f(x)$. Seja $\tilde{S}$ a separatriz de mesmo grau dessa distribuição ideal. Em vista da definição, a probabilidade de que um qualquer valor x escolhido ao acaso seja menor que $\tilde{S}$ é p. A separatriz S da amostra situar-se-á a uma distância $\delta x = \delta S$ de $\tilde{S}$, e o número de elementos da amostra acima de $\tilde{S}$ é $np + d$. Se y corresponde ao valor médio de x no intervalo dS, então, a menos de infinitésimos de ordem superior, temos $y\,\delta S = d$.

E' esta a equação que liga a variação dS do valor da separatriz da amostra ao desvio por amostragem d da freqüência acima de S. Daí

$$dS = \frac{d}{y} \quad \text{e} \quad \sigma_s^2 = \frac{1}{y^2}\,\sigma_d^2$$

Ora, o êrro padrão da freqüência de uma categoria definida pela proporção p é (§ 3.2) $\sigma_d = \sqrt{n\,p\,q}$; por conseguinte $\sigma_s = \sqrt{n\,p\,q}/y$.

Se a variável x se distribui normalmente, o valor de y é dado por $y = n\,[\phi(u)]_p/\sigma$ onde $[\phi(u)]_p$ é a ordenada que corresponde à divisão da área da curva normal na proporção p. Logo

$$\sigma_s = \frac{\sigma}{[\phi(u)]_p}\sqrt{\frac{pq}{n}}.$$

Para a mediana, $p=1/2$ e $\phi(u)=0.3989$; donde $\sigma_{Md}=1.253\,\frac{\sigma}{\sqrt{n}}$. Logo, a mediana tem um êrro padrão 1.253 vêzes maior que a média aritmética baseada na mesma amostra, conclusão, contudo, só estritamente válida para distribuições normais.

Para os quartis, $p = \frac{1}{4}$ (ou $\frac{3}{4}$) e $\phi\,(u) = 0.3178$; donde $\sigma_{Q1\ ou\ Q3} = 1.3626\,\frac{\sigma}{\sqrt{n}}$. Como na curva normal o desvio quartílico Q equivale a $0.6745\,\sigma$, basta multiplicar ambos os têrmos das fórmulas acima por êsse fator para termos os êrros prováveis. Assim

$$E.P._{Md} = 1.253\,\frac{Q}{\sqrt{n}}$$

$$E.P._{Q1\ ou\ Q3} = 1.3626\,\frac{Q}{\sqrt{n}}$$

B) Desvio Padrão.

O êrro padrão do desvio padrão s, quando essa estatística se baseia no momento de segunda ordem da distribuição, é aproximadamente dada por

$$\sigma_s = \frac{\sigma}{\sqrt{2n}} \qquad (3.9)$$

A distribuição não é normal, mas aproxima-se dêsse tipo com o crescer da amostra, como veremos adiante. Por exemplo, para $n=100$, já a distribuição tem as características $\beta_1=0.0051$ e $\beta_2=3.0000$.

Se a dispersão é medida pelo afastamento médio, temos que, para grandes amostras,

$$\sigma_{AM} = 1.068 \frac{\delta}{\sqrt{2n}} = 0.861 \frac{\sigma}{\sqrt{2n}}$$

sendo δ o valor do afastamento médio na população. Se, por conseguinte, estima-se σ a partir do afastamento médio, então o seu êrro padrão será $1.253\,\sigma_{AM} = 1.068 \frac{\sigma}{\sqrt{2n}}$, denotando a inconveniência dêsse processo de estimação.

2.13 Êrro padrão de uma função.

Importa também determinar o êrro padrão de uma função qualquer. Já cuidamos dêsse problema, referentemente às funções lineares, e vamos agora generalizá-lo, supondo porém que se trata de grandes amostras, estatísticamente independentes.

Seja $\lambda = f(\theta_1, \theta_2, \ldots \theta_m)$ a função ligando os valores dos parâmetros no universo, e sejam $l, t_1, t_2, \ldots t_m$ os valores correspondentes numa amostra aleatória. Então, os erros cometidos na determinação dessas constantes serão $l = \lambda + \delta\lambda$, $t_1 = \theta_1 + \delta\theta_1$, $t_2 = \theta_2 + \delta\theta_2, \ldots$, onde δ significa um desvio muito pequeno. Teremos $l = f(\theta_1 + \delta\theta_1, \theta_2 + \delta\theta_2, \ldots \theta_m + \delta\theta_m)$, e, se os erros forem tão pequenos que os seus produtos e potências possam ser desprezados, poderemos desenvolver essa expressão em série de TAYLOR, obtendo, aproximadamente,

$$\delta\lambda = \frac{\partial f}{\partial\theta_1}\delta\theta_1 + \frac{\partial f}{\partial\theta_2}\delta\theta_2 + \ldots + \frac{\partial f}{\partial\theta_m}\delta\theta_m.$$

O grau de aproximação implícita corresponde a desprezar quantidades da ordem de $1/n$ em comparação com a unidade, sendo n o número de elementos na amostra.

Tomando os valores médios de $\delta\lambda$, $\delta\theta_1$, $\delta\theta_2, \ldots$ para tôdas as amostras que se possam extrair do universo, teremos os quadrados dos correspondentes erros padrões $\sigma_l^2, \sigma_{t_1}^2, \sigma_{t_2}^2, \ldots\ldots$ Ora, observando que $\delta\lambda$ é uma função linear de $\delta\theta_1, \delta\theta_2, \ldots$, podemos aplicar a fórmula (2.6), obtendo

$$\sigma_l^2 = \left(\frac{\partial f}{\partial\theta_1}\right)^2 \sigma_{t_1}^2 = \left(\frac{\partial f}{\partial\theta_2}\right)^2 \sigma_{t_2}^2 + \ldots + \left(\frac{\partial f}{\partial\theta_m}\right)^2 \sigma_{t_m}^2. \qquad (2.10)$$

Nessa expressão, as derivadas parciais referem-se aos verdadeiros valores no universo. Apliquemo-la à determinação de algumas funções importantes.

A) Variância.

Temos $v = s^2$, $\theta_1 = \sigma$, $\frac{\partial f}{\partial\theta_1} = 2\sigma$. Logo $\sigma_v^2 = 4\sigma^2(\sigma_s^2)^2 = \frac{2\sigma^4}{n}$

e finalmente $\sigma_s = \sigma^2\sqrt{\frac{2}{n}}$.

B) Coeficiente de variação.

Como a média e o desvio padrão, estimados a partir da mesma amostra, são independentes, podemos aplicar o processo. Temos $\kappa = \frac{\sigma}{\mu}$ e portanto $\sigma_\kappa^2 = \frac{1}{\mu^2}\frac{\sigma^2}{2n} = \frac{\sigma^2}{\mu^4}\frac{\sigma^2}{n}$ ou seja $\sigma_k = \kappa\sqrt{\frac{2\kappa^2+1}{2n}}$.

Relativamente aos erros padrões que vimos de calcular, dois fatos devem ser notados. Primeiro, as fórmulas envolvem os valores dos parâmetros do universo, que, em geral, são desconhecidos. Sòmente para grandes amostras podemos substituir aos mesmos, com aproximação razoável, os valores deduzidos das amostras.

Em segundo lugar, a interpretação dêsses erros padrões, mediante a estimativa da probabilidade de se verificarem desvios maiores, depende da forma da distribuição, que nem sempre é conhecida. Muitas dessas distribuições, com o aumento do tamanho da amostra, tendem para o tipo normal, justificando-se assim o largo emprêgo dessa função na determinação da significância das constantes estatísticas.

2.14 Testes da normalidade do universo.

Para muitas estatísticas, a forma analítica de sua distribuição tem sido determinada no pressuposto de que as amostras provenham de um universo normal. Daí o interêsse em se poder discriminar até quando uma dada população, à vista da amostra que dela se tem, pode ser considerada normal ou não.

As duas constantes usualmente empregadas para caracterizar a forma das distribuições são:

para a assimetria $\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}$,

para a kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

Dada uma amostra, calculamos as estatísticas $\sqrt{b_1}$ e b_2, e devemos avaliar se elas desviam significantemente em relação aos correspondentes parâmetros da curva normal, que são $\sqrt{\beta_1} = 0$ e $\beta_2 = 3$. As distribuições dessas duas estatísticas não são ainda exatamente conhecidas, mas FISHER [12] calculou os seus quatro primeiros momentos, aos quais E. S. PEARSON ajustou curvas empíricas do sistema pearsoniano, e posteriormente GEARY e PEARSON [13] calcularam tabelas dando os valores correspondentes aos níveis de significância de 5% e 1%.

Para grandes amostras, $\sqrt{b_1}$ e b_2 aproximam-se da normalidade, com erros padrões respectivamente de $\sqrt{6/n}$ e $\sqrt{24/n}$; mas, para valores de $n < 1000$, devem-se preferir as tabelas de GEARY-PEARSON, sobretudo para b_2, cuja distribuição é fortemente assimétrica. Para $\sqrt{b_1}$, podemos considerar que o duplo do êrro padrão cai pràticamente no nível de significância de 5%.

---

[12] FISHER, R. A., "The moments of the distribution for normal samples of measures of departure from normality", *Proc. Roy. Soc. of London*, ser. A, vol. 130 (1930), pág. 16.

[13] GEARY, R. C., e PEARSON, E. S., *Tests of Normality* (Londres, Biometrika Office, 1938).

A assimetria também se pode estimar mediante o afastamento médio δ; tem, então, por medida

$$\alpha = \frac{\delta}{\sqrt{\mu_2}} = \frac{\Sigma|x - \bar{x}|}{[n\Sigma(x - \bar{x})^2]^{1/2}}$$

Para a curva normal, $\alpha = \sqrt{2/\pi} = 0.798$. As tabelas da distribuição desta estatística utilizam-se da mesma maneira que as de $\sqrt{b_1}$. Para pequenas amostras, é preferível o emprêgo de α.

Exemplo A. CROXTON e COWDEN (*Applied General Statistics*, pág. 273) dão a distribuição das distâncias de jogada em "baseball" de um grupo de 303 alunas de colégio, caracterizada por $\sqrt{b_1} = 0.102$, $b_2 = 2.77$. Pode-se considerar normal essa população ?

As tabelas de GEARY-PEARSON dão, para $n = 300$, os desvios correspondentes à probabilidade de 5% como $\sqrt{b_1} = 0.230$, $b_2 = 2.59$; logo justifica-se, no caso, a assimilação à curva normal.

Exemplo B. Suponhamos que um grupo de 300 alunas tivesse fornecido exatamente os valores $\sqrt{b_1} = 0.230$ e $b_2 = 2.59$, correspondentes ao nível de 5%. Confrontemos êsse resultado com o que se obtém interpretando os erros padrões dos coeficientes à luz da tabela da curva normal. Temos

$$\sigma_{\sqrt{b_1}} = \sqrt{\frac{6}{300}} = 0.141 \ , \ \sigma_{b_2} = \sqrt{\frac{24}{300}} = 0.283$$

No primeiro caso, o desvio reduzido é $\delta = \frac{0.230}{0.141} = 1.63$, a que corresponde a probabilidade de um desvio no sentido positivo $P\,(>\delta) = 0.052$. No segundo, temos $\delta = \frac{0.41}{0.283} = 1.45$ donde $P\,(>\delta) = 0.074$. Por conseguinte, em ambos os casos a aproximação normal sobrestima as probabilidades dos desvios considerados. Sobretudo para b_2 é preciso cautela na interpretação, pois, para um valor um pouco inferior a 2.59, a tabela de GEARY-PEARSON daria o desvio como significantemente diferente de 3, ao passo que a aproximação normal o consideraria como não significante, por ser a área à sua esquerda maior que 5%.

Se a curva fôsse leptokurtica, com $b_2 = 3.47$, correspondente ao nível de 5%, a aproximação normal daria $\delta = \frac{0.47}{0.283} = 1.66$ e $P\,(>\delta) = 0.049$, subestimando ligeiramente a probabilidade do desvio.

2.15 *O critério de* TCHEBYCHEFF.

O critério de TCHEBYCHEFF permite ajuizar sôbre a grandeza do êrro padrão, independentemente do conhecimento da natureza da distribuição da estatística em causa. Êle estabelece que: para uma distribuição qualquer $f(x)$, a probabilidade de um desvio relativo à média maior que um múltiplo $|\delta|$ do desvio padrão é, no máximo, igual a $\frac{1}{\delta^2}$.

Com efeito, essa probabilidade é

$$P_\delta = \int_{-\infty}^{-\delta\sigma} f(x)\,dx + \int_{\delta\sigma}^{\infty} f(x)\,dx$$

$$\leq \int_{-\infty}^{-\delta\sigma} \frac{x^2}{\delta^2\sigma^2} f(x)\,dx + \int_{\delta\sigma}^{\infty} \frac{x^2}{\delta^2\sigma^2} f(x)\,dx \leq \int_{-\infty}^{\infty} \frac{x^2}{\delta^2\sigma^2} f(x)\,dx,$$

pois que $\frac{x^2}{\delta^2\sigma^2} \geq 1$ no intervalo abrangido pela integração, e $f(x)$ é essencialmente positiva. Notando que a última integral iguala

$$\frac{1}{\delta^2\sigma^2}\int_{-\infty}^{-\infty} x^2 f(x)\,dx = \frac{1}{\delta^2},$$

segue-se que $P_\delta \leq \frac{1}{\delta^2}$. (2.11)

Êsse resultado também é conhecido sob a denominação de *desigualdade* ou *teorema de* TCHEBYCHEFF; na realidade, êle fôra antecipado por BIENAYMÉ.

A importância do critério de TCHEBYCHEFF deflui de ser válido para qualquer tipo de distribuição; mas, devido a essa extrema generalidade, êle pouco pode informar a respeito da distribuição particular com que lidamos. Isto é, os limites decorrentes do critério são tão afastados, que perdem o valor prático na interpretação das estatísticas.

2.16 Generalizações do critério de TCHEBYCHEFF.

Diversas generalizações têm sido propostas para êsse critério. Assim, PEARSON mostrou que, para distribuições contínuas,

$$P_\delta \leq \frac{\mu_{2k}}{\delta^{2k}\sigma^{2k}}$$

desigualdade, porém, que ainda não conduz a resultados práticos.

Outra generalização é devida a CAMP,[14] que, para distribuições unimodais, com a moda a menos de σ da média (assimetria $\leqslant 1$), estabelece que

$$P_\delta < \frac{\alpha_{2k}}{\left(\frac{\delta + 2k\delta}{2k}\right)^{2k}}$$

mais um resto, geralmente desprezível. Aí, α_{2k} representa o momento reduzido $\frac{\mu_{2k}}{\sigma^{2k}}$. Para $k = 1$, temos

$$P_\delta < \frac{1}{2.25\,\sigma^2}, \qquad (2.12)$$

também conhecida como *desigualdade de* CAMP-MEIDELL.

Comparemos os resultados que se obtém pela aplicação dessas desigualdades e da integral da curva normal. Se na de CAMP-MEIDELL fizermos $\delta = 0.667$ e na de TCHEBYCHEFF $\delta = 1$, resulta $P_\delta = 1$; isto é, a primeira não nos fornece nenhuma informação sôbre desvios compreendidos nos limites positivos e negativos de um êrro provável, e a de TCHEBYCHEFF nos limites de um êrro padrão.

O nível de significância de 5% corresponde, na curva normal, a um desvio $|\delta| = 1.96$ vêzes o desvio padrão; para a desigualdade de CAMP-MEIDELL êle corresponde a $2.98\,\sigma$ e para a de TCHEBYCHEFF $4.47\,\sigma$. Para o nível de significância de 1%, os valores correspondentes são $2.58\,\sigma$, $6.67\,\sigma$ e $10\,\sigma$.

---

[14] CAMP, B. H., "A new generalization of Tchebycheff's statistical inequality", *Bull. Amer. Mathem. Soc.*, vol. 28 (1922), pág. 427.

Os limites fornecidos por essas desigualdades podem, ainda, ser melhorados. CAMP mostrou que, se a distribuição podia ser assimilada a uma curva de GRAM-CHARLIER (até o têrmo de 4.ª ordem) ou a um dos tipos fundamentais de PEARSON, com valores usuais de α_3 e α_4 (isto é, $|\alpha_3| < 1.22$, $2 \leq \alpha_4 \leq 4$), então P_δ era menor que o valor máximo constante da tabela:

δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Max. P_δ	.20	.11	.05	.02	.01	.005	.002	.001

Êsses valores foram escolhidos suficientemente altos para que se possam aplicar indiferentemente aos dois tipos de funções de freqüência.

Exemplo. Na distribuição referida no Ex. 2.14 A, o desvio padrão é $\sigma = 20.95$ pés. Suponhamos que se escolhe ao acaso nesse grupo, 121 alunas; qual é a probabilidade de que o desvio padrão dêsse subgrupo difira de σ mais que 7.5 pés?

Não conhecendo a forma analítica da distribuição de σ, vamos aplicar o critério de TCHEBYCHEFF. Temos

$$\sigma_s = \frac{20.95}{\sqrt{121}} = 1.905\,, \qquad \delta = \frac{7.95}{1.905} = 3.94$$

e portanto $\frac{1}{\delta^2} = 0.064$ e $P_\delta \leq 0.064$.

Contudo, para amostra dêsse tamanho, já a distribuição de σ pode-se assimilar à normal. Ora, a tabela dessa curva nos dá, para êsse desvio,

$$P_\delta = 0.00008.$$

Vê-se que o critério de TCHEBYCHEFF pouco esclarece sôbre a probabilidade dos desvios. Empregando a desigualdade de CAMP-MEIDELL, teriamos

$$P_\delta < \frac{1}{2.25\ (3.94)^2} < 0.029.$$

Recorrendo à tabela de CAMP, obteriamos melhor aproximação, $P_\delta < 0.005$, mas, mesmo assim, longe da decorrente da curva normal.

[15] CAMP, B. H., *The Mathematical Part of Elementary Statistics* (Boston, 1931), pág. 258.

Capítulo III

A DISTRIBUIÇÃO BINOMIAL E SUAS APROXIMAÇÕES

3.1 A distribuição binomial.

Consideremos um universo, cujos elementos estão sujeitos a uma classificação dicotômica, isto é, caracterizam-se pela presença ou ausência de um certo atributo. Seja p a proporção dos elementos portadores dêsse atributo no universo. Em uma amostra de n elementos, um certo número x deles possuirá o atributo em questão, o qual faltará nos n-x restantes, sendo que x pode assumir todos os valores inteiros de 0 a n.

Por analogia com os problemas de esquemas de urnas, costuma-se considerar a escolha de um elemento ao acaso como um acontecimento, que será favorável ou contrário, conforme possua ou não o atributo. A proporção p é a probabilidade do acontecimento favorável numa tiragem.

Se p mantém-se constante durante as tiragens, entre tôdas as possíveis amostras de n elementos, a freqüência relativa daquelas que constam exatamente de x acontecimentos favoráveis será $C_x^n\ p^x\ q^{n-x}$, isto é, o $(x+1)^{esimo}$ têrmo do desenvolvimento de $(q+p)^n$. O desenvolvimento dêsse binômio

$$(q+p)^n = q^n + pq^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} + \ldots$$

dá-nos, pois, a distribuição das amostras possíveis de n elementos colhidos num universo infinito caracterizado por p.

E' essa a distribuição binomial, estudada por Jacques Bernoulli em sua *Ars Conjectandi,* publicada pòstumamente em 1713, e que constitui a primeira distribuição teórica introduzida na estatística. As séries de variáveis descontínuas, cujas classes têm freqüências proporcionais aos têrmos do binômio, também se chamam *bernoullianas.*

A importância do desenvolvimento binomial é considerável em tôda a ciência aplicada, como verificação da independência de alternativas e perequação matemática de dados de observação. Freqüentemente deparamos fenômenos sob formas mùtuamente exclusivas, como a sobrevivência ou a morte, o ataque ou não de uma doença, a eficácia ou não de um tratamento, o sexo num grupo ou numa linhagem, etc., que caem no âmbito dessa distribuição.

Para $p = q = \frac{1}{2}$, a distribuição é simétrica em tôrno de $x = \frac{n}{2}$, pois $C_x^n = C_{n-x}^n$. Para $p \neq q$ ela é assimétrica. Para calcular o valor mais provável de x, escrevamos que a razão do têrmo geral do binômio para o antecedente e o conseqüente deve ser igual ou maior que a unidade; obtém-se as desigualdades

$$\frac{n-x+1}{x}\,\frac{p}{q} \geqq 1 \quad \text{e} \quad \frac{x+1}{n-x}\,\frac{q}{p} \geqq 1.$$

Chamando de x_m o valor modal, devemos ter pois $np-q \leqslant x_m \leqslant np+p$. Se $np-q$ é um número inteiro, $np+p$ também o será, pois a sua diferença é igual à unidade; há, então, duas modas. Caso contrário, haverá sempre um número inteiro entre êsses limites, que será a moda única. Tomando as freqüências relativas, virá $p-\frac{q}{n} \leqslant \frac{x_m}{n} \leqslant p+\frac{p}{n}$.

Para valores crescentes de n, ter-se-á $x_m \to np$, que é uma das formas da conhecida "lei dos grandes números".

3.2 *Momentos da distribuição binomial.*

Para calcular os momentos da distribuição binomial, notemos que, se designarmos por x a realização do acontecimento, x terá o valor 1 quando êle fôr favorável e 0 quando contrário. A função característica dessa grandeza aleatória [1] será òbviamente $pe^t + q$. Repitamos a prova n vêzes; o número de realizações do acontecimento favorável aleatória x, cuja função característica é $\psi(t) = (pe^t + q)^n$.

Consideremos a função característica dos momentos em relação a np como origem; temos

$$\psi(t) = e^{-npt}\left(pe^t + q\right)^n = \left(qe^{-pt} + pe^{qt}\right)^n \qquad (3.1)$$

Desenvolvendo a expressão entre parênteses, segundo as potências de t, vem

$$\psi(t) = \left[1 + pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + pq(1-3pq)\frac{t^4}{4!} + \ldots\right]^n,$$

donde se obtém a função geratriz de semi-invariantes

$$L(t) = n \log_e \left[1 + pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + \ldots\right]$$

$$= n\left[pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + pq(1-6pq)\frac{t^4}{4!} + \ldots\right]$$

Os semi-invariantes são pois $\lambda_1=0$, $\lambda_2=npq$, $\lambda_3=npq\,(q-p)$, $\lambda_4=npq\,(1-6pq)$,... Segue-se que np é o número provável de acontecimentos favoráveis; e os momentos da distribuição em relação à média são

$$\mu_2 = npq,\ \mu_3 = npq(q-p),\ \mu_4 = npq\left[1 + 3(n-2)pq\right],\ \ldots$$

Daí os coeficientes de assimetria e kurtosis, $\sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}$, $\beta_1 = \frac{1-6pq}{\sqrt{npq}} + 3$.

Entre os momentos da distribuição binomial existe uma relação recorrente, que podemos estabelecer derivando a função característica $\psi(t) = \left(qe^{(q-1)t} + (1-q)e^{qt}\right)^n = U^n$ em relação a q. Vem

$$\Sigma \frac{t^k}{k!}\frac{\partial \mu_k}{\partial q} = nU^{n-1}\frac{\partial U}{\partial q} = nU^{n-1}\left[tU - \frac{1}{pq}\frac{\partial U}{\partial t}\right] = ntU^n - \frac{1}{pq}\frac{\partial U^n}{\partial t},$$

---

[1] No caso de variáveis descontínuas, a função característica define-se como $\psi(t) = \sum_i p_i e^{tx_i}$, onde p_i é a probabilidade de se obter o valor x_i.

donde se conclui

$$\mu_{k+1} = pq\left[nk\mu_{k-1} - \frac{\partial \mu_k}{\partial q}\right], \qquad (3.2)$$

fórmula que permite calcular os momentos sucessivos, desde $k \geqslant 1$.

3.3 *Êrro padrão de uma proporção.*

Em muitas investigações estatísticas, os dados apresentam-se sob a forma de proporções de atributos observados. A distribuição dêsses dados, suposta a existência de um sistema causal caracterizado por uma probabilidade constante, será ainda binomial, mas numa escala diferente, pois que a variável é então x/n. Devido a essa mudança de escala, o semi-invariante λ_k vem dividido n^k; e segue-se que

$$\mu_1' = p, \; \sigma = \sqrt{\frac{pq}{n}}.$$

A primeira equação é importante, pois mostra que, em média, as proporções observadas nas amostras são iguais à proporção p do universo; noutros têrmos, p deduzido da amostra é uma estimativa justa.

Note-se que o êrro padrão do número de acontecimentos x, quando p mantém-se constante, cresce com o tamanho da amostra proporcionalmente a $\sqrt{n}$, enquanto que o êrro padrão da proporção decresce proporcionalmente a $\frac{1}{\sqrt{n}}$.

3.4 *Teorema de* Bernoulli.

A segunda equação conduz à imediata demonstração do teorema de Bernoulli. Seja P_δ a probabilidade de que x/n caia fora do intervalo $(p-\epsilon, p+\epsilon)$ sendo $\epsilon > 0$. Considerando ϵ como um múltiplo do desvio padrão das freqüências relativas, $\epsilon = \delta\left(\frac{pq}{n}\right)^{1/2}$ teremos, de conformidade com o critério de Tchebycheff, $P_\delta \leq \frac{1}{\delta^2}$ ou seja, $P_\delta \leq \frac{p(1-p)}{n\epsilon^2}$.

Para um dado ϵ, P_δ pode-se tornar tão pequeno quanto se queira, pelo aumento de n. Noutros têrmos, a probabilidade de ser a diferença $\left(\frac{x}{n}\right) - p$, entre a freqüência relativa de um acontecimento e sua verdadeira probabilidade no universo, menor que qualquer quantidade dada, tende para a certeza quando o número de provas aumenta indefinidamente. E' o conhecido *teorema de* Bernoulli.

3.5 *Verificação da casualidade nas séries empíricas.*

Se os elementos são de duas espécies e acham-se repartidos em grupos, a distribuição binomial permite verificar, numa dada série de observações, se são independentes e se sua distribuição pelos grupos é inteiramente casual.

Notemos que a distribuição binomial só contém dois parâmetros, e, portanto, apenas dois momentos são independentes. Mediante as duas equações $\bar{x} = np$, $s^2 = npq$ podemos ajustar as observações.

Estas nos dão apenas os valores de x e as freqüências correspondentes $\omega(x)$, mas não contém nem p nem o número teórico de observações n. Estes se deduzem de $q = \frac{s^2}{\bar{x}}$, $n = \frac{\bar{x}^2}{\bar{x} - s^2}$.

Para que êsses valores tenham significação, é preciso que $\bar{x} > s^2$, senão adviriam valores negativos de n, e valores p ou q sobrepassando os limites *0* e *1*. O emprêgo da série binomial não seria então legítimo.[2]

Se $\bar{x}$ verifica a condição, escolhe-se para n o maior dos inteiros, satisfazendo a relação acima, e daí resultam os valores teóricos correspondentes aos observados. A aderência do ajustamento seria contrastada pelo teste *chi-quadrado* (cap. V).

Outro método, devido a GULDBERG,[3] consiste no *contrôle da estabilidade normal*. Funda-se em que os valores teóricos obedecem à relação recorrente

$$\frac{\omega(x+1)}{\omega(x)} = \frac{(n-x)}{x+1}\,\frac{p}{q} \quad \text{ou} \quad \frac{(x+1)\,\omega(x+1)}{\omega(x)} + \frac{xp}{q} = \frac{np}{q}.$$

Substituindo p, q e n pelos valores empíricos, conforme as fórmulas dadas, obtêm-se

$$\frac{s^2}{\bar{x}^2}\,\frac{(x+1)\,\omega(x+1)}{\omega(x)} + x\,\frac{\bar{x} - s^2}{\bar{x}^2} = 1, \qquad (3.3)$$

equação que, dependendo apenas dos valores observados, permite verificar a adaptabilidade da distribuição binomial à série empírica.

3.6 Amostragem de atributos.

O desenvolvimento binomial pode ser imediatamente empregado na verificação de hipóteses estatísticas. Suponhamos que numa amostra de n elementos encontram-se x portadores de certo caráter ou atributo. A freqüência relativa x/n diferirá da probabilidade p que caracteriza o universo devido a erros de amostragem. Daí a questão: será o valor de x/n, correspondente à amostra, compatível com o de p na população?

A resposta a êsse quesito exige que avaliemos a probabilidade de que, em n tiragens ao acaso dêsse universo, o número de acontecimentos difira do valor provável $x = np$ tanto ou mais que uma certa grandeza $|d|$. Como o desenvolvimento do binômio $(q+p)^n$ nos dá as probabilidades de tôdas as figurações possíveis, temos que a probabilidade procurada resulta dos somatórios

$$P_d = \sum_{0}^{np-d} C_x^n p^n q^{n-x} = \sum_{np+d}^{n} C_x^n p^n q^{n-x}. \qquad (3.4)$$

Exemplo. Num grupo de igual número de crianças de ambos os sexos, atacadas de difteria, foram observados os seguintes óbitos: homens, 4; mulheres, 10. Pode-se concluir que a mortalidade seja igual para ambos os sexos?

Nessa hipótese, deveriamos observar uma freqüência de óbitos entre homens de $0.5 \times 14 = 7$. A probabilidade de desvios iguais ou maiores que $|d| = 3$ resulta assim

---

[2] Probabilidades negativas ou maiores que a unidade e números de observações negativos foram empregados, com significação puramente formal, por PEARSON (*Biometrika*, vol. IV, 1905) e Miss WHITTAKER (*Biometrika*, vol. 10, 1914).

[3] GULDBERG, A., "On discontinuous frequency-functions and statistical series", *Skandinavisk Aktuarietidskrift*, 1931, pág. 167.

$$P_d = 2 \sum_{0}^{4} C_x^{14} p^{14} = 2\left[1471\,(0.5)^{14} \right] = 0.18$$

pois que aqui $C_x^n = C_{n-x}^n$ e $p^x q^{n-x} = p^n$.

Logo, em mais de 18% dos casos verifica-se, por flutuações de amostragem, discrepâncias iguais ou maiores que a observada, pelo que não se infirma a hipótese de que a mortalidade por difteria independe do sexo.

A medida direta da probabilidade pelo desenvolvimento binomial é, contudo, trabalhosa, se o número de casos é elevado. E as dificuldades resultam, não apenas de ser grande o número de têrmos incluídos no somatório, mas também de serem muito pequenas as probabilidades correspondentes. Quando n é menor que *50*, essas probabilidades podem ser obtidas diretamente das *Tables of the Incomplete Beta Function*,[4] onde a soma dos r primeiros têrmos de $(q+p)^n$ é representado por $I_q\,(n-r+1, r)$. Para grandes valores de n, há vantagens de se estabelecerem métodos rápidos de cálculo.

3.7 Forma limite da distribuição binomial.

Para determinar a forma limite da distribuição binomial para valores crescentes de n, isto é, quando ao polígono substituimos uma curva de freqüência, vamo-nos apoiar no resultado seguinte, devido a TCHEBYCHEFF:

Seja $f(x)$ uma função contínua e integrável entre $-\infty$ e $+\infty$, admitindo derivadas $f'(x)$, $f''(x)$,... gozando das mesmas propriedades. Ponhamos

$$\psi_1(it) = \int_{-\infty}^{\infty} e^{itx} f'(x)dx,\ \psi_2(it) = \int_{-\infty}^{\infty} e^{itx} f''(x)dx,\ \ldots$$

onde $i = -1$. A integração por partes dá

$$\psi_1(it) = \left[f(x) e^{itx} \right]_{-\infty}^{\infty} - it \int_{-\infty}^{\infty} e^{itx} f(x)dx = (-it).\ \psi_0(t)\ ,$$

$$\psi_2(it) = (-it)\psi_1(it) = (-it)^2\, \psi_0(it),\ \ldots$$

Daí se conclui que a função $a_0\, f\,(x) + a_1\, f'\,(x) + a_2\, f''\,(x) + \ldots$ considerada como função de probabilidade, admite a função característica

$$\psi_0(t)\ [a_0 + (it)a_1 + (it)^2 a_2 + (it)^3 a_3 + \ldots].$$

Pôsto isto, consideremos a função característica da distribuição binomial $\psi(t) = (pe^t + q)^n$. Introduzindo a variável z definida por $x = np + z\sqrt{npq} = np + \sigma z$, teremos

$$\psi(t) = e^{-\frac{np}{\sigma}}\, \psi\left(\frac{t}{\sigma}\right) = \left[qe^{-\frac{p}{\sigma}t} + qe^{\frac{q}{\sigma}t} \right]^n_0$$

---

[4] PEARSON, K. (ed.), *Tables of the Incomplete Beta Function* (Londres, Biometrika Office, 1934).

Reportando-nos ao § 3.2, vemos que

$$\psi(t) = e^{n \log\left[1 + \frac{t^2}{2n} + \frac{q-p}{n\sigma}\frac{t^3}{3!} + \frac{1-3pq}{n^2 pq}\frac{t^4}{4!} + \ldots\right]}$$

$$= e^{\left[\frac{t^2}{2} - \frac{p-q}{\sigma}\frac{t^3}{3!} + \frac{1-6pq}{\sigma^2}\frac{t^4}{4!} + \ldots\right]}$$

$$= e^{\frac{t^2}{2}}\left[1 - \frac{p-q}{\sigma}\frac{t^3}{3!} + \frac{1-6pq}{\sigma^2}\frac{t^4}{4!} + \ldots\right]$$

isto é, obtemos um desenvolvimento da forma

$$\psi(t) = \psi_0(t)\,[1 - a_3 t^3 + a_4 t^4 - \ldots] .$$

Ora, $\psi_0(t) = e^{\frac{t^2}{2}}$ é a função característica correspondente à densidade de probabilidade $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$; donde se conclui que $\varphi(t)$ corresponde à função de probabilidade

$$f(z) = \phi(z) + a_3\phi^{III}(z) + a_4\phi^{IV}(z) + \ldots \qquad (3.5)$$

3.8 Aproximação pela função normal.

Se nem p nem q são muito pequenos, os coeficientes $a_3, a_4, \ldots$ tendem para zero quando n cresce indefinidamente. Por conseguinte, para valores elevados de n, podemos substituir ao polígono binomial a função normal

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} .$$

A aplicação à verificação de hipóteses é imediata, pois que então a probabilidade que o número de acontecimentos favoráveis x em n provas difira do valor provável np tanto ou mais que um desvio $|d|$ é dada por

$$P_\delta = 2\int_\delta^\infty \phi(z)dz = 1 - 2\int_0^\delta \phi(z)dz ,$$

onde $\delta = \frac{d - 1/2}{\sigma}$. A razão de termos corrigido d de ½ unidade resulta de que as freqüências binomiais correspondem aos pontos médios $x = 0, 1, 2, \ldots$ das classes do histograma ao qual se ajusta a curva normal, e a integração estende-se até os limites das classes. Se se deseja apenas a probabilidade de desvios maiores que $|d|$, deve-se tomar $\delta = \frac{d + 1/2}{\sigma}$.

Exemplo. O coeficiente de letalidade de certa doença é 1/3; em 450 casos de incidência da mesma, qual é a probabilidade de se verificarem no mínimo 131 e no máximo 179 óbitos?

Temos $np = 150$, $\sigma = 10$, e os desvios reduzidos $\delta_1 = {}^1/_{10}(131 - 150 - 0.5) = -1.95$, $\delta_2 = {}^1/_{10}(179 - 150 + 0.5) = 2.95$. Admitindo a aproximação binomial, segue-se que

$$Q_\delta = \int_{-1.95}^{2.95} \phi(z)dz = 0{,}9728 .$$

CAMP calculou diretamente os têrmos do desenvolvimento binomial, encontrando o verdadeiro valor *0.9735*.

Tratando-se de freqüência relativa, P_δ dá a probabilidade que a diferença entre a proporção de acontecimentos favoráveis verificada x/n e a verdadeira probabilidade p satisfaça a relação $\left|\frac{x}{n} - p\right| \leq \delta \left(\frac{pq}{n}\right)^{1/2}$.

Baseando-nos nas relações acima, podemos estabelecer aproximadamente limites fiduciais em relação a p. A solução é idêntica à exposta para a média numa amostragem de variáveis. Por exemplo, o intervalo fiducial de 95% cobrirá aproximadamente a amplitude $p - 1.96\,\sigma$ a $p + 1.96\,\sigma$.

3.9 Aproximação por uma série de GRAM-CHARLIER.

Vimos que, quando $p \neq q$, o polígono binomial era assimétrico. Ao lhe substituirmos a curva normal, a aproximação, satisfatória para os valores centrais, piora nas extremidades. Podemos melhorá-la utilizando no ajustamento uma curva assimétrica.

Retomemos o desenvolvimento em série de $f(z)$ do § 3.7. Introduzamos os polinômios de Hermite, definidos por

$$H_n = z^n - C_2^n z^{n-2} + 1.3\, C_4^n z^{n-4} - 1.3.5\, C_6^n z^{n-6} + \ldots$$

$$(-1)^n \frac{d^n}{dz^n} e^{-\frac{z^2}{2}} = e^{-\frac{z^2}{2}} H_n(z) .$$

Daí o desenvolvimento de $f(z)$ em série de GRAM-CHARLIER *(tipo A)*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} [1 - a_3 H_3(z) + a_4 H_4(z) - \ldots] . \qquad (3.6)$$

A probabilidade de que a variável z caia no intervalo 0 a δ, depende do sinal de δ, e resulta da integral

$$U(\delta) = \frac{1}{\sqrt{2\pi}} \int_0^\delta e^{-\frac{z^2}{2}} dz \pm \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} [a_3 H_3(z) - a_4 H_3(z) + \ldots] . \qquad (3.7)$$

Os valores dos coeficientes a_3, a_4, ... em função dos momentos reduzidos da distribuição obtém-se imediatamente, referindo-os à expressão de $\varphi(t)$. Temos

$$a_3 = \frac{1}{3!} \frac{p-q}{\sigma} = -\frac{\alpha_3}{3!}, \quad a_4 = \frac{1}{4!} \frac{1-6\,pq}{\sigma^3} = \frac{\alpha_4 - 3}{4!}, \ldots$$

Em geral, será suficiente conservar apenas o primeiro têrmo do desenvolvimento. Temos, assim, a forma recomendada por EDGEWORTH e BOWLEY

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \left[1 + \frac{\alpha_3}{3!} (z^3 - 3z)\right], \qquad (3.8)$$

e portanto

$$U(\delta) = \int_0^\delta \phi(z) dz \mp \frac{\alpha_3}{3!} \left[\phi'(z)\right]_0^\delta . \qquad (3.9)$$

O cálculo do têrmo corretivo faz-se mediante as tabelas das derivadas da função normal.[5] BOWLEY [6] dá diretamente os valores da função

$$g(\delta) = \frac{1}{3!}\left[\phi''(z)\right]_0^{\delta} = \frac{1}{3!\sqrt{2\pi}}\left[e^{-\frac{\delta^2}{2}}(\delta^2 - 1) + 1\right].$$

Exemplo. Determinemos, em segunda aproximação, a probabilidade referida no EX. 3.8.

Temos $\alpha_3 = \frac{1}{10}\left(\frac{2}{3} - \frac{1}{3}\right) = \frac{1}{30}$. Apliquemos a fórmula (3.9). Entrando com os valores $\delta_1 = -1.95$ e $\delta_2 = 2.95$ nas tabelas de BOWLEY, encontramos $g(\delta_1) = .0943$, $g(\delta_2) = .0731$.

O têrmo corretivo é pois $^1/_{30}(.0943 - .0731) = .0007$, donde se conclui a probabilidade $Q\delta = .7928 + .0007 = .9735$, idêntica ao valor verdadeiro.

3.10 Significância da diferença entre proporções.

Ocorre freqüentemente a necessidade de verificar a hipótese de que as proporções p_1 e p_2, obtidas em duas amostras de tamanhos n_1 e n_2, são compatíveis com a amostragem em um único universo, caracterizado por p.

Temos então que o valor provável da diferença $E(p_1 - p_2) = 0$, de modo que o teste será verificar se a diferença observada é significantemente diferente de zero. O êrro padrão da diferença, admitida a independência entre as amostras, é (§2.11)

$$\sigma_{p_1 - p_2} = \left(\sigma_{p_1}^2 + \sigma_{p_2}^2\right)^{1/2} = \left(pq\right)^{1/2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}. \qquad (3.10)$$

Logo, a razão $z = \frac{p_1 - p_2}{\sigma_{p_1 - p_2}}$ varia em tôrno de zero com desvio padrão unitário.

O cálculo dos momentos superiores da distribuição de z mostra que, para valores fixos de p e q, à medida que o tamanho da amostra cresce, $\alpha_3 \to 0$ e $\alpha_4 \to 3$. Mesmo para amostras de tamanho moderado, há uma razoável aproximação da forma normal. Então, a probabilidade de obtermos uma diferença numèricamente igual ou maior que $|\delta|$ é dada por $P_{\delta} = 2\int_{\delta}^{\infty} \phi(z)dz$, onde δ é a diferença observada em unidade do desvio padrão.

E. S. PEARSON indica as seguintes regras práticas: denotemos por n_1 a menor das amostras $(n_1 < n_2)$; se $n_1 p > 5$, pode-se usar a distribuição normal. Se $n_1 p \leq 5$, ela ainda pode ser usada, desde que $\alpha_3 < .02$; para valores maiores de α_3, o teste não inspira confiança.

As fórmulas que vimos de deduzir são exatas, isto é, elas se exprimem em função dos parâmetros do universo. Freqüentemente, êste valor é desconhecido, e temos de estimá-lo, tão exatamente quanto possível, a partir da amostra e de conformidade com a hipótese a ser verificada.

Por exemplo, a variância da diferença entre proporções em amostras independentes é $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$. Ao verificar a hipótese de que as amostras provêm do mesmo universo, ambos os p são iguais, e devemos, pois, estimá-lo baseando-nos no conjunto das duas amostras.

---

[5] GLOVER, J. W., *Tables of Applied Mathematics* (Ann Arbor, 1930); RIETZ, H. L., *Handbook of Mathematical Statistics* (Boston, 1924).

[6] BOWLEY, A. L., *Elements of Statistics* (5.ª ed., Londres, 1926), pág. 303.

Temos

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \qquad (3.11)$$

isto é, a melhor estimativa de p é a média ponderada das proporções nas amostras. Demonstra-se que $E(\hat{p}) = p$.

Ao contrário, se desejamos uma medida da dispersão provável das diferenças entre pares de amostras oriundas de diferentes universos, cada um dos p deve ser estimado em concordância com a respectiva amostra.

Exemplo. Uma tabulação dos ciclos econômicos, feita por MITCHELL (*Business Cycles*, pág. 400), mostra que, no período 1890-1905, em 7 países industriais observaram-se 43 ciclos, dos quais 17 com duração de 3 - 4 anos, e em 10 países não industriais observaram-se 62 ciclos, com 14 dessa duração. Pode-se inferir que as fôrças cíclicas são diferentes num e noutro tipo de economia?

A proporção de ciclos de 3 - 4 anos é, no primeiro caso, 39.5%, e no segundo, 22.6% Temos

$$\hat{p} = .295,\ \sigma^2_{p_1 - p_2} = \left(.295\right)\left(.705\right)\left(\frac{1}{43} + \frac{1}{62}\right) = .0906,\ z = 1.87,$$

e, entrando na tabela da curva normal, achamos $P\delta = .0614$.

A diferença não é significante. As observações não apoiam a hipótese formulada.

3.11 *A lei de* POISSON *ou dos fenômenos raros.*

Em muitos setores das ciências naturais encontramos fenômenos em que a probabilidade do acontecimento é extremamente pequena, dando origem aos chamados fenômenos raros. Assim, o número de pessoas centenárias que morrem num dado ano, o número de bactérias existentes numa quadrícula de lâmina de ensaio, o número de vêzes que, num dado período, o volume de negócios numa bôlsa ou a descarga de um rio excedem certo limite, etc.

Suponhamos, pois, que a probabilidade p seja extremamente pequena, mas n muito grande, de modo que o número provável de acontecimentos np tenha um valor finito μ.

O número x de realizações do acontecimento tem por função característica $\psi(t) = (pe^t + q)^n = [1 + p(e^t - 1)]^n$. Ponhamos $u = e^t - 1$; vem

$$\psi(t) = (1 + pu)^n = e^{n \log (1+pu)} = e^{n[pu - \frac{1}{2} p^2u^2 + \frac{1}{3} p^3u^3 - \ldots]}$$

e, como $np = \mu$, segue-se que

$$\psi(t) = e^{\mu u} \cdot e^{\mu \left(-p\frac{u^2}{2} + \frac{p^2u^3}{3} - \ldots\right)}$$

$$= e^{\mu u}\left(1 - \frac{\mu p}{2} u^2 + \frac{\mu p^2}{3} u^3 - \ldots\right). \qquad (3.12)$$

Como p é muito pequeno, podemos desprezar os têrmos que contêm suas potências sucessivas, e temos, em primeira aproximação, $\psi(t) = e^{\mu u} = e^{\mu(e^t - 1)}$. Ora, sabemos (§ 2.4 Ex.) que a variável aleatória admitindo essa função característica é a variável descontínua que toma os valores $x = 0, 1, 2, \ldots$ com probalidade

$$w(x) = e^{-\mu}\frac{\mu^x}{x!}. \qquad (3.13)$$

E' essa a *lei exponencial de* POISSON ou *dos fenômenos raros*, também chamada por VON BORTKIEWICZ de lei dos pequenos números. Os valores dessa função acham-se tabelados na coletânea de PEARSON.[7]

Contendo a função de POISSON um único parâmetro, μ, o seu ajustamento vai depender apenas do primeiro momento da distribuição empírica. Como critério para verificar se uma série de observações segue essa lei, podemos nos basear na relação entre as freqüências sucessivas

$$\frac{w(x+1)}{w(x)} = \frac{\mu}{x+1}, \quad \text{ou seja} \quad \frac{w(x+1)}{w(x)}(x+1) = \mu .$$

Os valores do primeiro membro, calculados sôbre os dados observados, devem ser pràticamente constantes.

Os têrmos da série

$$e^{-\mu}\left(1 + \mu + \frac{\mu^2}{2!} + \ldots + \frac{\mu^x}{x!}\right)$$

dão as probabilidades de se obterem exatamente *0, 1, 2, ...* ocorrências do fenômeno raro em n observações. Em geral, porém, interessa-nos conhecer a probabilidade de termos, em n observações, r ou menos acontecimentos favoráveis. Essa probabilidade resulta da soma dos têrmos da série supra até o de ordem $r+1$

$$P_r = e^{-\mu} \sum_{x=0}^{r} \frac{\mu^x}{x!}. \qquad (3.14)$$

A soma dos primeiros têrmos da série de POISSON é dada como $I\,(u,p)$ nas *Tables of the Incomplete Gamma Function*,[8] onde $u = \mu\sqrt{r}$ e $p = r - 1$, para todos os valores de *0* até *52*. A coletânea de PEARSON contém também uma tábua, preparada por Miss WHITTAKER,[9] dando as probabilidades de desvios em relação às freqüências teóricas, no pressuposto da distribuição poissoniana.[10] PEARSON aconselha o emprêgo dessa distribuição desde que $p < .03$.

Exemplo. RIETZ (*Mathematical Statistics*, pág. 43), considera o caso da experiência de uma companhia com um grupo de 10,000 assegurados de 30 anos de idade, em que se verificaram 6 óbitos por pneumonia. Qual é a probabilidade de, em casos análogos, verificarem-se não menos de 3 nem mais de 9 mortes?

As tabelas de Miss WHITTAKER dão imediatamente, em correspondência ao valor provável 6, que é de se esperar desvios menores que 3 em 6.20% dos casos, e maiores que 3 em 8.39%; donde a probabilidade procurada de 14.59%.

A aplicação da aproximação normal daria $\delta = 1.43$, $Q\delta = 0.847$, o que corresponde a desvios maiores ou menores que 3 em 15.3% dos casos. RIETZ calculou o valor verdadeiro de $Q\delta$ pelo desenvolvimento binomial, achando 0.854, o que corresponde à probabilidade de desvios de 14.6%. Como se vê, a aproximação de POISSON dá resultado pràticamente idêntico ao verdadeiro.

---

[7] PEARSON, K. (ed.), *Tables for Statisticians and Biometricians*, vol. 1 (3.º ed., Londres, 1930), tabela LI.

[8] PEARSON, K. (ed.), *Tables of the Incomplete Gamma Function* (Londres, 1922).

[9] PEARSON, K. (ed.), *Tables for Statisticians and Biometricians*, vol. 1 (3.º ed., Londres, 1930), tabela LII.

[10] Para um cálculo expedito, podem-se utilizar nomogramas, como os de Miss THORNDIKE, *Applications of Poisson's Probability Summation* (Bell Telephone Laboratories, Mon. B-220).

3.12 Forma limite da lei de POISSON.

Tomando o logarítmo de $\varphi(t)$, temos a função geratriz de semi-invariantes da lei de POISSON

$$L(t) = \mu(e^t - 1) = \mu\left[t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots\right]$$

Logo, todos os semi-invariantes λ_i são iguais ao valor médio μ. Em particular, a variância é igual à média. Daí se obtêm os coeficientes de assimetria e kurtosis, $\sqrt{\beta_1} = \frac{1}{\sqrt{\mu}}$ e $\beta_2 = \frac{1}{\mu} + 3$.

A amplitude da função de POISSON vai de 0 a ∞. Para $\mu < 1$, a forma do polígono de probabilidade é de *talho-J;* para $\mu \geqslant 1$, é de *talho-I*, e para grandes valores de μ tende a se tornar simétrica.

Com efeito, se μ cresce, a sua probabilidade $w(\mu) = \frac{\mu^\mu e^{-\mu}}{\mu!} \to \frac{1}{\sqrt{2\pi\mu}}$, ao mesmo tempo que $\sqrt{\beta_1} \to 0$ e $\beta_2 \to 3$; isto é, essas grandezas tendem para os valores característicos da distribuição gaussiana.

A forma limite da lei dos fenômenos raros para valores crescentes de μ pode ser deduzida mediante a sua função característica

$$\psi(t) = e^{\mu(e^t - 1)} = e^{\left(\mu t + \mu\frac{t^2}{2!} + \cdots\right)}$$

Mudando a escala das abscissas, de modo a tornar $\sqrt{\mu}$ igual à unidade, obtém-se

$$\psi(t) = e^{\left(\sqrt{\mu}t + \frac{t^2}{2!} + \frac{t^3}{3!\sqrt{\mu}} + \cdots\right)}$$

Em primeira aproximação, teremos, desprezando os têrmos do terceiro em diante, $\psi(t) = e^{\sqrt{\mu}t + \frac{t^2}{2!}}$, que é a função característica da lei normal

$$w(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x - \sqrt{\mu}\right)^2}. \qquad (3.15)$$

Para grandes valores de μ retornamos assim da lei dos fenômenos raros à distribuição gaussiana, com média igual a $\sqrt{\mu}$ e desvio padrão unitário.

Para obter uma aproximação geral da lei de POISSON, volvamos à expressão (3.12). Devemos procurar as leis de probabilidade correspondentes a funções características da forma $u^k e^{\mu u}$. Ora, sendo esta a derivada k^{esima} em relação a μ da função $e^{\mu u}$, é claro que a lei de probabilidade correspondente será a derivada k^{esima} de

$$\theta(x) = e^{-\mu}\frac{\mu^x}{x!}.$$

Temos $\theta_1(x) = \frac{\partial}{\partial\mu}\theta(x) = \frac{\mu^{x-1}}{(x-1)!}e^{-\mu} - \frac{\mu^x}{x!}e^{-\mu} = \theta(x-1) - \theta(x)$,

isto é, $\theta_1\ (x)$ é a diferença inversa primeira de $\theta\ (x)$, representada por $\nabla^1\ \theta\ (x)$. Do mesmo modo, $\theta_2(x)$, $\theta_3(x)$, ... serão as diferenças inversas sucessivas de $\theta\ (x)$.

Por conseguinte, a função de probabilidade $\omega(x)$, admitindo como função característica *(3.12)*, tem por expressão

$$w(x) = \theta(x) - \frac{\mu p}{2} \nabla^2\theta + \frac{\mu p^2}{3} \nabla^3\theta(x) - \ldots \qquad (3.16)$$

ou seja, é uma série de POISSON-CHARLIER (*tipo B*).

Capítulo IV

A DISPERSÃO LEXIANA. DISTRIBUIÇÕES DE CONTÁGIO

4.1 A teoria de Lexis.

A existência de fenômenos caracterizados por uma probabilidade constante nem sempre se verifica na natureza. Seria irreal, por exemplo, supôr a probabilidade de morte constante para todos os indivíduos de uma dada população e inalterável no decorrer do tempo.

Se a observação de um fenômeno nos fornece uma série de freqüências $x_1, x_2, \ldots$, baseadas em $n_1, n_2, \ldots$ provas ou casos, resta saber se as variações das freqüências relativas são puramente fortuitas ou atribuíveis a alguma causa específica. Não se verificando a constância nas probabilidades de realização do acontecimento, devemos examinar outras hipóteses, em que ela seja variável.

A teoria de Lexis [1] estabelece três espécies de normas para comparação da dispersão observada na série empírica, caracterizadas pelas seguintes propriedades:

1) a probabilidade do acontecimento permanece constante durante tôdas as observações; obtém-se a *série de* Bernoulli;

2) a probabilidade do acontecimento varia de uma prova a outra, dentro de uma série de n tiragens, mas essas probabilidades são constantes de uma a outra série de provas; obtém-se a *série de* Poisson;

3) a probabilidade do acontecimento é constante em tôdas as provas de uma série, mas varia de uma série a outra; obtém-se a *série de* Lexis.

O caso de Poisson corresponde, por exemplo, a termos várias regiões de condições mesológicas idênticas, povoadas por n indivíduos de diferentes idades, e portanto com probabilidades de mortes diferentes. No caso de Lexis, os grupos seriam indivíduos da mesma idade, e, pois, com a mesma probabilidade de morte, mas variável de uma região a outra.

A dedução das fórmulas torna-se mais compreensível assimilando os casos a esquemas de urnas, e admitindo o número de provas constante em cada série. No caso de Bernoulli, as tiragens fazem-se de uma única urna, contendo bolas brancas e pretas numa relação constante, isto é, com reposição após cada tiragem. Vimos que o número provável de bolas brancas extraídas em cada série era np, e a variância $\sigma_B^2 = npq$. A dispersão diz-se então *normal*.

4.2 O esquema de Poisson.

Consideremos n urnas, contendo bolas brancas e pretas, e de composições diferentes $p_1, p_2, \ldots p_n$, a composição p_i sendo representada pelo mesmo número que a probabilidade de tirar uma bola branca na urna de ordem i.

---

[1] A memória de Lexis, *Zur theorie der Massenerscheinungen*, foi publicada em 1877; um pouco antes, o atuário Dormoy propusera um coeficiente análogo, denominado de *divergência*, e baseado no afastamento médio.

Tiremos uma bola de cada uma das urnas, e repitamos a operação k vêzes; obtém-se uma série estatística, $x_1, x_2, \ldots x_k$, representando as freqüências de aparecimento da bola branca em cada série de tiragens. E' a dispersão teórica dessa série que vamos calcular.

A variância das probabilidades p_i é

$$\sigma_p^2 = \frac{1}{n} \sum_{i=1}^{n} (p_i - p)^2 \text{, onde } p_i = \frac{1}{n} (p_1 + p_2 + \ldots + p_n) .$$

Em uma série qualquer de tiragens, o número de aparições da bola branca é, em média, $p_1 + p_2 + \ldots + p_n = np$. O afastamento quadrático médio dessa grandeza aleatória é a soma dos afastamentos quadráticos médios relativos a cada tiragem, que se obtém fazendo $n = 1$ na fórmula habitual $np_i q_i$; temos, pois,

$$p_1q_1 + p_2q_2 + \ldots + p_nq_n = \sum_{i=1}^{n} p_iq_i .$$

Considerando tôdas as séries de tiragens, teremos a variância do esquema poissoniano

$$\sigma_P^2 = \sum_{i=1}^{n} p_iq_i = \sum_{i=1}^{n} \Big[p + (p_i - p) \Big] \Big[q - (p_i - p) \Big] = npq - \sum_{i=1}^{n} (p_i - p)^2$$

pois que $\sum_{i=1}^{n} (p_i - p) = 0.$

Conclui-se que a dispersão dos x é dada por

$$\sigma_P^2 = \sigma_B^2 - n\, \sigma_p^2 , \qquad (4.1)$$

o que mostra que a dispersão da série de Poisson é menor que a da série de Bernoulli correspondente, em que a probabilidade constante do acontecimento é igual à média aritmética das probabilidades variáveis observadas. A série diz-se, então, *hipo-normal*.

4.3 O esquema de Lexis.

Consideremos agora k urnas contendo bolas brancas e pretas, e de composições diferentes $p_1, p_2, \ldots p_k$. Em uma urna qualquer efetuemos n tiragens sucessivas com reposição; e repitamos a operação para as demais urnas.

Como anteriormente, a variância das probabilidades será

$$\sigma_p^2 = \frac{n}{k} \sum_{i=1}^{k} (p_i - p)^2 \text{, com } p = \frac{1}{n} (p_1 + p_2 + \ldots + p_k) .$$ O número de bolas brancas extraído da urna i é uma variável aleatória de média np_i e variância, em relação a essa média, $np_i q_i$. Calculemos o afastamento quadrático médio relativo a np; temos $np_i q_i + (np_i - np)^2$.

Repetindo as tiragens em cada uma das urnas, teremos uma seqüência de bolas brancas $x_1, x_2, \ldots x_k$, cuja média será np e cuja variância será

$$\sigma_L^2 = \frac{n}{k} \sum_{i=1}^{k} p_iq_i + \frac{n^2}{k} \sum_{i=1}^{k} (p_i - p)^2 .$$

Vimos, porém, que $\sum_{i=1}^{k} p_i q_i = kpq - \sum_{i=1}^{k} (p_i - p)^2$, donde

$$\sigma_L^2 = npq + \frac{n^2 - n}{k} \sum_{i=1}^{k} (p_i - p)^2 = \sigma_B^2 + n(n-1)\sigma_p^2 . \qquad (4.2)$$

Por conseguinte, a dispersão da série de LEXIS excede a da série de BERNOULLI correspondente de uma quantidade que cresce ràpidamente com o número de tiragens. A série diz-se *hiper-normal.*

4.4 Critérios de LEXIS *e* CHARLIER.

A discriminação entre os três tipos de dispersão faz-se pelo *critério de* LEXIS

$$L = \frac{s}{\sigma_B} . \qquad (4.3)$$

onde s é o desvio padrão calculado diretamente sôbre os dados observados, e σ_B é o valor teórico correspondente à hipótese de uma distribuição bernoulliana.

Êsse critério é, entretanto, sensível às flutuações nos valores das probabilidades p_i e cresce com o aumento do número de observações. Com efeito, temos $L^2 = 1 + \frac{n-1}{pq}\sigma_p^2$. Para obviar êsses inconvenientes, CHARLIER propôs outro coeficiente para medir a intensidade das influências perturbadoras. Considerando que, no esquema lexiano, temos aproximadamente

$\sigma_L^2 = \sigma_B^2 + n\sigma_p^2$, segue-se que $\left(\frac{\sigma_p^2}{p}\right)^2 = \frac{\sigma_L^2 - \sigma_B^2}{n^2 p^2}$.

Temos assim, expresso como percentagem, o *coeficiente de perturbação de* CHARLIER

$$C = \frac{100}{\bar{x}} \sqrt{s^2 - \sigma_B^2} , \qquad (4.4)$$

sendo $\bar{x}$ a média aritmética dos dados. Vê-se que êle mede a flutuação relativa das probabilidades, e independe de n, pois que agora

$$C = \frac{100}{p} \sigma .$$

Quando $L = 1$ ou $C = 0$, temos uma série normal ou de BERNOULLI; quando $L < 1$ ou C imaginário, ela é hiponormal ou de POISSON; quando $L > 1$ ou $C > 0$, ela é hipernormal ou de LEXIS.

Exemplo. S. N. MAGALHÃES dá (*Rev. Bras. Est.*, vol. 1, pg. 253) os coeficientes de masculinidade no Distrito Federal de 1890 a 1938, tendo por média $p = .5163$ e dispersão $s = .00947$. Para calcular a dispersão correspondente à hipótese bernoulliana, $\sigma_B = \sqrt{\frac{pq}{n}}$, notemos que, como o número total de nascimento varia de ano a ano, devemos tomar para n a média harmônica dos valores ânuos observados. Tem-se assim $\sigma_B = \sqrt{\frac{(.5163)\ (.4837)}{23570}} = .0035$.

Daí o coeficiente de dispersão de LEXIS $L = 2.914$, denotando a hipernormalidade da série.

Quanto ao coeficiente de perturbação de CHARLIER, temos

$$C = \frac{100}{.5163}\sqrt{(.00947)^2 - (.00325)^2} = 1.723 .$$

4.5 *A dispersão lexiana e as modernas teorias.*

A importância da teoria da dispersão de LEXIS decorre de que, com a sua introdução, poude-se esclarecer os efeitos resultantes da supressão da hipótese de uma probabilidade constante como causa dos fenômenos. Depois, constitui ela um elemento de ligação entre as teorias clássicas e os modernos testes de significância.

Veremos que o critério de LEXIS guarda estreita ligação com o teste CHI-QUADRADO. Por outro lado, a fórmula da dispersão no esquema de LEXIS mostra a variância desdobrada em duas componentes parciais. A componente bernoulliana pode ser atribuda ao acaso, pois ela aparece ainda quando a probabilidade do fenômeno é constante, enquanto que a componente lexiana representa o efeito das alterações sistemáticas da probabilidade entre uma prova e outra. Essa resolução da variância em componentes distintas foi generalizada por R. A. FISHER no método da *análise da variância*, permitindo a avaliação dos efeitos da componente aleatória e de várias componentes sistemáticas simultâneas.[2]

4.6 *Os fenômenos de contágio.*

A teoria de LEXIS pressupõe que todos os acontecimentos sejam independentes; um valor de s maior ou menor que σ_B pode, pois, indicar, como interpretação alternativa, que os acontecimentos são correlacionados positiva ou negativamente.

Tem-se observado, por exemplo, que as taxas de mortalidade apresentam dispersões consideràvelmente maiores que as correspondentes à lei de BERNOULLI. Uma explicação dessa hipernormalidade é que freqüentemente uma epidemia, uma calamidade climática, etc., provocam um exagerado aumento do número de falecimentos. Essas circunstâncias ocasionam a morte simultânea de grupos de indivíduos; e essas mortes, que se prendem a uma mesma causa, não se podem considerar independentes umas das outras.

A fim de representar matemàticamente tais circunstâncias, POLYA[3] generalizou o esquema de BERNOULLI, introduzindo a noção de *contágio*. Diz-se que há contágio numa série de tiragens ou provas quando o resultado de uma depende do resultado da prova precedente.

Consideremos uma urna contendo r bolas brancas e s bolas pretas $(r + s = N)$, de modo que a probabilidade inicial de extração de uma bola branca é $p = r/N$ e de uma preta $q = s/N$. Após cada tiragem, recoloca-se na urna $1 + \triangle$ bolas da côr tirada. Para $\triangle = 0$, temos o esquema de BERNOULLI; para $\triangle = r - 1$, o esquema de bolas sem reposição, isto é, de amostragem num universo finito. Para valores arbitrários de $\triangle$ teremos novas distribuições. Se $\triangle$ é positivo, um acontecimento atrai outro; então as probabilidades do aconteci-

---

[2] Sôbre o assunto, veja-se GEIRINGER, H., "Observations on Analysis of Variance Theory", *An. Mathem. Stat.*, vol. XIII (1942), pág. 350.

[3] POLYA, G., "Sur quelques points de la théorie des probabilités", *Annales de l'Institut Henri Poincaré*, vol. I (1932), pág. 117.

mento aumenta pela sua repetição, e o jôgo pode continuar indefinidamente. No caso contrário ($\triangle < 0$), há repulsão entre os acontecimentos, a probabilidade diminui com a repetição, dá-se uma mútua imunização; o jôgo dura até que a urna se esvasie.

4.7 A distribuição de contágio de POLYA.

Suponhamos que se fazem n tiragens, e deseja-se calcular a probabilidade $\omega(x)$ de tirar x bolas brancas e $n - x$ bolas pretas numa ordem qualquer. A probabilidade de só tirar bolas pretas é

$$\omega(o) = \frac{s}{N} \cdot \frac{s + \Delta}{N} \cdots \frac{s + (n-1)\Delta}{N + (n-1)\Delta} = \frac{\prod_{k=o}^{n-1} (s + k\Delta)}{\prod_{k=o}^{n-1} (N + k\Delta)}$$

A probabilidade de tirar primeiro x bolas brancas e depois $n - x$ pretas é

$$\frac{\prod_{i=o}^{x-1} (r + i\Delta) \prod_{j=x}^{n-x-1} (s + j\Delta)}{\prod_{k=o}^{n-1} (N + k\Delta)}$$

Para ter a probabilidade de uma figuração qualquer, basta multiplicar êsse resultado pelo coeficiente binomial C_x^n, pois que êste é o número de ordens possíveis em que os acontecimentos podem ocorrer em n provas.

Introduzamos as probabilidades p, q e $\triangle/N = \delta$, que será o *coeficiente de contágio* ou *repulsão* conforme seja positivo ou negativo. Obtém-se

$$\omega(x) = C_x^n \frac{\prod_{i=o}^{x-1} (p + i\delta) \prod_{j=o}^{n-x-1} (q + j\delta)}{\prod_{k=o}^{n-1} (1 + k\delta)}$$

Como $\omega_o = \prod_{k=o}^{n-1} \frac{q + k\delta}{1 + k\delta}$, vem $\omega(x) = C_x^n \, w(o) \frac{\prod_{i=o}^{x-1} (p + i\delta)}{\prod_{k=n-x}^{x-1} (q + k\delta)}$.

Notando que o denominador, escrito em ordem inversa, é

$$\prod_{i=o}^{x-1} [q + (n - 1 - i)\delta] ,$$

obtemos finalmente

$$\omega(x) = \frac{\omega(o)}{x!} \prod_{i=o}^{x-1} \frac{(p + i\delta)(n - i)}{q + (n - 1 - i)\delta} . \qquad (4.5)$$

A relação entre as freqüências sucessivas será dada por

$$\omega(x+1) = \omega(x)\,\frac{n-x}{x+1}\,\frac{p+x\delta}{q+(n-1-x)\delta}\,.$$

Para calcular os momentos,[4] escrevamos essa expressão sob a forma

$$[np + x(n\delta - p) - x^2\delta]\,\omega(x) = (x+1)\,[q + n\delta - (x+1)\delta]\,\omega(x+1)\,, \qquad (4.6)$$

e somenos para todos os valores de x. Notando que

$$\sum_{x=0}(x+1)^k\omega(x+1) = \sum_{x=1} x^k\omega(x) = \sum_{x=0} x^k\omega(x),$$ obtém-se $np + \mu_1'(n\delta - p) - \delta\mu_2' =$

$$= \mu_1'(q + n\delta) - \delta\mu_2'\,,$$ ou seja $\mu_1' = np\,.$

A expectância de x é, pois, independente do contágio ou repulsão; ela conserva sempre o valor bernoulliano.

Para obter a variância, multipliquemos novamente a relação (4.6) por x e somenos; vê-se, empregando a relação simbólica

$\Sigma(x) = \Sigma(x+1) - \Sigma$, que $(\mu_1')^2 + \mu_2'\,(n\delta - p) - \mu_3'\,\delta = \mu_2'\,(q + n\delta) - \mu_3' - (q + n\delta)\mu_1' + \mu_2'\,\delta,$

ou seja, $(\mu_1')^2 + \mu_1'\,(q + n\delta) = \mu_2'\,(1 + \delta)\,.$

Daí se conclui $$\sigma^2 = \frac{np}{1+\delta}\,(q + n\delta - n\delta p) = \frac{npq}{1+\delta}\,(1 + n\delta)\,. \qquad (4.7)$$

Logo, a dispersão aumenta com o contágio, diminui com a repulsão. A correlação positiva entre os acontecimentos provoca uma hipernormalidade, a correlação negativa uma hiponormalidade.

Sabemos que, nos fenômenos sociais, depara-se mais freqüentemente a dispersão hipernormal. Isso quer dizer que as conexões positivas são mais freqüentes na vida social que a independência ou repulsão.

4.8 Formas especiais. A distribuição hipergeométrica.

O esquema de POLYA gera várias formas especiais, segundo certos valores particulares da probabilidade p, do coeficiente de contágio δ e do número fictício de provas n.

Quando Δ—*1*, temos as tiragens sem reposição, ou seja, a amostragem num universo finito. Então

$$\omega(x) = C_x^n\,\frac{r!}{(r-x)!}\,\frac{s!}{[s-(n-x)]!}\,\frac{(N-n)!}{N!} = \frac{C_x^r\,C_{n-x}^s}{C_n^N} = \frac{C_x^{Np}\,C_{n-x}^{Nq}}{C_n^N} \qquad (4.8)$$

distribuição descontínua, chamada *hipergeométrica*, porque a probabilidade das diversas figurações é dada pelos têrmos de uma série hipergeométrica. A variância é então

$$\sigma^2 = npq\,\frac{N-n}{N-1} \qquad (4.9)$$

Se $n = N$, a distribuição degenera; quando se esgota a urna, temos absoluta certeza de conhecermos a sua composição. Para $p = q$, a distribuição é simétrica em tôrno de

$$\bar{x} = \frac{n}{2}, \text{ pois } \omega\left(\frac{n}{2} + z\right) = \omega\left(\frac{n}{2} - z\right).$$

---

[4] Sôbre o cálculo do momento de ordem n mediante a função característica, veja-se DIEULEFAIT, C. E., "Sui momenti delle distribuzioni ipergeometriche", *Giorn. dell'Ist. Ital. degli Attuari*, vol. 10 (1939), pág. 221.

Para $N \to \infty$ ela aproxima-se da distribuição normal com variância

$$\sigma^2 = npq\left(1 - \frac{n}{N}\right). \qquad (4.10)$$

POLYA estudou também os casos limites $n \to \infty$ para valores positivos de δ. Se p é aproximadamente igual a q, e $\delta \to 0$ com $n\delta \to d$, obtém-se uma curva gaussiana com variância $\sigma^2 = npq\,(1 + d)$; é o caso de contágio fraco. Se $\delta > 0$, $x = tn$, temos o contágio forte, e

$$\varphi(t) = \frac{\Gamma\left(\frac{1}{\delta}\right)}{\Gamma\left(\frac{p}{\delta}\right)\Gamma\left(\frac{q}{\delta}\right)}\, t^{p/\delta}\,(1 - t)^{\frac{q-\delta}{\delta}}.$$

E' esta uma curva do sistema pearsoniano.

4.9 Fenômenos raros com contágio.

A noção de contágio pode-se estender ao estudo dos fenômenos raros. Consideremos o esquema de POLYA para o caso especial de uma probabilidade pequena $(p \to 0)$ e um contágio fraco $(\delta \to 0)$. Admitamos que $n \to \infty$ com $np = \mu$, $n\delta = d$. Vimos que a expectância de x era independente do contágio; quanto à dispersão, obtêm-se (4.7) $\sigma^2 = \mu\,(1 + d)$, sendo pois superior à dispersão da lei clássica de POISSON. Introduzindo os valores de μ e d em (4.5), vem

$$\omega(x) = \frac{\omega(o)}{x!}\prod_{i=0}^{x-1}\frac{\left(\mu + id\right)\left(1 + \frac{i}{n}\right)}{1 + d - \frac{(i+1)\,d}{n}},$$

que, em vista da condição $n \to \infty$, se reduz a

$$\omega(x) = \frac{\omega(0)}{x!\,(1 + d)^x}\prod_{i=0}^{x-1}(\mu + id).$$

O coeficiente $\omega(0)$ determina-se pela condição de ser a área da curva de probabilidade igual à unidade, obtendo-se $\omega(0) = \left(\frac{1}{1 + d}\right)^{\frac{\mu}{d}}$.

Resulta finalmente

$$\omega(x) = \frac{[\mu + (x - 1)\,d]\,!}{x!\,(\mu - 1)!\,(1 + d)^{x + \frac{\mu}{d}}} \qquad (4.11)$$

para forma analtica da distribuição.

Capítulo V

CHI-QUADRADO E A VERIFICAÇÃO DE LEIS EMPÍRICAS

5.1 *A lei multinomial.*

A comparação entre duas distribuições tem sido feita, até aqui, através de seus índices característicos. Ora, salvo casos excepcionais, êsses índices não sintetizam tôdas as peculiaridades das mesmas, e há, pois, vantagem em se estabelecer um processo que permita a comparação global das distribuições. Foi o que conseguiu Karl Pearson com o seu *teste de aderência* (goodness of fit)[1].

Consideremos um evento caracterizado pela variável aleatória v, que admite k valores $v_1, v_2, \ldots v_k$, com probabilidades $p_1, p_2, \ldots p_k$, sendo $\sum_1^k p_i = 1$. A probabilidade de que, em n provas independentes, tenhamos m_1 eventos da classe v_1, m_2 da classe v_2, e assim por diante, é dado por

$$p\,[m_i] = \frac{n!}{m_1!\; m_2!\; \ldots\; m_k!}\; p_1^{m_1} \cdot p_2^{m_2} \cdot \ldots\; p_k^{m_k},$$

onde $[m_i]$ representa abreviadamente $(m_1, m_2, \ldots . m_k)$ e $\sum_1^k m_i = n$. E' êste o têrmo geral do desenvolvimento do multinômio $(p_1+p_2+\ldots+p_k)^n$. O desenvolvimento binomial, considerado anteriormente, resulta como caso particular para $k = 2$.

A fórmula acima é, contudo, pràticamente inusitável, e vamos transformá-la, admitindo que os m_i são suficientemente grandes para que possamos substituir os fatoriais pelos valores dados pela fórmula de Stirling

$$m! = (2\pi)^{\frac{1}{2}}\, m^{m+\frac{1}{2}}\, e^{-m}.$$

Introduzindo tais valores, obtém-se, após certas reduções,

$$p[m_i] = \frac{1}{(2\pi n)^{\frac{k-1}{2}}\; (p_1 \cdot p_2 \cdot \ldots\; p_k)^{\frac{1}{2}}} \prod_{i=1}^{k} \left(\frac{np_i}{m_i}\right)^{m_i + \frac{1}{2}}. \qquad (5.1)$$

Computemos os desvios de m_i em relação a suas expectâncias, e alteremos a escala de modo que $x_i = \frac{m_i - np_i}{\sqrt{np_i}}$. Com essa transfor-

---

[1] A memória fundamental de Pearson, "On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling", foi publicada no *Philosophical Magazine*, vol. 50 (1900), pág. 157.

mação, temos que

$$\left(\frac{np_i}{m_i}\right)^{m_i + \frac{1}{2}} = \left(1 + \frac{x_i}{\sqrt{np_i}}\right)^{-\left(np_i + x_i\sqrt{np_i} + \frac{1}{2}\right)}$$

Tomando o logarítmo do segundo membro e desenvolvendo-o em série, vem

$$\left(np_i + x_i\sqrt{np_i} + \frac{1}{2}\right) \log_e\left(1 + \frac{x_i}{\sqrt{np_i}}\right) = x_i\sqrt{np_i} + \frac{1}{2}x_i^2 + \ldots.$$

onde os têrmos seguintes, contendo em denominador as potências de np_i, podem ser desprezados. Daí

$$\left(\frac{np_i}{m_i}\right)^{m_i + \frac{1}{2}} = e^{-x_i\sqrt{np_i} - \frac{1}{2}x_i^2}.$$

Substituindo êsses valores em (5.1), o resultado se simplifica, pois temos então em expoente

$$\sum_1^k x_i\sqrt{np_i} = \sum_1^k (m_i - np_i) = 0, \qquad (5.2)$$

e a fórmula se reduz a

$$p[m_i] = \frac{1}{(2\pi n)^{\frac{k-1}{2}}\,(p_1 . p_2 . \ldots p_k)^{\frac{1}{2}}}\, e^{-\sum_1^k x_i^2}.$$

Ora, em vista da relação linear (5.2), apenas $k-1$ dos desvios x_i são independentes, digamos os primeiros $k-1$. Imaginemos os m e x como coordenadas retangulares num hiper-espaço euclidiano. Seja R a região no espaço x, sujeita àquela restrição, correspondendo à região R_m no espaço m; como os m_i são essencialmente positivos, a variação unitária de m_i corresponderá a $\triangle x_i = (np_i)^{\frac{1}{2}}$, e daí, em virtude do teorema sôbre a existência de integrais definidas,

$$\lim_{n\to\infty} \sum_R p(x_1, x_2, \ldots x_k)\,\triangle x_1\,\triangle x_2 \ldots \triangle x_{k-1} = \frac{1}{(2\pi)^{\frac{k-1}{2}}\,(p_k)^{\frac{1}{2}}} \int_R e^{-\frac{1}{2}\sum_1^k x_i}\, dx, \qquad (5.3)$$

a somação estendendo-se, para um dado n, a todos os pontos da região R correspondente à de R_m para os quais se define $p[m_i]$. Tem-se uma integral k-múltipla, onde $dx = dx_1\, dx_2 \ldots dx_{k-1}$.

5.2 A distribuição de Chi-quadrado.

Obtivemos, assim, a distribuição de probabilidade do conjunto de desvios entre os m_i e suas expectâncias como uma função da expressão quadrática

$$\chi^2 = \sum_1^k x_i^2. \qquad (5.4)$$

Foi êsse o índice que PEARSON adotou para medir a aderência dos valores observados m_i aos valores teóricos. Cumpre agora transformar a diferencial (5.3) numa diferencial do próprio χ^2.

No espaço k-dimensional dos x_i, a relação (5.4) define um sistema de hiper-esferas tendo os centros coincidentes com a origem. Calculemos a integral quando se passa da hiper-esfera de raio χ à de raio $\chi + d\chi$, sujeita à restrição (5.2). Sendo esta uma equação linear homogênea relativamente às discrepâncias, ela representa um hiperplano passando pelo centro comum das esferas, e a integração não se estende a todo o volume entre as duas esferas concêntricas, mas à porção do espaço k—*1* dimensional compreendido entre ambas.

Ora, o volume V de uma hiper-esfera k-dimensional de raio r é $V = Cr^k$, sendo a constante C independente de r; logo, o volume entre as duas esferas concêntricas é aproximadamente $dV = C'r^{k-1}dr$. Aplicando essa fórmula ao caso vertente, e ponde $k' = k - 1$, vemos que a probabilidade de que χ esteja compreendido entre χ e $\chi + d\chi$ é dada por

$$dp = Ke^{-\frac{\chi^2}{2}}\chi^{k'-1}d\chi.$$

A constante K independe de χ, e deve ser tal que a soma de tôdas as probabilidades possíveis iguale a unidade; isto é, devemos ter

$$K\int_0^\infty e^{-\frac{\chi^2}{2}}\chi^{k'-1}d\chi = 1.$$

Mediante a transformação $\frac{\chi^2}{2} = u$, a integral se reconduz à função Gama, e vem

$$\frac{1}{K} = 2^{\frac{k'-2}{2}}\Gamma\left(\frac{k'}{2}\right).$$

Obtém-se finalmente a distribuição de χ^2

$$F(\chi^2)\,d(\chi^2) = \frac{e^{-\frac{\chi^2}{2}}(\chi^2)^{\frac{k'-2}{2}}\,d(\chi^2)}{2^{\frac{k'}{2}}\,\Gamma\left(\frac{k'}{2}\right)}. \qquad (5.5)$$

5.3 Graus de liberdade.

Supusemos que as freqüências teóricas das diversas classes eram conhecidas *a priori*. Nem sempre essa circunstância se verifica. Ao ajustar uma lei matemática aos dados de observação, os parâmetros são estimados sôbre a amostra colhida, o que importa em forçar a igualdade entre os momentos, ou outros índices estatísticos, das distribuições teórica e empírica.

A restrição (5.2) traduz a igualdade entre as freqüências totais. Se também usamos a média dos dados para determinar uma das cons-

tantes da função interpolatriz, teremos $\Sigma a_i m_i = n \Sigma a_i p_i$ sendo a_i o atributo relativo à classe i. Essa relação pode-se escrever

$$\Sigma a_i (m_i - np_i) = 0 \qquad (5.5)$$

Se usamos outros momentos de ordem superior na fixação dos parâmetros, resultarão outras equações dêsse tipo, em que apenas os coeficientes a_i diferirão. Por exemplo, para o desvio padrão temos que substituí-lo por a_i^2. De qualquer modo, essas restrições equivalem a tantas outras equações do tipo (5.6), que são lineares homogêneas em relação às discrepâncias.

Na representação geométrica usada, cada uma dessas equações indica um hiper-plano passando pelo centro comum das esferas. O cálculo do volume dV desta deve-se, pois, restringir à porção situada na intersecção das superfícies definidas por (5.6). Ora, cada uma das equações baixa de uma unidade o número de dimensões do espaço em que se efetua a integração. Se, pois, temos q condições, isto é, se a nossa lei teórica contém q parâmetros, dV passa a ser um elemento de volume num espaço de k-q dimensões. A equação (5.5) prevalece, desde que se ponha $k' = k$-q. Então k' representa os *graus de liberdade* do sistema.

E' êste um ponto capital. PEARSON, ao introduzir a estatística χ^2, observou que apenas k—1 dos desvios eram independentes, e pois o expoente da distribuição devia-se tomar como $k' = k - 1$. Mas essa regra foi aplicada errôneamente para os demais casos, em que se reconstrói o universo, calculando os seus parâmetros a partir da amostra observada. Caberia a R. A. FISHER [2] assinalar, quase um quarto de século depois, êsse êrro de interpretação, e dar a exata distribuição de χ^2, levando em conta os seus graus de liberdade.

5.4 *Aditividade de* χ^2.

Se adicionarmos várias quantidades que têm distribuição χ^2, a soma também terá essa distribuição, com um número de graus de liberdade igual à soma dos graus de liberdade das componentes.

Para prová-lo, consideremos $k_1 + k_2 + \ldots + k_m$ variáveis x_i. As somas dos quadrados das variáveis, isto é,

$$\chi_1^2 = x_1^2 + x_2^2 + \ldots + x_{k_1}^2$$

$$\chi_2^2 = x_{k_1+1}^2 + \ldots\ldots + x_{k_1+k_2}^2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

possuem a distribuição χ^2 com k'_1, k'_2 ... k'_m graus de liberdade. Somando essas quantidades, vem

$$\chi_1^2 + \chi_2^2 + \ldots + \chi_m^2 = x_1^2 + x_2^2 + \ldots + x_{k_1+k_2+\ldots+k_m}^2$$

Mas essa soma possui a distribuição χ^2 com $k' = k_1 + k_2 + \ldots + k_m - m = k'_1 + k'_2 + \ldots + k'_m$ graus de liberdade, de confirmidade com a proposição enunciada.

---

[2] FISHER, R. A., "On the Interpretation of χ^2 for Contingency Tables and the Calculation of P", *Jour. Roy. Stat. Soc.*, vol. 85 (1922), pág. 87.

5.5 Aplicação e tabelas.

A aplicação de χ^2 é imediata a questões dêste gênero: dada uma série de observações, pode ela ser considerada como se originando de um universo de forma especificada, apenas por flutuações de amostragem? Isso importa em calcular a probabilidade de, na hipótese da série ser uma amostra aleatória de certo universo, obtermos discrepâncias entre os valores teóricos e os empíricos conduzindo a um valor de χ^2 igual ou maior que o obtido. Temos assim que calcular

$$P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} F(\chi^2)\, d(\chi^2).$$

Aqui, como só se trata de desvios positivos, o nível de significância de p por cento corresponde ao valor em que a ordenada corta uma área no ramo crescente da distribuição equivalente a p por cento.

O cálculo da integral acima, considerada como função de χ, pode se efetuar por sucessivas integrações por partes. Se k' é par, obtém-se

$$P = e^{-\frac{\chi^2}{2}} \left[1 + \frac{\chi^2}{2} - \frac{1}{2!}\left(\frac{\chi^2}{2}\right)^2 + \ldots + \frac{1}{\left(\frac{k'-2}{2}\right)!}\left(\frac{\chi^2}{2}\right)^{\frac{k'-2}{2}} \right],$$

e se k é ímpar, a fórmula termina por uma integral da função normal,

$$P = \sqrt{\frac{\pi}{2}}\, e^{-\frac{\chi^2}{2}} \left[\chi + \frac{\chi^3}{1.3} + \ldots + \frac{\chi^{k'-2}}{1.3.5 \ldots (k'-2)} \right] + \sqrt{\frac{2}{\pi}} \int_{\chi}^{\infty} e^{-\frac{\chi^2}{2}}\, d\chi.$$

Os valores dessas integrais foram calculados por ELDERTON e encontram-se nas *Tables for Statisticians*,[3] onde n' representa o número de classes k. R. A. FISHER [4] dá os valores de χ^2 correspondentes a certos valores da probabilidade P; a entrada na tabela faz-se de acôrdo com os graus de liberdade k'. Notemos que $F(\chi^2)$ é uma curva pearsoniana do tipo III, e a sua integral é na realidade uma integral Gama Incompleta; no seu cálculo podem-se, pois, usar as tabelas de PEARSON.[5] Aí, o valor de $1 - P$ é $I(u, p)$, onde

$$u = \frac{\chi^2}{\sqrt{2k}} \quad \text{e} \quad p = \frac{k-2}{2}.$$

Para valores elevados de k, isto é, $k' > 30$, pode-se considerar, segundo FISHER, que $\sqrt{2\chi^2}$ se distribua normalmente com média $\sqrt{2k'-1}$ e desvio padrão unitário. Logo, os valores de χ^2 obtém-se entrando na tabela da integral da curva normal com a quantidade

$$(2\chi^2)^{\frac{1}{2}} - (2k'-1)^{\frac{1}{2}}$$

---

[3] PEARSN, K., *Tables for Statisticians and Biometricians* (3.º ed., Londres, 1930), 1.º vol., tabela XII.

[4] FISHR, R. A., *Statistical Methods for Research Workers* (8.ª ed., Edimburgo, 1941), tab. III; FISHER, R. A., e YATES, F., *Statistical Tables* (Edimburgo, 1938), tab. IV.

[5] PEARSN, K. (ed.), *Tables of the Incomplete Gamma Function* (Londres, 1922).

Se denotarmos por u_p o ponto da distribuição normal, que corresponde ao seccionamento de uma área à direita de p por cento, a regra de FISHER equivale a tomar $\chi^2_{p,k'} = \frac{1}{2}\left[u_p + \sqrt{2k'-1}\right]^2$. Uma melhor aproximação é dada por WILSON e HILFERTY [6] como

$$\chi^2_{p,k'} = k'\left[1 - \frac{2}{9k'} + u_p\sqrt{\frac{2}{9k'}}\right]^3.$$

Comparemos essas aproximações. Para o nível de significância de 5 por cento e $k' = 30$, a primeira fórmula dá $\chi^2 = 43.487$; a segunda, $\chi^2 = 43.749$. O valor exato é $\chi^2 = 43.773$.

5.6 Verificação de leis empíricas.

Ao verificarmos a hipótese de que a série de observações da variável v provém de um universo de forma funcional $f(v, \theta_1, \theta_2, \ldots \theta_q)$, algumas precauções devem ser observadas na aplicação do teste χ^2.

Primeiramente, nenhuma das classes deve conter poucos itens, porque na dedução dessa distribuição admitimos que os fatoriais pudessem ser substituídos pela aproximação de STIRLING. A freqüência mínima de uma classe deve ser *10*, e, para satisfazer êsse preceito, as classes extremas da série observada, que em geral apresentam diminutas freqüências, são recombinadas em grupos maiores.

Também o número de classes não deve ser muito grande, pois para $k \to \infty$ o desenvolvimento de P em série de potências mostra que $P \to 1$. Uma regra prática é tomar $k < 20$.[7]

Notemos que a região crítica para verificar a hipótese nula é o ramo superior da distribuição; isto é, o nível de significância de p por cento corresponde a uma área no ramo à direita equivalente a essa percentagem. Em geral, admite-se que valores de χ^2 superiores ao nível de $p = .05$ indicam uma discrepância real. Contudo, valores excepcionalmente elevados de χ^2 devem ser considerados com suspeição, por denotarem, seja o uso de fórmulas inadequadas, seja o fato da amostra não ter sido colhida estritamente ao acaso.

Exemplo. Numa experiência de cruzamento com duas variedades de milho, obtiveram-se grãos com as freqüências de coloração constantes da tabela abaixo. Verificar se a segregação corresponde à proporção teórica 9:3:3:1.

FREQUENCIA	Amarelo	Alaranj.	Avermel.	Vermelho
Observada f_o	480	155	162	51
Teórica f_t	477	159	159	53
$\frac{(f_o - f_t)}{f_t}$	.0189	.1006	.0566	.0755

[6] WILSON, E. B., e HILFERTY, M. M., *Proc. Nat. Acad. of Sciences*, vol. 17 (1931), pág. 68.

[7] Sôbre o número e intervalo "natural" para as classes, veja-se MANN, H. B. e WALD, A., "On the choice of the number of class intervals in the Chi Square Test", *Ann. Mthem. Stat.*, vol. 13 (1942), pág. 306; GUMBEL, E. J., "On the reliability of the classical Chi-quare Test". *Ann. Mathem. Stat.*, vol. 14 (1943), pág. 253.

Temos $\chi^2 = .2516$, com $k' = 3$. A tabela mostra que desvios como estes verificam-se em mais de 95 por cento dos casos, e portanto a distribuição teórica corresponde à mendeliana.

5.7 *Teste de independência. Tabelas de contingência.*

A aplicação do teste chi-quadrado estende-se imediatamente ao caso em que se quer verificar a independência, ou não, de dois, ou mais princípios de classificação. Suponhamos que, num mesmo grupo de indivíduos, observam-se simultâneamente duas espécies de atributos e registram-se as respectivas freqüências; obtém-se uma *tabela de contingência.*

Seja f_{ij} a freqüência observada na célula ij, f_i a soma das freqüências na linha i, f_j a soma das freqüências na coluna j N o número total de elementos. A probabilidade de que um elemento qualquer caia na linha i é $\frac{f_i}{N}$, e que caia na coluna j é $\frac{f_j}{N}$; se os dois atributos são independentes, a probabilidade de que caia na célula ij é $\frac{f_i f_j}{N^2}$, e a freqüência teórica correspondente $\frac{f_i f_j}{N}$. Podemos agora comparar as discrepâncias entre as freqüências observadas e teóricas mediante o teste χ^2, calculando-se

$$\chi^2 = \Sigma \frac{\left(f_{ij} - \frac{f_i f_j}{N}\right)^2}{\frac{f_i f_j}{N}}. \qquad (5.7)$$

No caso vertente, as probabilidades correspondentes às diversas células não são conhecidas *a priori*, mas resultam das freqüências marginais da tabela; isto é, o cálculo efetuado força os totais marginais dos valores teóricos e empíricos a igualarem-se, o que diminui o número de desvios independentes. Se a tabela fôr de $l \times h$ elementos, apenas $(l-1)\ (h-1)$ são independentes. Logo, na interpretação de χ^2, a entrada na tabela far-se-á com um número de graus de liberdade igual a $k' = (l-1)\ (h-1)$.

5.8 *Coeficiente de contingência.*

O teste χ^2 permite verificar a associação entre os atributos observados, mas não nos dá uma medida da intensidade da mesma.

Como χ^2 mede a dispersão global de N indivíduos, relativamente aos valores teóricos, PEARSON toma a dispersão média por indivíduo $\phi^2 = \frac{\chi^2}{N}$, que denomina contingência quadrática média. Para uma distribuição normal de correlação a duas variáveis, subdividida em classes de amplitude infinitamente pequena, PEARSON estabeleceu as relações

$$\phi^2 = \rho^2 (1 - \rho^2) \quad , \quad \rho^2 = \phi^2 / (1 + \phi^2),$$

sendo ρ o coeficiente de correlação do universo, pelo que, por analogia, definiu o coeficiente de contingência quadrático médio como $C^2 = \phi^2/(1 + \phi^2)$. Evidentemente

$$C^2 = \frac{\chi^2}{N + \chi^2}.$$

5.9 Relação entre os métodos de LEXIS e χ^2.

Observa R. A. FISHER [8] que a descoberta de χ^2 na realidade completa o método de LEXIS. Com efeito, êste último não permitia distinguir, quando se obtinha para uma série empírica o critério $L \leqq 1$, se a desigualdade era geral ou devida apenas a flutuações de amostragem, o que se consegue calculando o χ^2.

Se temos uma série de freqüências f_i $(i = 1, 2, \ldots k)$, observado em amostras de n elementos derivadas de uma população em que a probabilidade do acontecimentos favorável presume-se ser uma constante p, então o critério de LEXIS é, notando que a média verdadeira é np,

$$L^2 = \frac{\frac{1}{k}\sum_{1}^{k}(f_i - np)^2}{npq}$$

Por outro lado, a hipótese de uma probabilidade constante pode ser verificada pelo método de χ^2, observando que então a expectância de f_i é np, e a de $n - f_i$ é $n - np$, donde

$$\chi^2 = \sum_{i=1}^{k}\left[\frac{(f_i - np)^2}{np} + \frac{\left\{(n - f_i) - (n - np)\right\}^2}{n - np}\right],$$

com k graus de liberdade. Ora, temos $p + q = 1$, e a expressão reduz-se a

$$\chi^2 = \frac{\sum_{1}^{k}(f_i - np)^2}{npq}, \qquad (5.8)$$

que é o índice de dispersão baseado na distribuição binomial. Segue-se imediatamente que o critério de LEXIS L^2 equivale a $\frac{\chi^2}{k}$.

Se a probabilidade p fôr estimada a partir dos dados observados, $p = \frac{\Sigma f_i}{nk}$, o número de graus de liberdade será $k - 1$. Neste caso, seria preferível modificar L^2, substituindo o fator k por $k - 1$, de modo que a sua expectância permanecesse igual à unidade.

O índice de dispersão para a distribuição exponencial de POISSON será

$$\chi^2 = \frac{\sum_{1}^{k}(f_i - np)^2}{np}, \qquad (5.9)$$

pois que então a variância é igual à média.

Exemplo. No § 4.4 Ex., vimos que o critério de LEXIS para a série de coeficientes de masculinidade no Distrito Federal era $L = 2.914$ para 48 observações. Temos portanto $\chi^2 = 48\ (2.914) = 407.96$, com $k' = 47$. Não abrangendo as tabelas êsse valor, vamos reduzir χ^2 a desvios estandardizados da curva normal pela fórmula $(2\chi^2)^{1/2} - (2k' - 1)^{1/2} = 28.56 - 9.64 = 18.92$.

A probabilidade de desvios dessa ordem é infinitésima.

---

[8] FISHER, R. A., *Statistical Methods for Research Workers*, 8.ª ed., pág. 78.

5.10 Significância e limites fiduciais de uma proporção.

O teste χ^2 aplica-se de pronto à medida da significância de uma proporção. Suponhamos que se tenha apenas uma alternativa, isto é, que os indivíduos se achem classificados segundo a presença ou ausência de certo atributo. A fórmula (5.8) nos dá

$$\chi^2 = \frac{(f - np)^2}{npq} = \frac{\left(\frac{f}{n} - p\right)^2}{\frac{pq}{n}}. \qquad (5.9)$$

Comumenta a significância de uma proporção determina-se comparando a diferença entre ela e o valor teórico com o seu êrro padrão (§ 3.8); o valor de χ^2 que obtivemos é exatamente o quadrado dessa relação. Por conseguinte, as tabelas de χ^2 podem ser utilizadas para determinar a significância de uma proporção em vez das da curva normal.

Aliás, era êsse um resultado de se esperar. Se fizermos $k' = 1$ na distribuição de χ^2, ela se reduz a $(2\pi)^{-\frac{1}{2}} e^{-\frac{\chi^2}{2}} d\chi$, que é a distribuição normal.

O processo é interessante porque permite uma estima mais rigorosa dos limites fiduciais de p.[9] E' claro que o sentido das diferenças entre os valores observados e os teóricos não importa, havendo em todos os casos contribuição para o aumento de χ^2 e diminuição de sua probabilidade P. Se quisermos os limites correspondentes ao nível de 95%, podemos pôr na fórmula acima χ^2 igual ao valor correspondente a $P = .05$, seja χ^2_0. Tem-se então $npq\,\chi^2_0 = (f - np)^2$, ou seja,

$$p^2(\chi_0^2 + n) - p(\chi_0^2 + 2f) + \frac{f^2}{n} = 0, \qquad (5.11)$$

equação quadrática, cujas raízes dão os limites fiduciais p' e p''. Notar-se-á que êsses limites, em contraste com o processo aproximado do § 4.3, não são equidistantes em relação à percentagem observada.

Exemplo. Numa experiência sôbre a duração de dormentes tratados com creosoto, verificou-se que, após 20 anos de uso, 22 dormentes de um lote de 50 estavam ainda em boas condições. enquanto anteriores experiências com dormentes não tratados indicavam que apenas 27% deveriam subsistir. E' eficiente o tratamento pelo creosoto ? Entre que limites pode-se esperar, com uma probabilidade de 95%, que oscilará a percentagem de aproveitamento dos dormentes tratados ?

Temos $f/n = .44$, $p = .27$, e portanto $\chi^2 = \frac{50\,(.44 - .27)^2}{(.27)\,(.73)} = 7.34$.

Para $k' = 1$ grau de liberdade, a probabilidade dêsse χ^2 é menor que 0.01. A proporção de dormentes tratados que subsistem é significantemente maior que a dos não tratados; logo a aplicação do creosoto é vantajosa.

Para estabelecer os limites fiduciais da percentagem de aproveitamento devemos entrar na equação (5.11) com o valor de $\chi_0^2 = 3.841$, correspondente ao nível de 0.05. Obtém-se $53.84\,p^2 - 47.84\,p + 9.68 = 0$, que, resolvida, nos dá as duas raízes $p' = .312$ e $p'' = .577$, que não são simétricas em relação à percentagem encontrada de 0.44.

---

[9] Outro método exato deve-se a CLOPPER, C. J., e PEARSON, E. S., "The Use of Confidence or Fiducial Limits in the Case of the Binomial", *Biometrika*, vol. 26, pág. 404.

O processo habitual (§ 3.8) levaria, notando que no caso $s_p = \sqrt{\frac{(.44)(.56)}{50}} = .07$, aos limites fiduciais $p' = .44 - (1.96 \times .07) = .303$ e $p'' = .577$.

5.11 Teste de homogeneidade.

O cálculo do índice de dispersão (§ 5.9) importa, na realidade, em verificar a homogeneidade dos dados em relação à proporção teórica p. Êle também equivale ao cálculo de χ^2 baseado numa tabela de contingência de $2 \times k$ elementos. Podemos generalizar o método, tornando-o aplicável aos casos em que as amostras sucessivas não têm o mesmo tamanho.

Seja f_i a freqüência do atributo observado na amostra i, de tamanho n_i. Se considerarmos ùnicamente a proporção resultante do conjunto global de observações, $p = \frac{\Sigma f_i}{\Sigma n_i}$, pode-se ter a impressão de uma menor variabilidade do que a realmente existente entre as diferentes amostras. Imagine-se, por exemplo, que os valores de χ^2 para cada uma delas situam-se próximo ao nível de significância, mas que as razões f_i/n_i variam num sentido e outro em tôrno do valor teórico p. Ao totalizarmos êsses valores, as diferenças tendem a se cancelar, e daí resulta para a proporção global um χ^2 correspondendo a uma probabilidade exagerada.

Para obviar êsse inconveniente, apoiar-nos-emos na propriedade aditiva de χ^2. Os valores de χ^2 relativos a cada uma das amostras são adicionados, e a probabilidade de $\chi^2 = \chi_1^2 + \chi_2^2 + \ldots$ resultante é obtida da tabela com um número de graus de liberdade igual ao número de amostras.

Exemplo. MENDEL observou o número de grãos redondos e angulosos obtidos numa hibridização de ervilhas, obtendo os resultados constantes da tabela abaixo. Verificar se os resultados correspondem à expectação mendeliana de $p = .75$.

PLANTA	Num. grãos n	Grãos red. f	$\frac{f}{n}$	χ^2
A	57	45	.7895	.47
B	29	19	.6552	1.39
C	32	26	.8125	.67
D	32	22	.6875	.67
E	34	28	.8235	.98
Total	184	140	.7609	

A percentagem observada sôbre o resultado global $p = .7609$ corresponde a $\chi^2 = .098$ com 1 grau de liberdade, ou seja a uma probabilidade de desvios maiores de $P = .76$, confirmando-se a hipótese. Nenhum dos resultados parciais a infirma, pois χ^2 para o nível de significância de 5% é 3.841. Contudo, baseando-nos nesses χ^2 parciais, que constam da última coluna, obtemos para o experimento total $\chi^2 = 4.18$, com 5 graus de liberdade, o que corresponde a apenas $P = .51$.

5.12 Tabelas de contingência 2x2. Significância da diferença entre proporções.

O caso particular de tabelas de contingência relativas a dois atributos, ambos classificados segundo a alternativa de sua presença-ausência, é de grande importância prática. Então, há apenas um grau de liberdade; isto é, basta calcular a diferença entre o vàlor teórico e o observado em uma das células, as discrepâncias nas demais obtendo-se somando ou subtraindo aquela diferença das freqüências observadas. Ou então, designando por a, b, c, d as freqüências celulares e N a total, obtém-se χ^2 diretamente pela fórmula

$$\chi^2 = \frac{N\,(ad - bc)^2}{(a + b)\,(c + d)\,(a + c)\,(b + d)}. \qquad (5.12)$$

A êsse caso se reconduz o teste de significância da diferença entre duas proporções. Basta convertê-las em freqüências e aplicar a fórmula acima. Obtém-se resultados idênticos aos expostos no § 3.10, em virtude da redução da distribuição de χ^2 à normal neste caso.

5.13 Correção de continuidade de YATES.

A distribuição χ^2 é uma função contínua enquanto que a distribuição de freqüências numa tabela de contingência é sempre descontínua. Os testes baseados em χ^2 são, pois, aproximados, e supõe-se que a amostra seja suficientemente grande para que se possa admitir a continuidade entre as freqüências nas diversas células.

Uma correção para essa falta de continuidade foi estabelecida por YATES [10] para o caso de uma tabela de 2x2 elementos. Neste caso, χ^2 pode ser calculado a partir da freqüência observada numa só célula e os totais marginais. Suponhamos que aquela freqüência seja muito pequena; a probabilidade de termos desvios menores é, na realidade, a soma de um número limitado de probabilidades finitas, enquanto que χ^2 corresponde à fração do ramo superior de uma curva contínua. O cálculo da integral deve-se estender, pois, até o limite da classe cujo ponto médio é a freqüência observada; isto é, devemos acrescer ½ unidade a essa freqüência para obter o limite de integração. Feita essa retificação, as freqüências das demais células são corrigidas de modo a manterem-se constantes os totais marginais. Noutros têrmos, a correção de YATES importa em calcular χ^2, não sôbre os valores observados, mas sôbre aquêles que se obteriam se as freqüências fôssem menos extremas do que são de meia unidade.

Isso equivale a substituir à fórmula (5.12) esta outra

$$\chi_c^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a + b)\,(c + d)\,(a + c)\,(b + d)}, \qquad (5.13)$$

sendo que $N/2$ sempre reduz o valor numérico de $|ad - bc|$.

Para tabelas 2 x 3 a correção de continuidade já não é tão importante; a elevação do número de graus de liberdade também aumenta as combinações possíveis, e melhora a aderência à distribuição contínua.

---

[10] YATES, F., "Contingency Tables involving small numbers and the χ^2 test", *Supplement Jour. Roy. Stat. Soc.*, vol. 1 (1934), pág. 217; IRWIN, J. O., "Tests of significance for differences between percentages based on small numbers", *Metron*, vol. 12 (1935), pág. 83.

Quando o menor dos valores cai abaixo de 10, e preferìvelmente até 100, deve-se também levar em consideração a assimetria da distribuição resultante. FISHER e YATES [11] dão tabelas contendo os valores de χ^2 para os níveis de 5 a 1% em cada um dos ramos da curva.

Em casos de dúvida, pode-se recorrer ao cálculo exato da probabilidade de se obter um qualquer conjunto de valores celulares *a, b, c, d.* Demonstra-se [12] que essa probabilidade é

$$\frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{N!\,a!\,b!\,c!\,d!}.$$

Se *a* representa a menor das freqüências, a probabilidade de termos a figuração observada ou outras mais extremas é dada pela soma das probabilidades calculadas como acima, em que se supõe a freqüência *a* decrescer dêsse valor até 0, mantendo-se constantes os totais marginais.

Exemplo. Num inquérito sôbre a prática habitual de atos religiosos, obtiveram-se as respostas constantes do quadro abaixo. Verificar se a religiosidade difere significantemente entre os dois sexos.

(i) Freq. observadas

SEXO	Praticam	Não prat.	Total
Masculino.....	68	32	100
Feminino.....	52	48	100
Total.....	120	80	200

(ii) Freq. corrigidas

SEXO	Praticam	Não prat.	Total
Masculino.....	67.5	32.5	100
Feminino.....	52.5	47.5	100
Total.....	120	80	200

Calculando χ^2 diretamente sôbre os valores tabulares, teremos, para a primeira tabela

$$\chi^2 = \frac{[(68 \times 48) - (32 \times 52)]^2 \times 200}{120 \times 80 \times 100 \times 100} = 5.33,$$

a que corresponde, com 1 grau de liberdade, a probabilidade $P = .022$. A diferença de prática religiosa é, pois, significante.

Introduzindo a correção de YATES na tabela (ii), teremos $\chi_c^2 = 4.688$, donde $P = .034$. Assim, a falta da correção de continuidade sobrestima a significância da diferença.

Podiamos ter utilizado a tabela de FISHER e YATES. Calcularíamos $\chi_c = 2.165$, a menor expectância $m = \frac{100 \times 80}{200} = 40$ e a razão $p = \frac{40}{80} = .50$. Como a menor freqüência observada, 32, está abaixo de m, devemos utilizar o ramo inferior da distribuição e a tabela nos dá o valor crítico 1.96. O calculado é superior a êsse valor, confirmando-se a significância da variação de religiosidade com o sexo.

5.14 Combinação de probabilidades de testes de significância.

Outro uso de chi-quadrado, indicado por R. A. FISHER, refere-se ao caso de termos vários testes de significância independentes, dos quais apenas poucos ou nenhum é individualmente significante, mas que, em conjunto, dão a impressão de significância, que se quer verificar.

---

[11] FISHER, R. A., e YATES, F., *Statistical Tables*, tab. VIII.

[12] FISHER, R. A., *Statistical Methods for Research Workers*, 8.ª ed., pág. 95.

Sejam [13] $p_1, p_2, \ldots p_k$ as probabilidades dos diferentes testes, não sendo necessário que êles se refiram à mesma estatística para as várias amostras. Como a distribuição de p não depende de uma particular distribuição da estatística, da qual êle é a integral de probabilidade, podemos admitir que p se possa obter como a integral de probabilidade de uma qualquer distribuição contínua. FISHER admite que esta seja a distribuição χ^2 com 2 graus de liberdade.

Temos então que a integral de probabilidade é

$$\int_{\chi}^{\infty} e^{-\frac{1}{2}\chi^2} \chi \, d\chi = e^{-\frac{1}{2}\chi^2}.$$

Se $p = e^{-\frac{1}{2}\chi^2}$, vemos que $-2 \log_e p$ segue a mesma distribuição que χ^2 com 2 graus de liberdade, e

$$\log_e [p] = \log_e p_1 + \log_e p_2 + \ldots + \log_e p_k = -\frac{1}{2}(\chi_1^2 + \chi_2^2 + \ldots + \chi_k^2).$$

Em vista da propriedade aditiva de χ^2, a expressão entre parênteses segue esta distribuição com $2k$ graus de liberdade.

Por conseguinte, as probabilidades de k diferentes testes podem ser combinadas num teste de conjunto, cuja significância será verificada mediante a tabela de χ^2 com $2k$ graus de liberdade.

Exemplo. Numa experiência sôbre novo método de ensino, 3 turmas ficaram a cargo de professôres diferentes, que, medindo o aproveitamento por testes diversos, obtiveram, para probabilidade de uma diferença igual ou maior, relativamente ao método ordinário de ensino, os valores de *0.09, 0.12* e *0.06* respectivamente. Que conclusões se podem firmar a respeito do novo método?

Evidentemente, êsses resultados não podem ser englobados num experimento único, por terem sido diferentes os testes de aproveitamento usados.

Nenhuma das turmas revelou uma diferença significante, embora tôdas confirmassem a superioridade do novo método. Usando o processo descrito, teriamos

$$\begin{aligned} \log_e .09 &= \bar{3}.592 \\ \log_e .12 &= \bar{3}.888 \\ \log_e .06 &= \bar{3}.187 \\ \log_e &= \bar{8}.667 = -7.333, \end{aligned}$$

donde $\chi^2 = 14.666$ com 6 graus de liberdade. A tabela mostra que então $P < .05$, de modo que o resultado coletivo é significante.

[13] Cf. NAIR, K. R., "A note on the exact distribution of λ_n", *Sankyā-The Indiam Journ. Stat.*, vol. 3 (1937), pág. 171.

Capítulo VI

A DISTRIBUIÇÃO DE STUDENT

6.1 *A significância da média.*

O processo habitual de determinação da significância da média, que expusemos no Cap. II, pressupõe o conhecimento da variância do universo. Na realidade, raramente êste elemento é dado *a priori*, e antes cumpre calculá-lo a partir da amostra.

Se dispomos de N observações $x_1, x_2, \ldots x_N$, a melhor estimativa da variância do universo é, como veremos adiante, $s^2 = \frac{\sum_1^N (x_i - \bar{x})^2}{N-1}$.

Então, s será o êrro padrão de uma observação, e $\frac{s}{\sqrt{N}}$ o êrro padrão da média. Para verificar se a média da amostra $\bar{x}$ difere significantemente da média μ do universo, utiliza-se a estatística

$$t = \frac{(\bar{x} - \mu)\sqrt{N}}{s}. \qquad (6.1)$$

Durante mais de um século, admitiu-se que esta estatística tinha distribuição normal, e a avaliação da probabilidade de um determinado desvio fazia-se mediante a tabela das áreas da curva normal. E' esta uma conclusão errônea, pois tanto o numerador como o denominador são funções das observações, e pois estão sujeitos a erros de amostragem. Só em 1908 W. S. Gosset, mais conhecido sob o pseudônimo de "Student" [1], assinalou essa circunstância e determinou, embora empìricamente, a verdadeira distribuição de t. A dedução rigorosa da forma analítica dessa distribuição foi obtida em 1925 por R. A. Fisher.[2] Na realidade, Student lidou com a variável z, definida pela relação $z = t\,(N\text{-}1)^{-\frac{1}{2}}$

6.2 *A distribuição de* Student.

Consideremos um universo caracterizado pela distribuição normal $df = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$. Se dele colhermos uma amostra aleatória de N elementos $(x_1, x_2, \ldots x_N)$, a probabilidade de que esta amostra caia no elemento de volume $dv = dx_1\, dx_2 \ldots dx_N$ é

$$dF = (2\pi\sigma^2)^{-\frac{N}{2}} e^{-\frac{1}{2}\sum_1^N \left(\frac{x_i-\mu}{\sigma}\right)^2} dv. \qquad (6.2)$$

Em vista da identidade $\Sigma (x_i - \mu)^2 = \Sigma(x_i - \bar{x})^2 + N(\bar{x} - \mu)^2$, a diferencial dF se reduz a

$$dF = (2\pi\sigma^2)^{-\frac{N}{2}} e^{-\frac{1}{2\sigma^2}[N(\bar{x}-\mu)^2 + (N-1)s^2]} dv. \qquad (6.3)$$

---

[1] Student, "The probable error of a mean", *Biometrika*, vol. 6 (1908), pág. 1.

[2] Fisher, R. A., "Applications of "Student" 's distribution", *Metron*, vol. 5 (1925), pág. 90.

6.3 Independência das distribuições de $\bar{x}$ *e* s^2.

Escrevamos a equação (6.3) sob a forma

$$dF = k_1 \; e^{-\frac{N}{2\sigma^2}(\bar{x}-\mu)^2} \; d\bar{x} \times k_2 \; e^{-\frac{(N-1)s^2}{2\sigma^2}} \; (s^2)^{\frac{N-3}{2}} \; d(s^2) \; . \qquad (6.8)$$

O primeiro fator é uma função apenas de $\bar{x}$, $G(\bar{x})$, e o segundo apenas de s^2, $H(s^2)$; isto é, $F(\bar{x}, s^2) = G(\bar{x})\, H(s^2)$. Quer isso dizer que a média e a variância se distribuem independentemente, no sentido do cálculo de probabilidades. GEARY [3] provou que a condição necessária e suficiente para que essas duas estatísticas tenham distribuições independentes é que a variável x se distribua normalmente no universo a que pertence a amostra.

$G(\bar{x})$ nos dá a distribuição por amostragem de médias de um universo normal. Pela condição da área $\int_{-\infty}^{\infty} G(\bar{x})\, d\bar{x} = 1$, e vem $k_1 = \left(\frac{2\pi\sigma^2}{N}\right)^{-\frac{1}{2}}$, donde

$$G(\bar{x}) = \left(\frac{2\pi\sigma^2}{N}\right)^{-\frac{1}{2}} \; e^{-\frac{N(\bar{x}-\mu)^2}{2\sigma^2}} \; , \qquad (6.9)$$

que é uma distribuição normal com média igual a μ e desvio padrão igual a $\sigma/\sqrt{N}$, resultado já conhecido.

A distribuição da variância

$$H(s^2) = k_2 \; e^{-\frac{(N-1)s^2}{2\sigma^2}} \; (s^2)^{\frac{N-3}{2}} \; d(s^2) \, , \qquad (6.10)$$

será estudada no capítulo seguinte.

6.4 Propriedades e tabelas da distribuição de t.

A propriedade fundamental da distribuição de t, como também da de z, é que não depende da variância σ^2 do universo, e pode, pois, ser usada quando esta grandeza é desconhecida. Isso é importante quando se lida com pequenas amostras, devido aos erros na estima de σ^2.

A equação da curva de t mostra que ela é simétrica em tôrno de $t = 0$, a que corresponde a ordenada máxima; quando $t \to \pm \infty$, a curva torna-se asimptótica ao eixo das abscissas. A distribuição de t assemelha-se à normal, da qual se aproxima para valores crescentes de n. Contudo, para um dado n, ela é platykúrtica, e, portanto, para valores suficientemente grandes de n, as suas ordenadas sobrepassam as da curva normal. Isso implica que a prática habitual de calcular a probabilidade dos desvios mediante a curva normal erra no sentido de exagerar a significância dos grandes desvios, imputando-lhe probabilidades escassas.

Para verificar a normalização da curva de t, escrevamo-la sob a forma

$$F(t) = K\left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \, ,$$

---

[3] GEARY, R. C., "The Distribution of "Student's" Ratio for Non-Normal Samples", *Supplement Jour. Roy. Stat. Soc.*, vol. 3 (1936), pág. 90.

onde $K = \dfrac{\Gamma(n)}{2^{n-1}\, n^{\frac{1}{2}} \left[\Gamma\left(\frac{n}{2}\right)\right]^2}$. Tomando os logarítmos de ambos os têrmos, vem

$$\log_e F(t) = \log_e K - \frac{n+1}{2} \log_e \left(1 + \frac{t^2}{n}\right).$$

Desenvolvendo o último têrmo em série, segundo as potências de $\frac{t^2}{n}$, a qual converge uniformemente quando $n > t^2$, obtemos

$$-\frac{n+1}{2} \log_e \left(1 + \frac{t^2}{n}\right) = -\frac{t^2}{n} + \ldots$$

Por outro lado, substituindo a $\log_e \Gamma(n)$ e $\log_e \Gamma\left(\frac{n}{2}\right)$ os seus desenvolvimentos em série de STIRLING,[4] resulta, após certas reduções,

$$\log_e F(t) = -\frac{1}{2} \log_e 2\pi - \frac{z^2}{2} + \frac{z^4 - 2z^2 - 1}{4n} + \ldots$$

Daí se conclui $\lim_{n \to \infty} \log_e F(t) = -\frac{1}{2} \log_e 2\pi - \frac{z^2}{2}$, e, pela continuidade da função exponencial, $\lim_{n \to \infty} F(t) = (2\pi)^{-\frac{1}{2}}\, e^{-\frac{t^2}{2}}$, que é a função normal.

Para avaliar a probabilidade de se ter um desvio igual ou maior que t, devemos calcular a integral $P_t = K \int_t^{\infty} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} dt$.

Pela substituição $tan\, \theta = t\sqrt{n}$, essa integral se reduz à forma $\int_0^{\frac{\pi}{2}} cos^{n-1} \theta\, d\theta$, que se integra por partes, obtendo-se

para n par: $$P_t = \frac{1}{2} - \frac{1}{2} sen\, \theta \left\{1 + \frac{1}{2} cos\, \theta + \frac{1.3}{2.4} cos^4 \theta + \ldots\right\}$$

para n ímpar: $$P_t = \frac{1}{2} - \frac{\theta}{\pi} - \frac{sen\, \theta}{\pi} \left\{cos\, \theta + \frac{2}{3} cos^3 \theta + \frac{2.4}{3.5} cos^5 \theta + \ldots\right\}$$

Como no caso da distribuição de χ^2, o cálculo da integral para n ímpar envolve uma função transcendente, que é aqui uma função circular inversa.

STUDENT [5] publicou extensas tabelas da integral $\int_{-\infty}^{t} F(t)\, dt = 1 - P_t$, com os argumentos t e n. Mais práticas são, contudo, as tabelas de FISHER,[6] que dão os valores de t para os argumentos P_t e n.

---

[4] WHITTAKER, E. T., e WATSON, G. N., *A Course of Modern Analysis* (3.ª ed., Cambridge, 1920), pág. 252.

[5] "STUDENT", "New tables for testing the significance of observations", *Metron*, vol. 5 (1925), pág. 105.

[6] FISHER, R. A., *Statistical Methods for Research Worker's*, tab. IV; FISHER, R. A., e YATES, F., *Statistical Tables*, tab. III.

Para os valores de $n > 30$, STUDENT mostrou que a distribuição de z tende para a forma normal com desvio padrão igual a $(N—3)^{-\frac{1}{2}}$ Daí se conclui que podemos então usar as tabelas da curva normal, tomando $t\left(\frac{n-2}{n}\right)^{\frac{1}{2}}$ como desvio reduzido. DEMING e BIRGE [7] sugerem tomar-se $t\left(\frac{n-\frac{1}{2}}{n}\right)^{\frac{1}{2}}$ que dá melhor aproximação que a anterior para valores centrais $(P_t > .30)$, mas pior para valores extremos. HOTELLING e FRANKEL [8] mostraram que, quando $n \geqq t^2$, é pràticamente suficiente tratar $t\left(1-\frac{t^2+1}{4n}\right)$ como distribuido normalmente, o que dispensa o uso da própria tabela de t. Outra aproximação é devida a FISHER,[9] envolvendo porém o uso de tabelas auxiliares.

Acima de 100, admite-se que t se distribui normalmente com desvio padrão unitário. Para uma avaliação de P_t, pode ser utilizado o nomograma de NEKRASSOFF.[10]

6.5 *Significância e limites fiduciais da média.*

A aplicação da distribuição de STUDENT na verificação da significância da média, ou no estabelecimento de seus limites fiduciais, opera-se nas mesmas linhas expostas relativamente às grandes amostras. Dada a média do universo donde se presume derivar a amostra, calcula-se a estatística t, e, conforme o nível de significância em que caia P_t, é a hipótese confirmada ou rejeitada.

Se se trata de estabelecer limites fiduciais da média, a tabela nos dá, para o nível de significância de p por cento escolhido, o valor t_o satisfazendo a relação $P\ (|t| \geqq t_o) = p_o$, donde se conclui $|x - \mu| \geqq t_o \frac{s}{\sqrt{N}}$, e portanto os limites $\mu' = \bar{x} - t_o \frac{s}{\sqrt{N}}$ e $\mu'' = \bar{x} + t_o \frac{s}{\sqrt{N}}$.

Exemplo A. Retomemos o Ex. 2.9 A, supondo não conhecido o êrro padrão da determinação experimental de densidade. A que resultado se chega?

Temos de estimar s mediante os dados observados, obtendo-se $s = 1.097$. Daí $s_{\bar{x}} = \frac{1.097}{\sqrt{4}} = .548$ e $t = \frac{.25}{.548} = .456$. Reportando-nos à tabela de t com $n = 3$ graus de liberdade, vemos que P_t está compreendido entre 0.60 e 0.70, podendo-se, pois, concluir que a amostra é de ouro.

Exemplo B. Suponhamos que as determinações de densidade tivessem fornecido os valores 19.31, 19.20, 18.90, 18.79, cuja média é ainda 19.05.

---

[7] DEMING, W. E., e BIRGE, R. T., "On the Statistical Theory of Errors", *Reviews of Modern Physics*, vol. 6 (1934), pág. 130.

[8] HOTELLING, H., e FRANKEL, L. R., "The transformation of statistics to simplify their distribution", *Ann. Mathem. Stat.*, vol. 9 (1938), pág. 89.

[9] FISHER, R. A., "Expansion of "STUDENT" 's integral in powers of n^{-1}.", *Metron*, vol. 5 (1925), pág. 109.

[10] NEKRASSOFF, V. A., "Nomography in Applications of Statistics", *Metron*, vol. 8 (1930), pág. 95.

A dispersão é agora muito menor, obtendo-se a estimativa $s = .245$, e portanto $s_{\bar{x}} = .122$. Daí $t = \frac{.25}{.122} = 2.05$, a que corresponde, com 3 graus de liberdade, uma probabilidade $P_t > .10$. A diferença não é significante. Mas, se seguirmos a prática habitual de referir o desvio 2.05 à tabela da curva normal, vamos achar uma probabilidade $P = .04$, indicando uma diferença significante. No primeiro caso a amostra seria considerada de ouro, no segundo não.

6.6 *A distribuição de* t *em universos não-normais.*

A distribuição de STUDENT foi deduzida supondo-se normal a população donde deriva a amostra. Abre-se, pois, a questão de se saber se, no inadimplemento dessa condição, deve-se preferir êsse processo ou o clássico, baseado na distribuição normal. E isso é de grande relevância prática, pois muitas vêzes nada conhecemos sôbre a forma do universo.

Da normalidade do universo resultam as propriedades seguintes, utilizadas na dedução da distribuição de t: (*i*) a distribuição da média é normal; (ii) a da variância é uma curva do tipo III de PEARSON; (iii) $\bar{x}$ e s^2 não são correlacionados. Ora, quando o universo não é normal, se a primeira condição ainda se verifica aproximadamente, o mesmo não acontece com as demais, se o tamanho da amostra é pequeno.

NEYMAN [11] determinou a regressão da variância sôbre a média para amostras derivadas de universos caracterizados pelos parâmetros β_1 e β_2, mostrando que é aproximadamente parabólica, o sinal do têrmo quadrático dependendo de $\beta_2 - \beta_1 - 3$. Se essa expressão é nula, isto é, se β_1 e β_2 são pontos de uma reta no plano $\beta_1 \beta_2$, a regressão é linear. Para os pontos situados acima dessa linha, a parábola de regressão no plano $\bar{x} - \mu$ e s^2 é côncava no sentido positivo do eixo s^2, e viceversa. O efeito dessa regressão sôbre a distribuição de z é óbvio. Para grandes valores de $\bar{x} - \mu$, o valor de s^2 tende a ser menor que o seu valor médio, e $|z|$ maior; logo, haverá um maior adensamento de valores de z além do particular $|z|$ considerado, do que no caso do universo normal. Por outro lado, para valores de $\bar{x} - \mu$ próximos de zero há rarefação.

Tais efeitos foram verificados teòricamente por RIDER, e experimentalmente por SHEWHART e WINTERS, e NEYMAN e PEARSON, para amostras de 4 elementos extraídos de universos retangulares e triangulares. Estes estatísticos mostraram ainda que uma assimetria positiva do universo provoca uma assimetria negativa da distribuição de z, e inversamente. Outros estudos são devidos a BARTLETT, GEARY, RIETZ, e outros, e levam à conclusão de que a aplicação da teoria de STUDENT dá melhor resultado que a teoria clássica para grande número de distribuições não-normais, mas que há falhas, atribuíveis sobretudo ao grau e à natureza da correlação enter $\bar{x} - \mu$ e s^2. De qualquer modo, obtém-se melhores resultados quando se calcula a probabilidade de que t caia no intervalo $-t_o$ a $+t_o$ do que quando essa probabilidade se refere ao intervalo t_o a ∞. No primeiro caso, há uma compensação das deficiências dos valores de t no intervalo $-\infty$ a t_o contra um excesso no intervalo t_o a $+\infty$.

---

[11] NEYMAN, J., "On the correlation of the mean and the variance in samples from an "infinite" population", *Biometrika*, vol. 18 (1926), pág. 401.

6.7 Generalização do uso da distribuição de STUDENT.

Reportando-nos ao exposto no § 6.3, notemos que R. A. FISHER [12] assinalou que a distribuição de STUDENT compõe-se das duas distribuições independentes: (i) de $\frac{(\bar{x}-\mu)\sqrt{N}}{\sigma}$, distribuída normalmente com média igual a zero e variância unitária; (ii) de $\frac{(N-1)\,s^2}{\sigma^2} = \chi^2$, distribuída [13] segundo $k\,e^{-\frac{1}{2}\chi^2}\left(\frac{\chi^2}{2}\right)^{\frac{N-3}{2}} d(\chi^2)$, de modo que

$$t = \frac{(\bar{x}-\mu)\sqrt{N}}{\sigma} \div \left(\frac{\chi^2}{N-1}\right)^{1/2}. \qquad (6.11)$$

Isso mostra que o emprêgo da distribuição de STUDENT pode-se generalizar a todos os casos que importam na comparação de uma variável distribuída normalmente com uma estimativa de seu êrro padrão de distribuição independente, sendo o número de graus de liberdade igual ao utilizado nessa estimativa.

Dêste modo, poude-se estender o uso dessa distribuição a problemas referentes à significância da diferença entre médias e à verificação de hipóteses relativas aos coeficientes de regressão.

6.8 Significância da diferença entre médias.

Para determinar a significância da diferença entre duas médias, verificamos a hipótese de que elas provenham da mesma população normal. Sejam $\bar{x}_1$ e $\bar{x}_2$ as médias, e s_1 e s_2 os desvios padrões de duas amostras de tamanhos N_1 e N_2 respectivamente. Supondo-as não correlacionadas, a variância da diferença entre essas médias será $\sigma^2\left(\frac{1}{N_1}+\frac{1}{N_2}\right) = \sigma^2\frac{N_1+N_2}{N_1N_2}$, e a estatística

$$\frac{\bar{x}_1-\bar{x}_2}{\sigma}\left(\frac{N_1N_2}{N_1+N_2}\right)^{\frac{1}{2}}$$

distribui-se normalmente com desvio padrão unitário. Como não conhecemos σ, vamos estimá-lo combinando as dispersões de ambas as amostras, isto é, tomando

$$s^2 = \frac{\sum_1^{N_1}(x_1-\bar{x}_1)^2 + \sum_1^{N_2}(x_2-\bar{x}_2)^2}{N_1+N_2-2} = \frac{n_1s_1^2+n_2s_2^2}{n_1+n_2},$$

que tem a distribuição χ^2 com $N_1+N_2-2 = n_1+n_2$ graus de liberdade. Conseqüentemente,

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sigma}\left(\frac{N_1N_2}{N_1+N_2}\right)^{\frac{1}{2}} \div \frac{s}{\sigma} = (\bar{x}_1-\bar{x}_2)\left[\frac{N_1N_2(n_1+n_2)}{(N_1+N_2)(n_1s_1^2+n_2s_2^2)}\right]^{\frac{1}{2}} \qquad (6.12)$$

terá a distribuição de STUDENT com n_1+n_2 graus de liberdade.

---

[12] FISHER, R. A., "Applications of "STUDENT" 's distribution", *Metron*, vol. 5 (1925), pág. 94.

[13] Essa grandeza que não é o mesmo χ^2 usado nos testes de aderência de ajustamentos, tem porém a mesma distribuição.

Se ambas as amostras são do mesmo tamanho, $n_1 = n_2$, a fórmula simplifica-se para

$$t = (\bar{x}_1 - \bar{x}_2)\left(\frac{N}{s_1^2 + s_2^2}\right)^{\frac{1}{2}} \qquad (6.13)$$

com $n = 2\,(N-1)$ graus de liberdade.

Se se trata de grandes amostras, pode-se pôr aproximadamente $n_1 = N_1$ e $n_2 = N_2$, e obtém-se a fórmula habitual

$$t = (\bar{x}_1 - \bar{x}_2)\left(\frac{N_1 N_2}{N_1 s_1^2 + N_2 s_2^2}\right)^{\frac{1}{2}}. \qquad (6.15)$$

Como t tem distribuição assimptótica normal, justifica-se o emprêgo das tabelas da curva normal para grandes amostras.

Exemplo. (WISHART) A fim de confrontar dois tipos de segadeiras, um trigal foi dividido em secções longitudinais, e cada duas secções contíguas tratadas pelas duas máquinas. Obteve-se a produção constante da tabela abaixo, e pede-se verificar a signicância das médias.

Secção	Máquina A	Máquina B	D
1	8.0	5.6	3.5
2	8.4	7.4	1.0
3	8.0	7.3	0.7
4	6.4	6.4	0.0
5	8.6	7.5	1.1
6	7.7	6.1	1.6
7	7.7	6.6	1.1
8	5.6	6.0	— 0.4
9	5.6	5.5	0.1
10	6.2	5.5	0.7
Total	72.2	63.9	8.3
Média	7.22	6.39	0.83
$\Sigma(x-\bar{x})^2$	11.936	5.569	6.001

Temos primeiramente que estimar a variância comum, que, no caso, é $s^2 = \dfrac{11.936 + 5.569}{10 + 10 - 2} = .9725$, $s = .9862$. Pôsto isto, temos $t = \dfrac{7.22 - 6.39}{.9862}\sqrt{5} = 1.89$. Entrando na tabela de t com $n = 18$ graus de liberdade, vemos que não atinge o nível de 5%, e pois não há superioridade significante de uma segadeira sôbre outra.

6.9 *Diferença entre medidas emparelhadas.*

A fim de diminuir a variabilidade das medidas, e assim aumentar a precisão dos experimentos, podemos escolher os casos aos *pares*, de modo que ambos os membros de cada par estejam sujeitos às mesmas influências perturbadoras. Por exemplo, ao estudar a eficiência de dois métodos de ensino, podemos formar as turmas com alunos que tenham obtido o mesmo escore num teste de inteligência; às vêzes, comparamos as notas obtidas pelos mesmos alunos antes e depois da introdução de novo sistema.

Nesses casos, calculamos as diferenças $d = x_1 - x_2$ entre as medidas de cada par, e verificamos se a média da distribuição dessas diferenças, isto é, a diferença média $\bar{d}$, difere significantemente de zero. Temos

$$t = \bar{d} \left[\frac{N\,(N-1)}{\Sigma\,(d-\bar{d})^2} \right]^{\frac{1}{2}}. \qquad (6.15)$$

A estima da variância de d baseia-se agora em $N-1$ graus de liberdade. Se tivessemos considerado os x_1 e os x_2 como duas amostras separadas, a comparação de suas médias apoiar-se-ia em $2\,(N-1)$ graus de liberdade, e resta saber se o aumento de precisão de $\bar{d}$ resultante da redução da variabilidade pelo emparelhamento das medidas é ou não contrabalançado pela perda de precisão devida ao abaixamento de 50% nos graus de liberdade.

Exemplo. Reportando-nos ao Ex. 6.8, vemos que, o fato de terem sido trabalhadas as secções aos *pares* com os dois tipos de segadeiras, permite fazer a comparação mediante as diferenças d, que constam da última coluna do quadro.

Tem-se $\bar{d} = .83$, $t = .83 \sqrt{\dfrac{90}{6.001}} = 3.22$. A tabela de t, com $n = 9$, mostra que êsse valor está aproximadamente no nível de 0.01, e portanto o novo cálculo revela a superioridade da máquina A.

Superficialmente, podiamos esperar que ambos os processos conduzissem ao mesmo resultado. Isso não acontece porque cada uma das medidas da primeira amostra liga-se, de um certo modo, à da segunda, isto é, elas são correlacionadas, enquanto que o processo habitual as pressupõe independentes.

Por conseqüência, devemos corrigir a estimativa da variância comum s^2 do efeito dessa correlação, usando

$$s^2 = \frac{\Sigma\,(x_1 - \bar{x}_1)^2 + \Sigma\,(x_2 - \bar{x})^2 - 2\,\Sigma\,(x_1 - \bar{x}_1)\,(x_2 - \bar{x}_2)}{2\,(N-1)}.$$

E' fácil de ver que o numerador corresponde agora a $\Sigma\,(d-\bar{d})^2$, e os graus de liberdade se reduzem a $N-1$, pois $N-1$ são absorvidos pela estimativa da covariância.

No exemplo tratado, teriamos $s^2 = \dfrac{11.936 + 5.569 - 2\,(5.752)}{18} = .3334$,

e portanto $t = 3.22$, como no segundo processo.

Nada se pode afirmar antecipadamente sôbre a eficiência de um ou outro processo, no caso de medidas emparelhadas. Se há uma correlação positiva entre as variáveis, ou formando as diferenças, ou corrigindo a estimativa de s^2, reduzimos a variância da diferença $\bar{x}_1 - \bar{x}_2$, e, pois, aumentamos a precisão do experimento. Mas êsse ganho pode ser anulado pela perca da metade dos primitivos graus de liberdade.

6.10 *Universos com diferentes variâncias. O teste* Fisher-Behrens.

A aplicação de t como teste da significância da diferença entre médias importa na aceitação ou rejeição da hipótese de que duas populações normais têm a mesma média $\bar{x}_1 = \bar{x}_2$ e a mesma variância

$\sigma^2_{x_1} = \sigma^2_{x_2}$. Se $\sigma^2_{x_1} \neq \sigma^2_{x_2}$, qualquer inferência relativa à validade da hipótese da igualdade das médias é questionável, pois um valor alto de t pode traduzir antes diferenças entre as variâncias que entre as médias. Veremos no capítulo seguinte processos que permitem discernir se as variâncias são ou não iguais. Só no primeiro caso podemos utilizar, a rigor, o teste t.

Suponhamos, por conseguinte, diferentes as variâncias das duas populações. Sejam $s_{\bar{x}_1}$ e $s_{\bar{x}_2}$ as estimativas dos erros padrões das duas médias; t_1 e t_2 quantidades seguindo a distribuição de STUDENT. Então R. A. FISHER [14] mostrou que a estatística

$$d = \bar{x}_2 - \bar{x}_1 \left(s^2_{\bar{x}_1} - s^2_{\bar{x}_2} \right)^{\frac{1}{2}} = t_2 \cos\theta - t_1 \operatorname{sen}\theta \qquad (6.16)$$

onde $\tan\theta = \dfrac{s_{\bar{x}_1}}{s_{\bar{x}_2}}$, devida originàriamente a BEHRENS, fornecia um teste para verificar a hipótese de que a diferença entre as médias das duas populações era nula. Para tanto, devemos considerar a distribuição de $t_2 \cos\theta - t_1 \operatorname{sen}\theta$ em função de $n_1 = N_1 - 1$, $n_2 = N_2 - 1$ e θ, a fim de avaliar a probabilidade de que d exceda um certo valor. A forma analítica da distribuição é difícil de se obter, mas podem ser calculados os valores correspondentes a dados níveis de significância. Para a aplicação dêsse teste, SUKHATME [15] calculou tabelas dando os valores de d para o nível de 5 por cento.

A comparação dos valores a 5% de d e t, baseados em $n_1 + n_2$ graus de liberdade, mostra que os valores de d são maiores que os de t, evidenciando as conclusões errôneas a que pode levar a aplicação do teste t em vez do d, especialmente quando $n_1 = n_2 = n$. É-se então levado rejeitar a hipótese nula, embora sendo verdadeira, quando o valor calculado cai entre $d_{5\%}$ e $t_{5\%}$. Pelo teste t a conclusão é que as duas amostras não provêm da mesma população normal, ao passo que o teste d indica essa possibilidade. Freqüentemente, o exagerado valor de t que se encontra é devido à diversidade das variâncias.

Exemplo. As percentagens médias de cinzas em dois carvões A e B, conforme ensaios tabelados no Ex. 7.12, são $\bar{x}_1 = 9.8$ e $\bar{x}_2 = 7.9$. Verificar se a diferença é significante.

Nesse Ex. 7.12 veremos que as variâncias $s_1^2 = 18.398$ e $s_2^2 = 3.476$ dos dois carvões diferem significantemente, e por isso não podemos empregar com segurança o teste t. Para o teste FISHER-BEHRENS, calculemos

$$s^2_{\bar{x}_1} = 2.628, \quad s^2_{\bar{x}_2} = .574, \quad \tan\theta = \frac{1.621}{.761} = 2.127, \quad \theta = 65^\circ, \quad d = \frac{7.9 - 9.8}{\sqrt{2.628 - .579}} = -1.061 .$$

Entrando na tabela de SUKHATME, vemos que o valor de d para o nível de significância de 5% e os valores $n_1 = n_2 = 6$ e $\theta = 60^\circ$ é 2.436, evidenciando, pois, que os dois carvões não diferem quanto ao teor médio de cinza.

---

[14] FISHER, R. A., "The Fiducial Argument in Statistical Inference", *Annals of Eugenics*, vol. 6 (1935). pág. 391.

[15] SUKHATME, P. V., "On FISHER and BEHRENS test of significance for the diference in means of two normal samples", *Sankhya — The Indian Jour. Stat.*, vol. 4 (1938), pág. 39.

CAPÍTULO VII

A ESTIMAÇÃO E COMPARAÇÃO DE VARIÂNCIAS

7.1 A distribuição da variância.

Vimos, no capítulo anterior (§ 6.3), que a distribuição da variância, em amostras oriundas de um universo normal, era dada por

$$H(s^2) = k_2 \, e^{-\frac{ns^2}{2\sigma^2}} (s^2)^{\frac{n-2}{2}} \, d(s^2), \qquad (7.1)$$

onde $n = N - 1$ são os graus de liberdade usados na estimação de σ^2.

E' essa uma curva do tipo III de PEARSON, com amplitude limitada de um lado a $s^2 = 0$, e estendendo-se até o infinito do outro. O valor da constante k_2 determina-se mediante a condição da área

$$k_2 = \int_0^\infty e^{\frac{ns^2}{2\sigma^2}} (s^2)^{\frac{n-2}{2}} \, d(s^2) = 1.$$

Pela substituição $x = \dfrac{ns^2}{2\sigma^2}$, obtém-se

$$\frac{1}{k_2} = \left(\frac{2\sigma^2}{n}\right)^{\frac{n}{2}} \int_0^\infty x^{\frac{n-2}{2}} e^{-x} \, dx = \left(\frac{n}{2\sigma_2}\right)^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right),$$

e portanto

$$H(s^2) = \frac{n^{\frac{n}{2}}}{(2\sigma^2)^{\frac{n}{2}} \, \Gamma\left(\frac{n}{2}\right)} e^{-\frac{ns^2}{2\sigma^2}} (s^2)^{\frac{n-2}{2}} \, d(s^2). \qquad (7.2)$$

Essa distribuição foi obtida pelo astrônomo alemão HELMERT em 1876; mas o seu trabalho [1] passou desapercebido, e só em 1908 foi a distribuição redescoberta empìricamente por STUDENT.

Podemos reconduzir a distribuição de s^2 a outra, já conhecida. Ponhamos $\chi^2 = \dfrac{\Sigma (x_i - \bar{x})^2}{\sigma^2} = \dfrac{ns^2}{\sigma^2}$. Obtém-se

$$H(s^2) = \left[2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)\right]^{-1} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{n-2}{2}} \, d(\chi^2), \qquad (7.3)$$

que é a distribuição chi-quadrado.

---

[1] Publicado nos *Astronomische Nachrichten* (vol. 88, pág. 122) sob o título "Die Genauigkeit der Formel von Peters zur Berechung des wahrscheinlichen Beobachtungsfehlers".

A vantagem dessa transformação é que, na verificação de hipóteses relativas a s^2, podemos empregar tabelas da distribuição χ^2. Utilizando a expressão (7.2), teriamos de recorrer a outras tabelas, de manuseio muito mais incômodo. Com efeito, a probabilidade de s^2 exceder um certo valor, ou seja, $x = \frac{n s^2}{2\sigma^2}$ exceder o valor correspondente, é dada por

$$P = 1 - \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^x x^{\frac{n-2}{2}} e^{-x} dx = 1 - \frac{\Gamma_x\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},$$

onde $\Gamma_x(x)$ é a função Gama Incompleta, tabelada por PEARSON.[2]

Exemplo. A variabilidade do teor de cinza de um carvão, tipo A, obtida em 7 análises, é de $s^2 = 18.398$ (ver Ex. 7.12). Sabendo-se que os anteriores fornecimentos tinham uma variabilidade média de $\sigma^2 = 10.560$, pode-se admitir que esta partida tenha sido obtida nas mesmas condições?

Temos que $\chi^2 = \frac{6 \times 18.398}{10.560} = 10.450$. Para 6 graus de liberdade, a tabela de χ^2 nos dá a probabilidade dêsse desvio $P = .10$, denotando que não houve alteração significante na homogeneidade do carvão.

7.2 *Estimativa ótima da variância.*

Apliquemos o método da máxima verossimilhança (§ 1.5) para obter a estimativa ótima da variância. A equação (7.2) nos mostra que a verossimilhança é proporcional a $\sigma^{-n} e^{-\frac{n s^2}{2\sigma^2}}$. Para determinar o valor máximo dessa expressão, tomemos primeiramente o seu logarítmo $-n \log_e \sigma - \frac{n s^2}{2\sigma^2}$, e igualemos a zero a sua derivada em relação a σ. Designando por $\hat{\sigma}^2$ a *estimativa ótima*, temos $-n + \frac{n s^2}{\hat{\sigma}^2} = 0$, ou seja $\hat{\sigma}^2 = s^2$.

Quer isso dizer que a melhor estimativa da variância do universo obtém-se calculando a variância da amostra com $n = N - 1$ graus de liberdade. A razão de adotar o divisor $N - 1$ para a soma dos quadrados das discrepâncias, e não N, é que a estimativa da média $\bar{x}$ nessa expressão faz-se com os mesmos dados da amostra, reduzindo de uma unidade o número de variações independentes.

7.3 *Graus de liberdade na distribuição de* s^2.

Suponhamos, porém, que tivessemos um conhecimento prévio do valor da média do universo μ. Não mais subsistindo a exigência do ponto P situar-se no hiper-plano definido pelo valor de x (vide § 6.2), as superfícies de densidade constante serão hiper-esferas a N-dimensões. A diferencial de volume será proporcional a $s^{N-1} ds$, e é fácil de ver que se obtém então a mesma equação (7.2), desde que se substitua n por N.

---

[2] PEARSON, K. (ed.), *Tables of the Incomplete Gamma Function* (Londres, 1922).

Aplicando o método da máxima verossimilhança, segue-se que a estimativa ótima da variância basear-se-á em N graus de liberdade, isto é $\hat{\sigma}^2 = \dfrac{\Sigma (x_i - \mu)^2}{N}$.

Ordinàriamente, adota-se essa mesma fórmula ainda quando não se conhece μ, mas apenas $\bar{x}$. Se a amostra é suficientemente grande, a diferença entre N e $N-1$ é de somenos importância; mas se é pequena, resultam diferenças apreciáveis.

Consideremos, ao invés, o caso em que, além da relação entre a média da amostra e seus valores constituintes, existem p relações independentes entre $x_1, x_2, \ldots x_N$. O ponto P representativo da amostra, além de jazer no hiper-plano $\Sigma x_i = N\bar{x}$, também pertence a p outros hiper-planos. A superfície da hiper-esfera, de centro em M, à qual pertence P, terá, então, dimensões iguais a $N - p - 2$. Se definirmos a variância como $s^2 = \dfrac{1}{n} \sum_1^N (x_i - \bar{x})^2$, com $n = N - p - 1$, obteremos a mesma equação (7.2) que é, assim, a forma geral da distribuição de s^2. A estimativa ótima será aqui baseada em $N - p - 1$ graus de liberdade.

Os graus de liberdade do sistema equivalem, pois, ao número de observações dadas, menos o número de relações independentes que entre elas existem, levando-se em conta que a média da população também é estimada a partir da amostra.

7.4 *Expectância de* s^2 *em universos não-normais.*

A dedução acima fundamenta-se na forma analítica do universo, que supusemos normal. Há interêsse em determinar a expectância de s^2, quando essa forma é desconhecida.

Consideremos amostras independentes de tamanho N da variável u, sendo u a discrepância relativa à média do universo, isto é, $u = x - \mu$. Definamos a variância da amostra como habitualmente se faz,

$$'s^2 = \frac{1}{N} \sum_1^N (u_i - \bar{u}_i)^2 = \frac{1}{N} \sum_1^N u_i^2 - \left[\frac{1}{N} \sum_1^N u_i\right]^2$$

$$= \frac{1}{N} \sum_1^N u_i^2 - \frac{1}{N^2} \sum_1^N u_i^2 - \frac{1}{N^2} \sum_{i \neq j} u_i u_j .$$

Daí se obtém a expectância

$$E('s^2) = \frac{1}{N} E\left\{\sum_1^N u_i^2\right\} - \frac{1}{N^2} E\left\{\sum_1^N u_i^2\right\} - \frac{1}{N^2} E\left\{\sum_{i \neq j} u_i u_j\right\}$$

$$= \frac{1}{N} \sum_1^N \left\{E\, u_i^2\right\} - \frac{1}{N^2} \sum_1^N \left\{E\, u_i^2\right\} - \frac{1}{N^2} \sum_{i \neq j} \left\{E\,(u_i u_j)\right\}$$

Notando que a expectância do produto das discrepâncias de duas variáveis independentes relativamente às próprias expectâncias é nula, $E(u_i u_j) = E(x_i - \mu_i)\ E(x_j - \mu_j) = 0$, obtém-se finalmente

$$E('s^2) = \frac{N-1}{N} \sigma^2. \qquad (7.4)$$

Por conseguinte, a expectância de $'s^2$ para um universo arbitrário é igual à variância do universo multiplicada pelo fator $\frac{N-1}{N}$.

Tal resultado era de se esperar. Para cada amostra, μ constitui uma origem arbitrária. Como a soma dos quadrados das discrepâncias é mínima em relação à média, a média de todos os valores $'s^2$, calculados em relação às médias das respectivas amostras $\bar{x}$, será menor que σ^2, calculado em relação a μ.

Segue-se que a estatística $'s^2$, obtida pela divisão da soma dos quadrados das discrepâncias por N, não é *justa* (§ 1.5). Como $E\left\{\frac{N}{N-1}\,'s^2\right\} = \sigma^2$, a estimativa justa $\hat{\sigma}^2$ da variância do universo será

$$\hat{\sigma}^2 = \frac{N}{N-1}\,'s^2 = \frac{\sum_1^N (x_i - \bar{x})^2}{N-1}.$$

Demonstra-se que $\hat{\sigma}^2$ também é uma estatística *eficiente*, isto é, nenhuma outra estimativa de σ^2 tem menor êrro de amostragem.

O fator $\frac{N}{N-1}$ denomina-se "*correção de* BESSEL", embora tenha sido usado primeiramente por GAUSS.

7.5 *Momentos da distribuição de* s^2.

Tomemos a distribuição de s^2 sob a forma

$$\left[\Gamma\left(\frac{n}{2}\right)\right]^{-1} e^{-\frac{\chi^2}{2}} \left(\frac{\chi^2}{2}\right)^{\frac{n-2}{2}} d\left(\frac{\chi^2}{2}\right).$$

A sua função característica será

$$\psi(t) = \left[\Gamma\left(\frac{n}{2}\right)\right]^{-1} \int_0^\infty e^{-\frac{\chi^2}{2}(1-t)} \left(\frac{\chi^2}{2}\right)^{\frac{n-2}{2}} d\left(\frac{\chi^2}{2}\right).$$

Pondo $\frac{\chi^2}{2}(1-t) = x$, vem

$$\psi(t) = \left[\Gamma\left(\frac{n}{2}\right)\right]^{-1} (1-t)^{-\frac{n}{2}} \int_0^\infty e^{-x}\, x^{\frac{n-2}{2}}\, dx = (1-t)^{-\frac{n}{2}},$$

Como $\frac{\chi^2}{2} = \frac{n}{2\sigma^2}$, a função característica de s^2 será $\psi(t) = \left(1 - \frac{2\sigma^2}{n}t\right)^{-\frac{n}{2}}$; e a função geratriz de semi-invariantes nos dá

$$L(t) = -\frac{n}{2}\log_e\left(1 - \frac{2\sigma^2}{n}t\right) = \frac{n}{2}\left(2\frac{\sigma^2 t}{n} + 4\frac{\sigma^4 t^2}{n} + \ldots\right) =$$

$$= \sigma^2 t + \frac{2\sigma^4}{n}\frac{t^2}{2!} + \ldots$$

O valor médio de s^2 é pois σ^2, a variância $\frac{2\sigma^4}{n}$; daí o êrro padrão $\sigma_{s^2} = \sigma^2\sqrt{\frac{2}{n}}$.

Quando n cresce, a função geratriz tende para equivalência assimptótica com $\psi(t) = e^{\sigma^2 t + \frac{\sigma^4 t^2}{n}}$, e portanto a distribuição de s^2 tende para a normalidade.

7.6 A distribuição de s e seus momentos.

A distribuição de s obtém-se imediatamente de (7.1), e resulta

$$H(s) = 2k_2\, e^{-\frac{ns^2}{2\sigma^2}} s^{n-1}\, ds = \frac{n^{\frac{n}{2}}}{2^{\frac{n-2}{2}}\, \sigma^n\, \Gamma\left(\frac{n}{2}\right)} e^{-\frac{ns^2}{2\sigma^2}} s^{n-1}\, ds . \qquad (7.5)$$

Anulando a derivada primeira dessa equação, acha-se o valor modal

$$\check{s} = \sigma \sqrt{\frac{n-1}{n}} .$$

Pela substituição $x = s - \check{s}$, verifica-se que a distribuição de x é, em primeira aproximação, uma curva normal com desvio padrão $\frac{\sigma}{\sqrt{2N}}$.

O cálculo do momento de ordem h da distribuição de s efetua-se a partir de sua definição

$$M_h = 2\, k_2 \int_0^\infty e^{-\frac{ns^2}{2\sigma^2}} s^{n-1+h}\, ds .$$

Substituindo a variável $s = \sqrt{x}$, vem

$$M_h = k_2 \int_0^\infty e^{-\frac{nx}{2\sigma^2}} x^{\frac{n-2+h}{2}}\, dx = k_2 \left(\frac{2\sigma^2}{n}\right)^{\frac{n+h}{2}} \Gamma\left(\frac{n+h}{2}\right) = \frac{\Gamma\left(\frac{n+h}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{2\sigma^2}{n}\right)^{\frac{h}{2}} .$$

Daí se conclui o valor médio de

$$E(s) = \frac{\left(\frac{2}{n}\right)^{\frac{1}{2}} \Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sigma = b(N)\, \sigma \qquad (7.6)$$

Resulta também que a estimativa *justa* de σ é $\hat{\sigma} = [b(N)]^{-1} s$, sendo que os valores de $b(N)$ acham-se tabelados.[3] E' costume[4] tomar-se como estimativa $\hat{\sigma} = \left[\frac{N}{N-1}\right]^{\frac{1}{2}} s$. Ora ROMANOWSKY mostrou que

$$b(N) = 1 - \frac{3}{4N} - \frac{7}{32N^2} - \ldots$$

[3] V. g., SHEWHART, W. A., *Economic Control of Quality of Manufactured Product* (Nova Iorque, 1931), pág. 185.

[4] KENNEY, J. F., *Mathematics of Statistics*, vol. 2, pág. 126.

A prática habitual corresponde a substituir essa expressão por

$$'b\,(N) = 1 - \frac{1}{2N} - \frac{1}{8N^2} - \ldots$$

A relação entre as duas expressões $\frac{b\,(N)}{'b\,(N)} = 1 - \frac{1}{4N} - \frac{11}{32N^2} - \ldots$ dá-nos a grandeza do êrro que assim se comete.

Vejamos agora o êrro padrão de s. Temos

$$\sigma_s^2 = E\,[s - E\,(s)]^2 = E\,(s^2) - [E\,(s)]^2 = \sigma^2\left[1 - \frac{2}{n}\,\frac{\Gamma^2\left(\frac{n+1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)}\right].$$

ROMANOWSKY [5] indicou o desenvolvimento $\frac{\Gamma^2\left(\frac{n+1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} = \frac{N}{2}\left[1 - \frac{3}{2N} + \right.$

$+ \frac{2}{8N^2} + \frac{3}{16N^3} + \ldots$, donde se conclui

$$\sigma_s = \sigma\left[\frac{1}{2(N-1)} - \frac{2}{8N(N-1)} - \frac{3}{16N^2(N-1)} + \ldots\right]^{\frac{1}{2}}.$$

A fórmula usualmente apresentada nos compêndios corresponde apenas à aproximação do primeiro têrmo $\sigma_s = \frac{\sigma}{\sqrt{2N}}$, que é interpretada como sendo o êrro do desvio padrão $\frac{1}{\sqrt{2}}$ vezes o da média. Para sua utilização na prática fazem-se mister mais duas suposições, que só são válidas para N suficientemente grande. Primeiro, a incógnita σ é substituída pelo s calculado; depois, admite-se que a distribuição de s é normal.[6]

7.7 *Variância de* s *em universos não-normais*

Suponhamos desconhecida a fórma analítica do universo, que admitimos apenas definido pelos seus momentos $\mu_2, \mu_3, \ldots$ referidos à média. Adotemos um sistema de coordenadas tal que o valor médio da variável x seja nulo; então $\mu_k = E\,x^k$.

A variância de uma amostra de tamanho N será

$$s^2 = \frac{\Sigma\, x_i^2}{N} - \Sigma\left(\frac{x_i}{N}\right)^2.$$

Calculemos a sua variância. Temos

$$\sigma_{s^2}^2 = E\,[s^2 - E\,(s^2)] = E\,(s^2)^2 - [E\,(s^2)]^2.$$

---

[5] ROMANOWSKY, V., "On the moments of standard deviation and of correlation coefficient in samples from normal", *Metron*, vol. 5 (1925), pág. 3.

[6] Cf. HOTELLING, H., "The Consistency and Ultimate Distribution of Optimum Statistics", *Trans. Amer. Mathem. Soc.*, vol. 32 (1930), pág. 851.

Ora, $E\,s^2 = E\,\frac{\Sigma x_i^2}{N} = \mu_2$; e $E\,(s^2)^2 = E\left(\frac{\Sigma x_i^2}{N}\right)^2 = \frac{1}{N^2}\left\{N\,E\,x_i^4 + N(N-1)\,E(x_i^2 x_j^2)\right\}$.

Como as amostras são independentes, $E\,(x_i^2 x_j^2) = (E^2\,x)^2 = \mu_2^2$, e portanto $E\,(s^2)^2 = \frac{1}{N}\left\{\mu_4 + (N-1)\,\mu_2^2\right\}$. Obtém-se finalmente

$$\sigma_{s^2}^2 = \frac{\mu_4 - \mu_2^2}{N}. \tag{7.7}$$

Para obter a variância do desvio padrão de uma amostra de tamanho N, notemos que, como a média de s em amostras suficientemente grandes difere tão pouco quanto se queira de σ, temos que o valor médio da variância será aproximadamente o valor de $(s - \sigma)^2 = \frac{(s^2 - \sigma^2)^2}{(s+\sigma)^2}$. O numerador tem por valor médio $\sigma_{s^2}^2$. Para grandes amostras, podemos substituir $s + \sigma$ por 2σ, e temos, assim, aproximadamente, $\sigma_s^2 = \sigma_{s^2}^2 / 4\,\sigma^2$, ou seja

$$\sigma_s^2 = \frac{\mu_4 - \mu_2^2}{4\,N\,\mu_2}. \tag{7.8}$$

Para uma população de distribuição normal, $\mu_4 = 3\,\mu_2^2$, e obtém-se

$$\sigma_{s^2} = \sqrt{\frac{2\,\mu_2}{N}} = \sigma\sqrt{\frac{2}{N}}, \qquad \sigma_s = \sqrt{\frac{\mu_2}{2N}} = \frac{\sigma}{\sqrt{2N}},$$

como anteriormente acháramos.

7.8 *A superfície de freqüência* u,s. *Teste* λ.

A distribuição (6.8) pode ser transformada de modo a dar a distribuição conjunta das discrepâncias das médias das amostras em relação á do universo, $u = \bar{x} - \mu$, e de s. Obtém-se uma função do tipo $F\,(u,s) = G\,(u)\,H\,(s)$, a qual define a superfície de freqüência u,s.[7]

O volume elementar $F\,(u,s)$ dá a percentagem de amostras cujas discrepâncias caem no intervalo $u \pm \frac{1}{2}\,du$ e cujos desvios padrões caem em $s \pm \frac{1}{2}\,ds$; integrando, obtém-se o volume sob essa superfície compreendido num contôrno fechado no plano u,s, que dá a proporção de amostras cujas discrepâncias e desvios padrões caem simultâneamente nos intervalos definidos pela figura de contôrno.

Em vista da independência de u e s, tôdas as secções planas $u = const.$ dessa superfície serão curvas assimétricas definidas pela equação $H\,(s)$; para valores crescentes de N, elas tendem para a normalidade, com centro em $s = \sigma$ e desvio padrão $\sigma/\sqrt{2N}$. As curvas $s = const.$ são normais, tôdas com centro em $u = 0$ e desvio padrão $\sigma/\sqrt{N}$. Quando N cresce, a superfície tende a se concentrar em tôrno do ponto $u = 0, s = \sigma$.

A posição de uma amostra no plano u,s só se pode fixar se conhecido o parâmetro μ do universo. A consideração da superfície u,s torna-se útil para a verificação de hipóteses relativas à média, ou ao des-

---

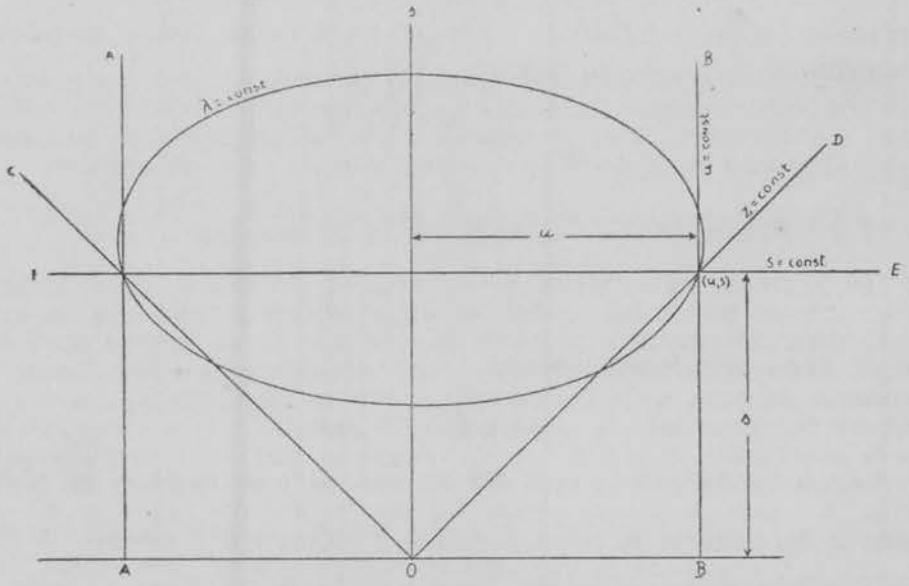
[7] Deming, W. E., e Birge, R. T., "On the Statistical Theory of Errors", *Rev. Modern Physics*, vol. 6 (1934), pág. 130.

vio padrão, ou a ambos. Dados os valores hipotéticos de μ e σ, marcamos a posição do ponto representativo da amostra (u,s). Por êsse ponto podem ser traçados vários contornos que dividem o volume simètricamente em relação ao eixo dos s. Êsses contornos são:

$$\pm u = const., \quad s = const., \quad \pm z = \frac{u}{s} = const.,$$

$$\lambda = \left(\frac{s}{\sigma}\right)^N e^{-\frac{N}{\sigma^2}(u+s^2-\sigma^2)} = const.$$

Os três primeiros são retas prolongando-se até o infinito: o último consiste em curvas ovais fechadas, rodeando o ponto máximo da superfície de coordenadas $u = 0$, $s = \sigma\sqrt{\frac{N-2}{N}}$.



A fração de volume exterior aos contornos simétricos *AA* e *BB* equivale à probabilidade de se obter uma amostra de N elementos, cuja média difira da do universo mais que dado valor de u. Temos assim o *teste u* (§2.9).[8] Tomando a fração do volume além do contôrno *EE*, temos a probabilidade de se obter uma amostra com desvio padrão maior que um dado s. O cálculo dêsse volume conduz à função Gama Incompleta (§ 7.1), pois $P_s = 1 - \Gamma_x\left(\frac{N-1}{2}\right)\Big/\Gamma\left(\frac{N-1}{2}\right)$, sendo $x = \frac{n s^2}{2\sigma^2}$.

Os contornos *CO* e *CD* são traçados de modo a formar os ângulos $\pm arc\ tan\ \frac{u}{s}$ com o eixo dos s. A fração P_z do volume exterior a êsses contornos mede a probabilidade de se obter uma amostra onde a razão $z = \frac{u}{s}$ seja maior que a razão entre o valor hipotético de u e o s

[8] Notemos que o u definido no § 2.9 equivale ao atual $\frac{u\sqrt{n}}{\sigma}$.

observado. Chega-se à função Beta Incompleta, ou, mais precisamente, à integral da distribuição de STUDENT (§ 6.4). Notemos que os contornos z são os únicos independentes de σ.

Os testes descritos dependem de contornos que se estendem até o infinito; por isso, para extrair tôdas as informações referentes ao universo donde porém a amostra, devemos recorrer às vêzes a mais de um deles. Notável contribuição deve-se a NEYMAN e PEARSON,[9] cujo *teste* λ baseia-se numa única família de contornos fechados, as curvas λ [10]. Ao longo de uma dessas curvas, permanece constante a razão entre a freqüência de um dado ponto da superfície u,s para o valor máximo correspondente ao ponto de coordenadas $(0,\sigma)$. A fração de volume compreendido pela superfície u,s, exterior ao contôrno λ, é dada por

$$P_\lambda = K \iint \left(\frac{s}{\sigma}\right)^{N-2} e^{-\frac{N}{2\sigma^2}(u^2+s^2)} \, du \, ds ,$$

a integração se realizando externamente à curva λ. NEYMAN e PEARSON publicaram tabelas e diagramas facultando a aplicação dêsse teste, que permite distinguir, quando P_λ é excepcionalmente pequeno, se a infirmação da hipótese se deve atribuir a μ ou σ, ou a ambos.

7.9 A estimação de σ a partir de várias amostras.

Se tivermos várias séries de observações, do mesmo grau de precisão, mas de diferentes médias, devemos avaliar a variância do conjunto para determinar a precisão das médias. Se as médias dos universos, como geralmente acontece, são desconhecidas, calculamos a variância de cada série em relação à sua média. Seja $N_i = x_i + 1$ o número de elementos e s_i^2 a variância da amostra i, e m o número de séries observadas $(i = 1, 2, \ldots m)$. Segue-se que $n_i s_i^2$ dividido pela verdadeira variância σ^2 tem a distribuição χ^2 com n_i graus de liberdade; e, em virtude da aditividade dessa estatística, também $\frac{1}{\sigma^2} \sum_1^m n_i s_i^2$ tem a distribuição χ^2 com um número de graus de liberdade igual a $\Sigma n_i = N_1 + N_2 + \ldots N_m - m$, isto é, a diferença entre o número total de observações e o número de séries.

A verossimilhança é aqui proporcional a $\sigma^{-\Sigma x_i} e^{-\frac{\Sigma n_i s_i^2}{2\sigma^2}}$, e, por conseguinte, a estimativa ótima de σ, obtida pelo método da máxima verossimilhança, será

$$\hat{\sigma}^2 = \frac{\Sigma n_i s_i^2}{\Sigma n_i} . \qquad (7.9)$$

Se a verdadeira média é conhecida para cada caso, é a partir dela que se computarão os devios, e a fórmula será válida, substituindo n_i por N_i.

---

[9] NEYMAN, J., e PEARSON, E. S., "On the use and interpretation of certain test criteria for purposes of statistical inference", *Biometrika*, vol. 20 A (1928), pág. 175.

[10] Outro sistema de curvas fechadas, os contornos δ, fornecem também um teste pelo cálculo de P_δ; mas a diferença entre êste e P_λ é insignificante e o último é preferível por certas razões teóricas.

Quando a variância difere de uma série para outra, mas numa determinada proporção, então a variância da série i será $k_i \sigma^2$, e a variável $\Sigma \frac{n_i s_i^2}{k_i \sigma^2}$ terá a distribuição χ^2 com $\Sigma k_i n_i$ graus de liberdade. A estimativa ótima da variância será $\hat{\sigma}^2 = \frac{\Sigma n_i s_i^2}{\Sigma k_i n_i}$.

7.10 Comparação de duas variâncias. A distribuição F.

O desenvolvimento de um teste de significância para a diferença entre as variâncias de duas amostras aleatórias, além de sua importância intrínseca, é capital para a exata comparação entre duas médias pelo teste t. Com efeito, uma das hipóteses aí implícitas é que a variância de duas amostras podiam ser utilizadas conjuntamente como estimativa da variância do universo, e precisamos verificar até que ponto tal suposição é real.

Para tal fim, R. A. FISHER [11] considerou, não a própria diferença entre as variâncias s_1^2 e s_2^2, mas a sua relação s_1^2/s_2^2. Além das dificuldades analíticas que assim são aplainadas, demonstra-se que essa estatística corresponde ao princípio da *razão de verossimilhança* (likelihood ratio), estabelecido por NEYMAN e PEARSON.

Sejam n_1 e n_2 os graus de liberdade das estimativas das variâncias das duas amostras, obtidas independentemente; sabemos que as expressões $\frac{n_1 s_1^2}{\sigma_1^2}$ e $\frac{n_2 s_2^2}{\sigma_2^2}$ têm distribuição χ_1^2 e χ_2^2 com n_1 e n_2 graus de liberdade respectivamente. Para a hipótese nula, isto é, de que ambas as amostras provenham de universos com a mesma variância, temos que s_1^2/s_2^2 se distribui como $F = \frac{n_2}{n_1} \frac{\chi_1^2}{\chi_2^2}$.

E' essa a distribuição da *análise da variância,* também chamada da *razão de variâncias* (variance ratio),[12] à qual se reconduzem todos os problemas oriundos daquele método de análise, a notação F tendo sido adotada em honra de seu inventor, R. A. FISHER.

Como χ_1^2 e χ_2^2 se distribuem independentemente, a sua distribuição conjunta obtém-se como o produto de ambas as distribuições, donde

$$\left[2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)\right]^{-1} (\chi_1^2)^{\frac{n_1-2}{2}} (\chi_2^2)^{\frac{n_2-2}{2}} e^{-\frac{1}{2}(\chi_1^2+\chi_2^2)} d\chi_1^2 d\chi_2^2. \quad (7.10)$$

Introduzamos as novas variáveis definidas por $\frac{\chi_1^2}{\chi_2^2} = \frac{n_1}{n_2} F$ e $\chi_1^2 + \chi_2^2 = \chi^2$.

Daí se conclui $\chi_1^2 = \chi^2 \frac{n_1}{n_2} F \left(1 + \frac{n_1}{n_2} F\right)^{-1}$ e $\chi_2^2 = \chi^2 \left(1 + \frac{n_1}{n_2} F\right)^{-1}$.

---

[11] FISHER, R. R., "On a distribution yielding the error functions of several well known statistics", *Proc. Int. Math. Congress,* Toronto (1924), vol. 2, pág. 806.

[12] FISHER, R. A., e YATES, F., *Statitsical Tables,* pág. 2.

Para obter a distribuição conjunta de χ^2 e F, temos de substituir χ_1^2 e χ_2^2 por seus valores dados pelas equações acima, e dx_1^2, dx_2^2 pela expressão $J \frac{(\chi_1^2, \chi_2^2)}{(F, \chi^2)} \, dF \, d\chi^2$, onde J é o Jacobiano da transformação

$$J = \begin{vmatrix} \frac{\partial \chi_1^2}{\partial F} & \frac{\partial \chi_1^2}{\partial \chi^2} \\ \frac{\partial \chi_2^2}{\partial F} & \frac{\partial \chi_2^2}{\partial \chi^2} \end{vmatrix}$$

No caso vertente, temos

$$J = \begin{vmatrix} \chi^2 \frac{n_1}{n_2} \left(1 + \frac{n_1}{n_2} F\right)^{-2} & \frac{n_1}{n_2} F \left(1 + \frac{n_1}{n_2} F\right)^{-1} \\ -\chi^2 \frac{n_1}{n_2} \left(1 + \frac{n_1}{n_2} F\right)^{-2} & \left(1 + \frac{n_1}{n_2} F\right)^{-1} \end{vmatrix} = \chi^2 \frac{n_1}{n_2} \left(1 + \frac{n_1}{n_2} F\right).$$

Substituindo êsse valor na equação (7.10), obtém-se

$$\left[2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)\right]^{-1} \left(\frac{n_1}{n_2} F\right)^{\frac{n_1-2}{2}} \left(1 + \frac{n_1}{n_2} F\right)^{-\frac{n_1+n_2}{2}} (\chi^2)^{\frac{n_1+n_2-2}{2}} e^{-\frac{\chi^2}{2}} \frac{n_1}{n_2} d\chi^2 \, dF$$

$$= \left[\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)\right]^{-1} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} F^{\frac{n_1-2}{2}} \left(1 + \frac{n_1}{n_2}\right)^{-\frac{n_1+n_2}{2}} \left(\frac{\chi^2}{2}\right)^{\frac{n_1+n_2-2}{2}} e^{-\frac{\chi^2}{2}} d\left(\frac{\chi^2}{2}\right) dF.$$

Para obter agora a distribuição de uma das variáveis, devemos integrar a distribuição conjunta em relação à outra variável em todo o seu intervalo de variação. Ora, F dá-nos exatamente a distribuição procurada; quanto a $\frac{\chi^2}{2}$, o seu intervalo é de 0 a $+\infty$, pois também êsses são os intervalos de χ_1^2 e χ_2^2. Daí temos que

$$dp = \left[\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)\right]^{-1} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} F^{\frac{n_1-2}{2}} \left(1 + \frac{n_1}{n_2} F\right)^{-\frac{n_1+n_2}{2}} dF$$

$$\int_0^\infty \left(\frac{\chi^2}{2}\right)^{\frac{n_1+n_2-2}{2}} e^{-\frac{\chi^2}{2}} d\left(\frac{\chi^2}{2}\right)$$

ou seja

$$dp = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} F^{\frac{n_1-2}{2}}}{(n_2 + n_1 F)^{\frac{n_1+n_2}{2}}} dF \qquad (7.11)$$

que é a distribuição da análise da variância. O seu intervalo de variação vai de 0 a $+\infty$, e, tal como a distribuição de t, ela não depende de σ.

7.11 *A distribuição* z *de* Fisher.

Na realidade, para verificar a significância da diferença de duas variâncias, R. A. Fisher não utilizou a razão s_1^2/s_2^2, mas a metade de seu logarítimo neperiano. Pondo [13] $z = \frac{1}{2} \log_e F$, tem-se $F = e^{2z}$, e portanto a diferencial de probabilidade (7.11) transforma-se em

$$\frac{2\, n_1^{\frac{n_1}{2}}\, n_2^{\frac{n_2}{2}}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{e^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{n_1+n_2}{2}}}\, dz, \qquad (7.12)$$

onde B representa a função Beta.

A variável z varia entre $-\infty$ e $+\infty$, sendo negativa quando s_1/s_2 é menor que a unidade, e recìprocamente. A sua distribuição é assimétrica, a menos que $n_1 = n_2$; mas a parte positiva da curva $z = \log_e s_1/s_2$ é a mesma que a parte negativa de $z = \log_e s_2/s_1$, de modo que basta o conhecimento das integrais de probabilidade do ramo positivo para o cálculo relativo a qualquer combinação de graus de liberdade.

A distribuição de z é extremamente geral, e compreende como casos particulares diversas outras. Por exemplo, para $n_2 = \infty$, $n_1 = n$, temos a distribuição χ^2 mediante substituição $e^{2z} = \frac{\chi^2}{n}$ ou $z = \frac{1}{2} \log_e \frac{\chi^2}{n}$. Para $n_1 = 1$, $n_2 = n$, temos a distribuição de Student com a substituição $z = \frac{1}{2} \log_e t^2$. Para $n_1 =$, $n_2 = \infty$, vem a distribuição normal mediante a transformação $z = \frac{1}{2} \log_e u^2$.

7.12 *Aplicação e tabelas das distribuições* F *e* z.

Para verificar a significância da diferença entre duas variâncias, devemos calcular a probabilidade de que F seja maior que o valor s_1^2/s_2^2 observado; isto é, devemos calcular a integral da função (7.11) entre êsse valor e $+\infty$. Alternativamente, toma-se $\frac{1}{2} \log_e \frac{s_1^2}{s_2^2}$, e integra-se a função (7.12) daí até $+\infty$.

Tabelas têm sido calculadas, dando os valores de F ou z correspondentes a vários valores de n_1 e n_2, e os níveis de 0.05, 0.01 e 0.001, isto é, os valores que seccionam uma parte do ramo positivo da curva de distribuição equivalente a essas frações.

As tabelas de z são devidas a Fisher [14] para os níveis de 0.05 e 0.01, e sua extensão a 0.001 a Colcord e Deming [15]; essa extensão facilita conjeturas sôbre a probabilidade P, quando z cai além do nível de 0.01. Nessas tabelas os últimos argumentos estão em progressão harmônica, e a interpolação é aproximadamente linear se se tomam os inversos de n_1 e n_2.

---

[13] Êsse z não se deve confundir com o utilizado por Student em sua distribuição.

[14] Fisher, R. A., *Statistical Methods for Research Workers*, tab. VI; Fisher, R. A., e Yates, F., *Statistical Tables*, tab. V.

[15] Colcord, C. G., e Deming, L. S., "The One-tenth Percent Level of *Z*", *Sankhya — The Indian Jour. Stat.*, vol. 2 (1936), pág. 423.

Para valores altos de n_1 e n_2 $(n_1 > 24, n_2 > 30)$, FISHER indica a aproximação $z = \frac{u}{\sqrt{h-1}} - \lambda\left(\frac{1}{n_1} - \frac{1}{n_2}\right)$, onde u é o desvio correspondente ao nível de significância em questão para um dos ramos da curva normal, h indica a média harmônica de n_1 e n_2, isto é, $\frac{2}{h} = \frac{1}{n_1} + \frac{1}{n_2}$, e $\lambda = \frac{x_2 + 2}{6}$. COCHRAN [16] obtém melhores resultados substituindo $\sqrt{h-1}$ por $\sqrt{h-\lambda'}$, onde $\lambda' = \frac{x^3 + 3}{6}$. Os valores dessas constantes estão consignados abaixo.

Nível sign.	.05	.01	.001
u	1.6449	2.3263	3.0902
λ	.7843	1.235	1.925
λ'	.95	1.40	2.09

Para o cálculo de z podem-se empregar, evidentemente, os logarítmos decimais; temos a relação

$$\frac{1}{2} \log_e \frac{s_1^2}{s_2^2} = 1.151 \log_{10} \frac{s_1^2}{s_2^2}.$$

Nas aplicações, torna-se mais prático o emprêgo das tabelas de F, que evitam o cálculo de logarítmos. Tais tabelas foram apresentadas por MAHALANOBIS [17] e por SNEDECOR. [18]

É usual a advertência nessas tábuas de que n_1 deve sempre corresponder à maior variância. Assinalou A. WALD em suas lições (*Columbia University*, 1941) que, nessas condições, a região crítica compreende todos os pontos que satisfazem ao menos uma das desigualdades

$$F = \frac{n_2}{n_1} \frac{\chi_1^2}{\chi_2^2} \geqq F_0 \quad \text{ou} \quad \frac{1}{F} = \frac{n_1}{n_2} \frac{\chi_2^2}{\chi_1^2} \geqq F_0',$$

onde F_0 corresponde aos valores críticos da tabela relativos a n_1 e n_2 graus de liberdade, e F'_0 os relativos a n_2 e n_1. Como essas desigualdades são mùtuamente exclusivas, a probabilidade de que ao menos uma delas seja satisfeita é igual à soma das probabilidades de ambas, isto é,

$$P = Prob.\left(F \geqq F_0 \quad ou \quad \frac{1}{F} \geqq F'_0\right) = Prob.\ (F \geqq F_0) + Prob.\left(\frac{1}{F} \geqq F'_0\right),$$

e como cada uma dessas probabilidades é igual ao nível de significância de 5% (ou 1%), segue-se que $P = 10\%$ (ou 2%). Por conseguinte, a aplicação estrita da regra corresponde ao uso da região crítica de 10% (ou 2%).

---

[16] COCHRAN, W. G., "Note on an approximate formula for the significance levels of Z", *Ann. Math. Stat.*, vol. 11 (1940), pág. 93.

[17] MAHALANOBIS, P. C., "Auxiliary tables for Fisher's Z-test for use in the Analysis of Variance", *Ind. Jour. Agric. Scien.*, vol. 2 (1932), pág. 679.

[18] SNEDECOR, G. W., *Calculation and Interpretation of Analysis of Variance and Covariance* (Ames, Iowa, 1934).

Na maioria dos problemas de análise da variância, temos a considerar apenas as alternativas $\sigma_1 \geqslant \sigma_2$, e portanto devemos usar apenas o ramo à direita da curva; as tabelas devem ser usadas sem se atender à advertência mencionada. Isto é, dividiremos sempre s_1^2 por s_2^2; se acontecer que s_1^2 seja menor que s_2^2, então não é necessário a consulta às tábuas, porque todos os valores críticos (para o ramo à direita) são maiores que a unidade. A hipótese é então rejeitada.

Exemplo. A tabela abaixo dá os resultados de ensaios para determinar as percentagens de cinza em amostras de carvões provenientes de duas minas, A e B. Pede-se comparar os dois carvões relativamente à homogeneidade.

Amostra	Carvão A	Carvão B
1	5.6	8.3
2	13.2	7.6
3	12.5	4.7
4	4.6	10.2
5	13.7	9.1
6	5.5	7.5
7	13.5	—
Média	9.0	7.9
Variância	18.398	3.476

Dos elementos fornecidos conclui-se $F = \dfrac{18.398}{3.476} = 5.293$. Para $n_1 = 6$, $n_2 = 5$, temos que o valor de F correspondente ao nível de significância de 5% é 4.95. Por conseguinte, a variabilidade do carvão A é significantemente maior que a do carvão B.

O teste poderia ter sido feito com a tabela z. Teríamos $\log_e s_1^2 - \log_e s_2^2 = 3.9122 - 1.2459 = 2.6663$, donde $z = 1.332$. A tabela nos dá, para $n_1 = 6$, $n_2 = 5$, o valor de $z = 0.7997$, donde se confirma a significância da diferença.

7.13 *Comparação entre variâncias de grandes amostras*

Quando tanto n_1 como n_2 são grandes números, ou, se moderados, são iguais ou quase iguais, a distribuição de z aproxima-se suficientemente da forma normal, com média zero e variância $\dfrac{1}{2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$. Podemos, pois, referir a estatística $u = \dfrac{z}{\left[\dfrac{1}{2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\right]^{\frac{1}{2}}}$ à tabela da curva normal.

Outro processo baseia-se em que a variância da diferença $w = s_1^2 - s_2^2$ é, de conformidade com a fórmula 2.6 e por se tratar de grandes amostras, $\sigma_w = \sigma\left(\dfrac{1}{2N_1} + \dfrac{1}{2N_2}\right)^{\frac{1}{2}}$. A estatística $u = \dfrac{s_1^2 - s_2^2}{\sigma_w}$ tem, então, distribuição normal, em tôrno da média zero e desvio padrão unitário. A estima de σ faz-se como indicado no § 7.9. Se as amostras são suficientemente grandes para que se possam confundir

n_1 e n_2 com N_1 e N_2, resulta finalmente $u = \dfrac{s_1^2 - s_2^2}{\left[\dfrac{s_1^2}{2N_2} + \dfrac{s_2^2}{2N_1}\right]^{\frac{1}{2}}}$, que se interpreta mediante a tabela da curva normal.

7.14 Significância das variações entre várias amostras.

Analisando dados experimentais, tem-se freqüentemente necessidade de verificar a homogeneidade de um conjunto de estimativas de variâncias. Por exemplo, se queremos combinar os resultados de várias experiências ou de várias fontes, devemos primeiro perquirir sôbre essa possibilidade; ou ainda, ao aplicar os processos de análise da variância, convém verificar essa homogeneidade, sôbre a qual êles se baseiam.

O problema foi abordado com caráter mais geral por NEYMAN e PEARSON [19] que procuraram critérios para verificar se um grupo de amostras, oriundas da mesma população normal, se diferenciam estatìsticamente em suas médias ou variâncias. Três foram as hipóteses consideradas e os respectivos testes de significância:

(i) Hipótese H_0: as amostras provêm de populações normais, tendo a mesma média e a mesma variância (teste L_0).

(ii) Hipótese H_1: as amostras provêm de populações tendo a mesma variância, independentemente de terem ou não a mesma média (teste L_1).

(iii) Hipótese H_2: as amostras provêm de populações tendo médias aproximadamente iguais, admitindo-se que as variâncias também sejam iguais (teste L_2).

Seja k o número de amostras consideradas, e n_t, $\bar{x}_t$ e s_t^2 o tamanho, a média e a variância da amostra t; então

$$\bar{x}_t = \frac{\sum_{i=1}^{n_t} (x_{ti})}{n_t}, \qquad s_t^2 = \frac{\sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2}{n_t}.$$

A variância média de tôdas as amostras será $s_a^2 = \dfrac{\sum_{t=1}^{k} (n_t s_t^2)}{N}$, onde $N = \sum_{t=1}^{k} n_t$. Quanto à variância geral de todos os valores, temos

$$\bar{x}_0 = \frac{\sum_{t=1}^{k} \sum_{i=1}^{n_t} (x_{ti})}{N}, \qquad s_0^2 = \frac{\sum_{t=1}^{k} \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_0)^2}{N}.$$

No caso particular em que tôdas as amostras são do mesmo tamanho, $n_1 = n_2 = \ldots = n_k = n, N = nk$, as estatísticas L tomam a forma simplificada

[19] O seu trabalho "On the problem of k samples" foi publicado no *Bull. Acad. Polonaise Scien. et Let.*, ser. A, 1931.

$$L_0 = \frac{1}{s_0^2}\left\{s_1^2 \cdot s_2^2 \cdot \ldots \cdot s_k^2\right\}^{1/k}$$

$$L_1 = \frac{1}{s_a^2}\left\{s_1^2 \cdot s_2^2 \cdot \ldots \cdot s_k^2\right\}^{1/k} \qquad (7.13)$$

$$L_2 = s_a^2 / s_0^2.$$

Estas estatísticas são iguais à unidade quando as hipóteses H_0, H_1 e H_2 são verdadeiras, e decrescem até zero à medida que erramos nessas afirmativas.

A forma exata da distribuição dessas estatísticas ainda não é conhecida, mas, baseados nos seus primeiros momentos, pode-se ajustar-lhes curvas do tipo I de PEARSON, e calcular os valores correspondentes aos vários níveis de significância. Essas tabelas foram preparadas por MAHALANOBIS [20] para L_0 e L_1, e por NAYER [21] para L_1, com maior amplitude de valores k e n.

Constata-se que $L_2 = \eta^2$, sendo η a razão de correlação de PEARSON. Também $L_2 = \left[1 + \frac{k-1}{N-k}\, e^{2z}\right]^{-1}$, onde z é a função usada por FISHER como teste da análise da variância. Quando $z \to 0$, $L_2 \to {}^{(N-k)}/_{(N-1)}$, do $z \to \infty$, $L_2 \to 0$. Assim, L_2 fornece apenas uma forma alternativa do teste z de FISHER.

A aplicação dos testes L_0 e L_1 faz-se como segue: se o valor de L_0 obtido da amostra não é significante, confirma-se a hipótese H_0, isto é, tôdas as amostras provêm de populações normais caracterizadas pela mesma média e mesma variância, quer dizer, idênticas; o teste L_1 e a análise da variância nenhuma informação podem acrescentar, pois êles apenas verificam a *natureza* da heterogeneidade das amostras.

Se L_0 é significante, cumpre discenir se a heterogeneidade é devida a diferenças nas médias, ou nas variâncias, ou em ambas. Aplicando o teste L_1, verificamos se a hipótese da igualdade das variâncias é plausível ou não. Se $L_1 \doteq 1$, a variabilidade pode-se considerar constante, o que permite aplicar com segurança o teste da análise da variância para verificar a variabilidade das médias. Ao contrário, se $L_1 = 0$, a variabilidade não é constante para as amostras, e se o teste z também é positivo, conclui-se que as amostras diferem tanto relativamente às médias como às variâncias.

Exemplo. Para comparar a eficiência de 3 métodos de ensino, realizaram-se experiência com 3 turmas de 20 alunos cada, constatando-se as seguintes variabilidades no teste final de aproveitamento: $s_1^2 = 74$, $s_2^2 = 132$, $s_3^2 = 97$. A variância do resultado global das 3 turmas é $s_0^2 = 107$. Podem-se considerar as turmas como homogêneas, isto é, derivando de uma população normal comum? Podem-se considerar como homogêneas em relação à variância, independentemente das médias de aproveitamento?

A primeira questão resolve-se pelo teste L_0, a segunda pelo L_1. Temos

$$\sqrt[3]{s_1^2 \cdot s_2^2 \cdot s_3^2} = \sqrt[3]{74 \times 132 \times 97} = 98.22, \quad s_a^2 = \frac{30}{90}(74 + 143 + 97) = 101.$$

---

[20] MAHALANOBIS, P. C., "Tables for *L*-Tests", *Sankhya — The Indian Jour. Stat.*, vol. 1 (1933), pág. 109.

[21] NAYER, P. N., "An investigation into the application of Neyman and Pearson's L_1 test", *Stat. Res. Memoirs*, vol. 1 (1936), pág. 38.

Daí $L_0 = \frac{98.22}{107} = .918$. A tabela nos dá o valor de L_0 correspondente ao nível de 5% para $k = 3$ e $n = 20$ como *0.8450;* donde se conclui que as amostras são heterogêneas.

Aplicando o segundo teste, temos $L_1 = \frac{98.22}{101} = .972$, enquanto que a tabela nos fornece o valor crítico *0.8980*. Segue-se que também as variâncias, quaisquer que sejam os aproveitamentos médios, diferem significativamente entre si.

7.15 Os testes L_1 *generalizado e* μ.

No caso geral, em que os tamanhos das amostras diferem, temos que

$$L_1 = \frac{\prod_t (s_t^2)^{\frac{n_t}{N}}}{\frac{1}{N} \sum_t (n_t s_t^2)}, \qquad (7.14)$$

isto é, L_1 é a razão da média geométrica ponderada das variâncias das amostras para a sua média aritmética. NAYER estudou a aproximação dessa estatística por uma curva pearsoniana. Seria impossível calcular tabelas para tôdas as combinações possíveis dos valores k e n_t; mas êle mostrou que podiam ser utilizadas as tabelas de L_1, calculadas na hipótese da constância dos n_t, fazendo-se a entrada com k e $\bar{n} = \frac{n_t}{k}$, isto é, com o tamanho médio das amostras. Admite-se que essa aproximação seja válida se nenhum dos n_t fôr menor do que 20, ou mesmo 15.

Quando o número de observações n_t de cada amostra é grande, NEYMAN e PEARSON mostraram que a distribuição de L_1 se confundia pràticamente com a de χ^2, se usarmos a transformação $L_1 = e^{-\frac{\chi^2}{N}}$ e entrarmos nas tabelas com $k - 1$ graus de liberdade. Os níveis de significância de L_1 podem então ser obtidos introduzindo os respectivos valores de χ^2 nessa equação. Notemos que a distribuição de L_1 só depende então de $N = \Sigma n_t$ e k, e não dos valores individuais de n_t. Essa aproximação é satisfatória para $n > 60$, isto é, além dos limites da tabela de NAYER.

Outro teste para verificar a homogeneidade de diversas variâncias foi proposto por BARTLETT [22]. Consideremos k populações normais com variância σ_t^2 $(t = 1, 2, \ldots k)$. Seja s_t^2 uma estimativa justa de σ_t^2 baseada em f_t graus de liberdade, e F o número total de graus de liberdade $F = \sum_t f_t$. A estatística μ de BARTLETT é dada por

$$-2 \log_e \mu = F \log_e \left\{ \sum_t (f_t s_t^2) / F \right\} - \sum_t f_t \log_e s_t^2 .$$

[22] BARTLETT, M. S., "Properties of sufficiency and statistical tests", *Proc. Roy. Soc. London*, ser. A, vol. 160 (1937), pág. 273.

Designemos o segundo membro dessa equação por x. Verifica-se que a distribuição de x aproxima-se da de χ^2, e BARTLETT sugere que o teste se faça com a estatística $\frac{x}{C}$, onde $C = 1 + \frac{1}{3\,(k-1)}\left(\sum_t \left(\frac{1}{f_t}\right) - \frac{1}{F}\right)$, entrando-se na tabela de χ^2 com $k - 1$ graus de liberdade. Como $C > 1$, dispensa-se o seu cálculo, se χ^2 para $C=1$ é significante.

Exemplo. Além dos carvões A e B, referidos no Ex. 7.12, foi examinado mais um carvão C, realizando-se 10 ensaios com os resultados: $\overline{x}_3 = 8.2$, $s^2_3 = 12.560$. Verificar se as três variâncias divergem significativamente.

Aplicando o teste μ, teriamos:

Carvão	s_t^2	f_t	$f_t s_t^2$	$\log_e s_t^2$	$n_t \log_e s_t^2$
A	18.398	6	110.388	2.9177	17.506
B	3.476	5	17.380	1.2459	6.230
C	12.560	9	113.040	2.5306	2.775
Total	—	20	240.808	—	46.511

Daí $x = 02 \log_e \frac{24.808}{20} - 46.511 = 3.185$. Para 2 graus de liberdade, o valor crítico de χ^2 a 5% é 5.991. Não é preciso, portanto, calcular o divisor C; as variâncias dos três tipos de carvão não diferem significativamente.

Capítulo VIII

A ANÁLISE DA VARIÂNCIA

8.1 A análise da variância.

A variação de uma certa grandeza pode-se, geralmente, atribuir a um certo número de causas principais, às quais se superpõem as ações de uma multidão de pequenas causas fortuitas. Por exemplo, o rendimento de uma certa variedade agrícola depende da variação da fertilidade do sólo, ou da diferente adubação aplicada ao mesmo, e também de várias causas acidentais, como a dissemelhança das sementes, os fatôres climáticos, etc. A qualidade de um certo produto industrial varia com o tipo de máquina ou a técnica de produção empregada, e ainda com outros fatôres acidentais, dependentes do operador, condições de trabalho, etc.

O processo da análise da variância, estabelecido por R. A. Fisher [1], tem precisamente por objetivo analisar a variação total de um material heterogêneo segundo as variações componentes devidas a fatôres independentes, possibilitando assim segregar aquêles que produzem variações significantes.

8.2 A decomposição da variância.

Consideremos o caso mais simples, da existência de um único fator principal. Seja N o número total de observações, classificadas em k classes de h elementos cada uma, segundo a intensidade dêsse fator, isto é, $N = hk$. Representemos por X_{ij} a medida da característica do i^{esimo} indivíduo da j^{esima} classe, por X_j a média da j^{esima} classe, por $\overline{X}$ a média geral. Podemos dispor os dados segundo a forma tabelar

$$\begin{array}{ccccc} X_{11} & \dots & X_{1j} & \dots & X_{1k} \\ \dots & \dots & \dots & \dots & \dots \\ X_{i1} & \dots & X_{ij} & \dots & X_{ik} \\ \dots & \dots & \dots & \dots & \dots \\ X_{h1} & \dots & X_{hj} & \dots & X_{hk} \end{array}$$

Temos, então, que a variação total dos elementos X_{ij} pode ser decomposta como segue

$$\sum_{j=1}^{k}\sum_{i=1}^{h}(X_{ij}-\overline{X})^2 = \sum_{j=1}^{k}\sum_{i=1}^{h}\left[(X_{ij}-\overline{X}_j)+(\overline{X}_j-\overline{X})\right]^2 = \sum_{1}^{k}\sum_{1}^{h}(X_{ij}-\overline{X})^2 +$$

$$+ 2\sum_{1}^{k}\sum_{1}^{h}(X_{ij}-\overline{X}_j)(\overline{X}_j-\overline{X}) + \sum_{1}^{k}\sum_{1}^{h}(\overline{X}_j-\overline{X})^2 = \sum_{1}^{k}\sum_{1}^{h}(X_{ij}-\overline{X}_j)^2 + h\sum_{1}^{k}(\overline{X}_j-\overline{X})^2, \qquad (8.1)$$

---

[1] Fisher, R. A., "On a distribution yielding the error functions of several well-known statistics", *Proc. Int. Mathem. Congress* (Toronto, 1924), pág. 805, e artigos subseqüentes; *Statistical Methods for Research Workers*, cap. VII.

o têrmo produto anulando-se na somação. Operou-se assim a decomposição em duas parcelas, a primeira representando a variação de cada elemento em tôrno da média da respectiva classe, a segunda a variação ponderada, segundo o número de elementos nas classes, da média das mesmas em tôrno da média geral. Chama-se à primeira componente de *variação intra-classes*, à segunda de *variação inter-classes*. Essa última componente representa a influência do fator principal considerado, a outra o efeito dos demais fatôres, aleatórios ou estranhos àquele fator.

Suponhamos que os X_{ij} representem observações independentes de uma população normal homogênea, de variância σ^2. Então, $\frac{\sum_{1}^{k}\sum_{1}^{h}(X_{ij}-\overline{X})^2}{hk-1}$ será, como sabemos, uma estimativa justa dessa variância, baseada em *hk—1* graus de liberdade.

Calculemos as expectâncias do segundo membro da identidade *(8.1)*. Temos, denotando por μ a média da população,

$$E(\overline{X}_j-\overline{X})^2 = E\left[\frac{1}{h}\sum_{i=1}^{h}(X_{ij}-\mu)-(\overline{X}-\mu)\right]^2$$

$$= E\left[\frac{1}{h^2}\sum_{i=1}^{h}(X_{ij}-\mu)^2 - \frac{2(\overline{X}-\mu)}{h}\sum_{i=1}^{h}(X_{ij}-\mu)+(\overline{X}-\mu)^2\right]$$

Ora, $E(X_{ij}-\mu)^2=\sigma^2$, $E(\overline{X}-\mu)(X_{ij}-\mu)=\frac{\sigma^2}{hk}$, $E(\overline{X}-\mu)=\frac{\sigma^2}{rs}$,

e portanto $E(\overline{X}_j-\overline{X}) = \frac{\sigma^2}{h}-\frac{2\sigma^2}{hk}+\frac{\sigma^2}{hk}=\frac{\sigma^2}{h}\left(\frac{k-1}{k}\right)$

Daí se segue que $E\left\{h\sum_{j=1}^{k}(\overline{X}_j-\overline{X})^2\right\}=\sigma^2(h-1)$

Assim, $\frac{h\,\Sigma(\overline{X}_j-\overline{X})^2}{k-1}$ é uma estimativa justa de σ^2. Como os X se distribuem normalmente, também $\overline{X}_j$, que depende apenas de uma soma dêles, terá essa distribuição; e a distribuição da grandeza considerada obtém-se conforme exposto no capítulo anterior. Ela é, com efeito, uma estimativa da variância da população baseada em *k—1* graus de liberdade, os quais resultam das K classes consideradas, um grau de liberdade tendo sido absorvido na estimativa de $\overline{X}$.

Quanto ao outro têrmo, temos

$$E(X_{ij}-\overline{X}_j)^2 = E\left[(X_{ij}-\mu)-(\overline{X}_j-\mu)\right]^2 = E(X_{ij}-\mu)^2 - 2E(X_{ij}-\mu)(\overline{X}_j-\mu)+E(\overline{X}_j-\mu)^2.$$

Ora, $E(X_{ij}-\mu)(\overline{X}_j-\mu)=\frac{\sigma^2}{h}$, $E(\overline{X}_j-\mu)=\frac{\sigma^2}{h}$,

donde $E(X_{ij}-\overline{X}_j)^2=\sigma^2\frac{h-1}{h}$ e portanto $E[\Sigma\Sigma(X_{ij}-\overline{X}_j)^2]=\sigma^2k(h-1)$

Conclui-se que $\frac{\Sigma\Sigma(X_{ij}-\overline{X}_j)^2}{k(h-1)}$ também é uma estimativa justa de σ^2, cuja distribuição nos é conhecida, por se tratar de função linear de variáveis normalmente distribuídas. Essa estimativa baseia-se em *k (h—1)* graus de liberdade, pois temos k classes com *h—1* graus de liberdade cada, sendo que um grau é absorvido na estimativa da média da classe.

Podemos, por conseqüência, escrever a identidade (8.1) sob a forma $(hk-1)\sigma^2 = k(h-1)\sigma^2 + (k-1)\sigma^2$, ou seja

$$hk - 1 = k(h-1) + (k-1) \qquad (8.2)$$

o que mostra que os graus de liberdade gosam da mesma propriedade aditiva que a soma de quadrados.

8.3 Significância de um conjunto de médias.

Os resultados precedentes podem ser sumariados sob a forma tabelar seguinte:

NATUR. VAR.	*Soma quadrados*	*Graus liberdade*	*Variância*
Inter-classes.....	$h\Sigma(\overline{X}_j - \overline{X})^2$	$k-1$	s_1^2
Intra-classes.....	$\Sigma\Sigma(X_{ij} - \overline{X}_j)^2$	$k(h-1)$	s_2^2
Total.....	$\Sigma\Sigma(X_{ij} - \overline{X})^2$	$hk-1$	

Se a amostra com que lidamos provém de uma população homogênea normal, de variância σ^2, cada uma dessas somas de quadrados, dividida pelos respectivos graus de liberdade, fornece uma estimativa justa dessa variância. Demonstra-se [2] que essas estimativas são independentes, pelo que podemos utilizar a teoria da distribuição de z ou F para verificar se diferem significantemente entre si.

Nessas considerações se apoia o teste da análise da variância. Mesmo lidando com uma população homogênea, é natural que as diversas estimativas de σ^2 difiram por erros de amostragem. O cálculo das estatísticas z ou F permitirá descriminar se há diferenciação significante entre as mesmas. Se a probabilidade do valor z ou F encontrado é muito pequena, rejeita-se a hipótese da homogeneidade do material.

A comparação faz-se relativamente à variação intra-classes, s_2^2. Com efeito, ela compreende tôdas as causas aleatórias e fatôres secundários que influem no fenômeno, depois de isolado o fator principal; isto é, a sua natureza é a mesma dos erros de observação, inerentes a tôdas as medidas físicas. Daí também ser denominada essa componente de *erro* ou *variação residual*. A estatística $F = s_1^2/s_2^2$ será, pois, o elemento aferidor da homogeneidade dos dados observacionais.

Se o teste é negativo, a heterogeneidade pode provir da variação das médias, ou da diversa variabilidade dos elementos nas classes, ou de ambos. Em geral, admite-se que a variância nas classes é constante, hipótese que, aliás, podemos verificar pelo teste L_1. Então, a análise da variância importa em verificar a significância de um conjunto de médias, constituindo assim uma generalização do teste t.

8.4 Simplificação de cálculos.

Cada uma das expressões que figura no quadro da análise da variância pode ser calculada diretamente, a partir de sua definição. Contudo, o cálculo da variação residual é trabalhoso, e costuma-se obtê-lo por diferença entre a total e a inter-classes.

---

[2] FISHER, R. A., "Aplications of "Student's distribution", *Metron*, vol. 5 (1925), pág. 97, COCHRAN, W. G., "The distribution of quadratic forms in a normal system", *Proc. of the Cambridge Phil. Soc.*, vol. 30 (1933-34), pág. 178.

Na prática, os cálculos são simplificados recorrendo-se à identidade seguinte $\Sigma (X - \overline{X})^2 = \Sigma X^2 - N\overline{X}^2 = \Sigma X^2 - T^2/N$, onde T representa a soma dos valores da variável para as N observações, isto é, $T = \Sigma\Sigma X_{ij}$.

Aplicando essa identidade, obtém-se as fórmulas simplificadas

$$\sum_i \sum_j (X_{ij} - \overline{X})^2 = \sum_j \sum_i X_{ij}^2 - \frac{T^2}{N}$$

$$h \sum_j (\overline{X}_j - \overline{X})^2 = \frac{\sum_j T_j^2}{h} - \frac{T^2}{N} \qquad (8.3)$$

$$\sum_j \sum_i (X_{ij} - \overline{X}_j)^2 = \sum_j \sum_i X^2 - \frac{\sum_j T_j^2}{h}.$$

onde T_j é o total dos valores da coluna j.

Podemos também referir a variável a uma origem arbitrária, ou dividí-la por uma constante arbitrária c, de modo a simplificar o cálculo dos quadrados e somas. Nestes casos, após a aplicação das fórmulas *8.3*, devem-se multiplicar as somas de quadrados resultantes por c^2.

Exemplo. Para verificar a qualidade de cimentos de 5 procedências, fizeram-se ensaios de ruptura em cinco briquetes de cada um, obtendo-se os resultados (em lbs.) do quadro (i). Verificar se as médias de carga de ruptura diferem significantemente.

(i)

A	B	C	D
518	508	554	555
500	574	598	567
538	528	579	560
510	534	538	538
544	538	544	540

(ii)

	A	B	C	D
	18	8	54	55
	0	74	98	67
	38	28	79	50
	10	34	38	38
	44	38	44	40
Totais.....	110	182	313	247

Considerando 500 como origem arbitrária, obtemos o quadro (ii). A soma total de valores é $T = 852$, e o têrmo corretivo $T^2/N = (852)^2/20 = 31290$. Tomando os quadrados dos elementos na tabela (ii), vêm $\Sigma\Sigma X^2 = \{(18)^2 + (8)^2 + \ldots + (40)^2\} = 47708$, donde se conclui a variação total $\Sigma\Sigma (X_{ij} - \overline{X})^2 = 47708 - 31295 = 11413$. Quanto à variação inter-classes, temos $\Sigma_j T_j^2/h = \frac{1}{5}\{(110)^2 + (182)^2 + (313)^2 + (247)^2\} = 30840$, e portanto $h \Sigma (\overline{X}_j - \overline{X})^2 = 4545$.

Temos assim o quadro da análise da variância:

NATUR. VAR.	Soma quadrados	Gráus lib.	Variância
Inter-classes.....	4545	3	1515
Residual.....	6868	16	429
Total.....	11413	19	

Calculemos a estatística $F = 1515/729 = 3.88$. Das tabelas se vê que o valor de F para o nível de 5% é 3.24, donde se conclui que a resistência média à ruptura dos diversos cimentos difere significantemente.

8.5 Tabelas com classes desiguais.

A análise da variância pode ser aplicada a tabelas, em que o número de elementos varia de uma classe a outra. Seja n_j o número de elementos na classe j ; então, a identidade *(8.1)* transforma-se em

$$\Sigma\Sigma(X_{ij} - \overline{X})^2 = \Sigma\Sigma(X_{ij} - \overline{X}_j)^2 + \Sigma_j n_j(\overline{X}_j - \overline{X})^2. \qquad (8.4)$$

As fórmulas *(8.3)* modificam-se consoantemente, devendo-se substituir nelas $\Sigma_j T_j^2/h$ por $\Sigma_j(T_j^2/n_j)$.

8.6 Análise da variância segundo duas componentes.

A análise da variância exposta pode ser fàcilmente estendida ao caso em que os elementos estão subordinados a dois critérios de classificação. Suponhamos, com efeito, que a disposição dos elementos segundo as linhas se faça obedecendo a um segundo critério, traduzindo a influência de outro fator principal.

A influência dêsse fator, que na análise precedente estava englobada na variação residual, deve agora ser isolada. Consideremos as expectâncias dos elementos X_{ij} relativamente às médias das colunas, isto é, os valores $X_{ij} - \overline{X}_j$. Imaginemo-los dispostos segundo um retângulo, do mesmo modo que anteriormente dispusemos os X_{ij}, e tratemo-los anàlogamente. Denotemos por $\overline{X}_{.j}$ a média da coluna j esima, por $\overline{X}_{i.}$ a média da lina i esima; isto é, o ponto denota a somação para todo os elementos que êle substitui.

Prevalece então a identidade

$$\begin{aligned}\Sigma\Sigma(X_{ij} - \overline{X}_{.j})^2 &= \Sigma\Sigma\left[(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{X}) + (\overline{X}_{i.} - \overline{X})\right]^2 \\ &= \Sigma\Sigma(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{X})^2 + k\Sigma_i(\overline{X}_{i.} - \overline{X})^2, \qquad (8.5)\end{aligned}$$

em vista do têrmo produto anular-se na somação.

A expectância do têrmo $k\Sigma_i(X_{i.} - \overline{X})^2$ é, de conformidade com a exposição anterior, $E\left\{k\Sigma_i(X_{i.} - \overline{X})^2\right\} = \sigma^2(h-1)$, de modo que $\frac{k\Sigma(\overline{X}_{i.} - \overline{X})^2}{h-1}$ é uma estimativa justa da variância da população σ^2, baseada em $h-1$ graus de liberdade.

Vejamos o têrmo resíduo. Temos

$$\begin{aligned}E(X_{ij} - \overline{X}_{.j} + \overline{X}_{i.} + \overline{X})^2 &= E\left\{(X_{ij} - \mu) - (\overline{X}_{.j} - \mu) - (X_{i.} - \mu) + \overline{X} - \mu)\right\}^2 \\ &= E\left\{(X_{ij} - \mu)^2 + (\overline{X}_{.j} - \mu)^2 + (\overline{X}_{i.} - \mu)^2 + (\overline{X} - \mu)^2 - 2(X_{ij} - \mu)(\overline{X}_{.j} - \mu)\right. \\ &\quad - 2(X_{ij} - \mu)(\overline{X}_{i.} - \mu) + 2(X_{ij} - \mu)(\overline{X} - \mu) - 2(\overline{X}_{.j} - \mu)(\overline{X} - \mu) \\ &\quad \left. - 2(\overline{X}_{i.} - \mu)(\overline{X} - \mu) + 2(\overline{X}_{.j} - \mu)(\overline{X}_{i.} - \mu)\right\}\end{aligned}$$

Recordando as expectâncias já calculadas, e que

$$E(X_{.j} - \mu)(\overline{X} - \mu),\ E(\overline{X}_{i.} - \mu)(\overline{X} - \mu) \quad \text{e} \quad E(\overline{X}_{.j} - \mu)(\overline{X}_{i.} - \mu)$$

são todos iguais a $\frac{\sigma^2}{hk}$, temos finalmente que

$$E(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{X}) = \sigma^2\left(1 - \frac{1}{h} - \frac{1}{k} - \frac{1}{hk}\right) = \frac{\sigma^2(h-1)(k-1)}{hk}$$

Por conseqüência, $\frac{E\left\{\Sigma\Sigma(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{X})^2\right\}}{(h-1)(k-1)}$ também constitui uma estimativa justa de σ^2, baseada em *(h—1) (k—1)* graus de liberdade. É fácil ver que êsses graus de liberdade são a diferença entre os totais *hk—1* e os absorvidos pelos dois outros têrmos calculados, isto é,

$$(hk - 1) = k(h - 1) + h(k - 1) + (h - 1)(k - 1). \qquad (8.6)$$

O quadro da análise da variância com dois critérios de classificação assume, pois, a forma.

NATUR. VAR.	*Soma quadrados*	*Gráus lib.*	*Variância*
Colunas	$h\Sigma_j(\overline{X}_{.j} - \overline{X})^2$	$k - 1$	s_1^2
Linhas	$k\Sigma_i(\overline{X}_{i.} - \overline{X})^2$	$h - 1$	s_2^2
Residual	$\Sigma\Sigma(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{X})^2$	$(h-1)(k-1)$	s_3^2
Total	$\Sigma\Sigma(X_{ij} - \overline{X})^2$	$hk - 1$	

Na hipótese de perfeita homogeneidade do material, tôdas as somas de quadrados, divididas pelos respectivos graus de liberdade, dão estimativas justas da variância da população. Para contrastar essa hipótese, comparam-se as variâncias devidas aos dois princípios de classificação, s_1^2 e s_2^2, com a residual s^2_3, devida a causas aleatórias ou estranhas àquêles princípios. Entrando nas tabelas de *z* ou *F*, verifica-se até que ponto as diferenças entre essas estimativas se podem considerar como fortuitas ou reais.

Exemplo. (SANDERS). Numa experiência para confrontar o rendimento de variedades de milho, utilizaram-se seis replicações e obtiveram-se os resultados seguintes (em lbs.):

BLOCOS	*VARIEDADES*					
	A	*B*	*C*	*D*	*E*	*Total*
1	*82.1*	*70.2*	*81.1*	*79.4*	*88.4*	*396.2*
2	*85.7*	*82.8*	*84.5*	*82.6*	*89.4*	*425.0*
3	*85.0*	*84.8*	*77.9*	*84.4*	*85.7*	*417.8*
4	*86.6*	*68.0*	*78.7*	*78.2*	*86.6*	*398.1*
5	*77.4*	*73.1*	*76.0*	*75.8*	*78.0*	*380.6*
6	*71.6*	*63.8*	*74.3*	*79.6*	*78.0*	*367.3*
Total	*484.4*	*448.0*	*472.5*	*480.0*	*501.1*	*2885.0*
Média	*81.40*	*78.83*	*78.75*	*80.0*	*88.52*	*79.5*

A soma total dos valores é 2385.0, e o têrmo corretivo 189,607.5. Pôsto isto, calculamos:

(a) soma dos quadrados para todo sos rendimentos $(82.1)^2 + (85.7)^2 + \ldots + (78.0)^2 = 190{,}667.24$, donde, deduzida a correção, temos que a soma dos quadrados das discrepâncias é 1,059.74;

(b) soma dos quadrados dos totais das variedades, dividida por 6, igual a 189,923.5, donde, deduzida a correção, temos a soma dos quadrados das discrepâncias para as variedades como 316.00;

(c) soma dos quadrados dos totais dos blocos, dividida por 5, igual a 190,081.11, donde, deduzida a correção, temos a soma dos quadrados das discrepâncias para os blocos como 473.61.

Com êsses elementos, formamos o quadro da análise da variância:

NATUR. VAR.	*Soma quadr.*	*Gráus lib.*	*Variância*	$\frac{1}{2} \log_e$
Variedades.....	*316.00*	*4*	*79.00*	*2.1847*
Blócos.....	*473.61*	*5*	*94.72*	*2.2754*
Residual.....	*270.13*	*20*	*13.51*	*1.3017*
Total.....	*1059.74*	*29*	—	*.8830*

Para verificar a significância dos resultados, usaremos neste caso o teste z, calculando para tal, a última coluna. Apenas temos de considerar o valor de z referente a variedades e residual, pois a classificação segundo blocos teve por finalidade isolar o fator da heterogeneidade do solo da componente do êrro. Então, $z = 2.1847/1.3017 = .8830$. Entrando na tabela com $n_1 = 4$, $n_2 = 20$, vemos que o valor de z no nível de 5% é .5265 e no de 1% é .7443. O valor encontrado é, pois, altamente significante, isto é, os rendimentos das diferentes variedades diferem significantemente.

8.7 Subdivisão da variância em mais de duas componentes.

O processo exposto pode-se generalizar, permitindo a subdivisão da variância em mais de duas componentes, além da residual. Sempre se decompõe a variação total $\Sigma\Sigma\,(X_{ij} - \overline{X})^2$ em um certo número de formas quadráticas que, divididas pelos correspondentes graus de liberdade, fornecem estimativas justas da variância da população. Porém, à medida que se introduzem novas componentes, a complexidade das fórmulas aumenta muito.

Atentemos para o caso de três componentes principais. Dispomos de $N = h\,k\,m$ valores observados, sujeitos a uma tríplice classificação segundo grupos, colunas e linhas. Sejam m os grupos, de k colunas e h linhas cada um, e denotemos por x_{ijl} o elemento constituinte da linha i da coluna j do grupo l.

Os valores serão distribuídos no quadro seguinte:

Colunas	*1*	*k*
Grupo i.....	X_{111}	X_{1k1}
	.....	
	X_{h11}	X_{hk1}
.....		
.....		
Grupo m.....	X_{11m}	X_{1km}
	.....	
	X_{h1m}	X_{hkm}

Despresando por um instante a circunstância das linhas estarem reunidas segundo um certo número de grupos, consideremos a tabela como consistindo de hm linhas. A análise da variância comum daria então

NATUR. VAR.	*Gráus lib.*
Colunas.....	$k - 1$
Linhas (tot.).....	$hm - 1$
Resíduo (tot.).....	$(k - 1)(hm - 1)$
Total.....	$hkm - 1$

Na realidade, os $hm - 1$ graus de liberdade atribuídos às linhas (tot.) podem-se decompor num outro quadro da análise da variância, levando em conta a subdivisão em grupos; temos

NATUR. VAR.	*Gráus lib.*
Linhas.....	$h - 1$
Grupos.....	$m - 1$
Interação.....	$(h - 1)(m - 1)$
Linhas (tot.).....	$hm - 1$

Aparece aqui uma noção nova, a da *interação linhas x grupos*, que representa a ação simultânea de dois fatôres principais sôbre o fenômeno em estudo. Isto é, a presença do fator G pode modificar a ação do outro fator L sôbre o fenômeno; é êsse *efeito diferencial* que aparece representado na análise da variância pelo têrmo interação $L \times G$.

Por sua vez, o têrmo resíduo (tot.) representa a interação entre estes três últimos elementos considerados e as colunas. Daí podermos decompô-lo como segue:

NATUR. VAR.		*Gráus lib.*
Interação	$C \times L$	$(k - 1)(h - 1)$
"	$C \times G$	$(k - 1)(m - 1)$
"	$C \times L \times G$	$(k - 1)(h - 1)(m - 1)$
Resíduo (tot.)....		$(k - 1)(hm - 1)$

O último têrmo, interação $C \times L \times G$, engloba o efeito dos três fatôres além das flutuações casuais; é o equivalente da variação residual na análise a dois componentes, e serve de têrmo de comparação para avaliar a significância dos fatôres considerados.

No cálculo prático de uma análise da variância segundo três critérios de classificação serão utilizadas as mesmas simplificações de cálculos expostas no § 3.4.

Observemos, finalmente, que temos exposto os princípios de classificação binária ou múltipla supondo a freqüência constante em cada uma das classes. Caso contrário, faz-se mister a introdução de certas modificações de cálculo, cuja exposição excede o âmbito dêste trabalho.[3]

[3] Cf. YATES, F., "The analysis of multiple classifications with unequal numbers in the different classes", *Jour. Amer. Stat. Assoc.*, vol. 29 (1934), pág. 51.

Capítulo IX

VERIFICAÇÃO DA INTERDEPENDÊNCIA ENTRE FENÔMENOS ANÁLISE DA COVARIÂNCIA

9.1 *Significância do coeficiente de correlação.*

A interdependência entre fenômenos pode ser medida, ou pelos coeficientes de correlação ou pelos coeficientes de regressão. Ambos estão sujeitos a flutuações por amostragem, e importa estabelecer testes permitindo a verificação de sua significância.

Consideremos, em primeiro lugar, a significância do coeficiente de correlação pearsoniano r, calculado sôbre uma amostra de N pares de valores (X,Y). Para verificar se êsse coeficiente difere significantemente de zero, devemos calcular a probabilidade de que êle possa provir, por flutuações de amostragem, de uma população caracterizada pelo parâmetro $\varrho = 0$. Si essa população tem uma distribuição $f(X,Y)$ *normal,* então a distribuição por amostragem de r é dada por

$$f(r) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\pi^{1/2}\,\Gamma\left(\frac{N-2}{2}\right)}\,(1-r^2)^{\frac{N-4}{2}}$$

As curvas correspondentes são simétricas em tôrno de $r = 0$, e tem-se

$$\sigma_r = (N-1)^{-1/2} \quad , \quad \alpha_3 = 3 - \frac{6}{N-1} \, .$$

Para valores elevados de N, a função torna-se pràticamente normal, e podemos pois referir a expressão

$$t = r\,(N-1)^{1/2} \tag{9.1}$$

à tabela das áreas da curva normal.

Para pequenos valores de N, a aproximação deixa de ser válida. Façamos, porém, a transformação $r = t/n^{1/2}\left(1 + \frac{t^2}{n}\right)^{1/2}$, $n = N - 2$. Obtém-se

$$f(r)\,dr = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{1/2}\,\Gamma\left(\frac{n}{2}\right)}\left[1 - \frac{t^2}{n\left(1+\frac{t^2}{n}\right)}\right]^{\frac{N-4}{2}} \frac{dt}{n^{1/2}\left(1+\frac{t^2}{n}\right)^{3/2}}$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{(\pi n)^{1/2}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \, ,$$

que é a distribuição de Student (6.7), com $n = N - 2$ graus de liberdade.

Portanto, para verificar se a população donde provém a amostra é correlacionada ou não, utilizamos a transformação inversa

$$t = \frac{r n^{\frac{1}{2}}}{(1 - r^2)^{\frac{1}{2}}} \quad , \quad n = N - 2 \; . \tag{9.2}$$

e utilizamos a tabela da distribuição de STUDENT.

Da mesma forma, si tivermos um coefiente de correlação parcial $r_{12 \cdot 34 \ldots k}$ baseado em N pares de valores, a verificação de sua significância pode ser obtida pela transformação precedente e a utilização da tabela de distribuição de STUDENT com $n = N - k$.

R. A. FISHER [1] calculou tabelas permitindo a aplicação direta dêste teste, às quais dão os valores de r para os níveis de significância de $P = .10, .05, .02,$ e $.01$, e para vários valores de n. Um exame dessas tabelas evidencia a precariedade da classificação costumeira do coeficiente de correlação como desprezível ($r < .30$); sensível ($.30 < r < .50$); médio ($.50 < r < .70$) e forte ($r > .70$). Um coeficiente de 0.30 é perfeitamente válido no nível de significância de 5%, para amostras de 40 elementos, enquanto um de 0.80 não o será para amostras de 4 elementos.

Aproveitemos êsses resultados par aconfrontar a validade da aproximação normal, quando se trata de pequenas amostras. Para $N=20$, $1/\sqrt{N-1} = .229$, e o coeficiente de correlação 0.449 corresponde ao nível de significância de 5%; a tabela de FISHER dá o valor exato de r para $N-2 = 18$ graus de liberdade como 0.444. Para $N = 10$ e $N = 5$, temos respectivamente, com a aproximação normal, os valores de r no nível de significância de 5% como 0.664 e 0.980, enquanto que a tabela de FISHER nos dá os valores exatos 0.632 e 0.878. Como se vê, a aproximação normal não conduz a erros exagerados, mesmo para amostras de 10 elementos; mas abaixo dêsse limite, ela exagera as condições de significância.

Exemplo. Num estudo biométrico feito por PEARL sôbre 292 inglêses adultos, obteve-se o coeficiente de correlação $+.12$ entre a estatura e o pêso do cérebro. Verificar se êle é significante, e, caso não seja, de que tamanho deveria ser a amostra para torná-lo altamente significante.

Usando a transformação (9.2) temos $t = \frac{.12}{.9928}\ 17.029 = 2.058$. As tabelas nos dão, no nível de 5%, $t = 1.958$, e no de 1%, $t = 2.593$. O coeficiente de correlação encontrado é significante, mas não altamente significante (nível 1%). Para alterar sua significância, o tamanho da amostra deveria ser $\sqrt{N-2} = 2.593\ (.9928/.12) = 21.4$, donde $N = (214)^2 + 2 = 460$.

9.2 *Aplicação da análise da variância*

A questão poderia, também, ser ventilada à luz da análise da variância. Representemos por Y' a estimativa da variável Y obtida mediante a equação de regressão linear $Y' = a + bX$. Veremos adiante (§ 9.5) que a variação total de Y em tôrno da média geral $\overline{Y}$ é igual à variação da reta de regressão em tôrno da média, mais a variação residual em tôrno daquela reta, isto é,

$$\Sigma (Y - \overline{Y})^2 = \Sigma (Y' - \overline{Y})^2 + \Sigma (Y - Y')^2,$$

[1] FISHER, R. A., *Statistical Methods for Research Workers*, tab. V A.

e que os graus de liberdade correspondentes à primeira e segunda parcelas são respectivamente *1* e *N-2*. Ora,

$$\Sigma (Y' - \overline{Y})^2 = b^2 \Sigma (X - \overline{X})^2 = \Sigma (X - \overline{X})(Y - \overline{Y}) / \Sigma (X - \overline{X})^2 = r^2 \Sigma (Y - \overline{Y})^2,$$

portanto $\Sigma (Y - Y')^2 = (1 - r^2) \Sigma (Y - \overline{Y})^2$.

A análise da variância correspondente será, pois,

NATUR. VAR.	*Soma quadrados*	*G. L.*
Regressão.....	$r^2 \Sigma (Y - \overline{Y})^2$	*1*
Residual.....	$(1 - r^2) \Sigma (Y - \overline{Y})^2$	$N - 2$
Total.....	$\Sigma (Y - \overline{Y})^2$	$N - 1$

Comparando a variação devida à regressão com a residual, estamos, *ipso facto,* verificando a significância do coeficiente de correlação. Daí o teste $F = \frac{r^2}{1 - r^2} (N - 2)$, que, referimos à tabela da distribuição da análise da variância, com $n_1 = 1$ e $n_2 = N\text{-}2$ graus de liberdade. Ora, vimos (§ 7.11) que, para $n_1 = 1$, a distribuição de z se reconduz à de STUDENT com a transformação $z = \frac{1}{2} \log_e t^2$, o que se consegue correlativamente com a distribuição F mediante a transformação $F = t^2$. Recaímos destarte na expressão (9.2).

9.3 O caso das populações correlacionadas ($\varrho \neq 0$).

Se a população se caracteriza por um coeficiente de correlação ϱ, a distribuição de ϱ obtida por R. A. FISHER [2] em 1915, assume uma forma muito mais complicada

$$\phi(r) = k (1 - \rho^2)^{\frac{1}{2}(N-1)} (1 - r^2)^{\frac{1}{2}(N-4)} \frac{d^{N-2}}{d(r\rho)^{N-2}} \left(\frac{\arccos(-r\rho)}{\sqrt{1 - r^2 \rho^2}} \right).$$

As curvas correspondentes são assimétricas, e nalguns casos de *talho-U*. Compreende-se a existência da assimetria, e que ela cresça com ϱ. Suponhamos o universo com $\varrho = .8$; a amplitude de variação de r acima de ϱ é apenas *0.2*, enquanto que abaixo de ϱ é *1.8*.

Para grandes amostras e baixos valores de ϱ a distribuição tende para a forma normal, com média igual a ϱ e desvio padrão $\sigma_r = (1 - \varrho^2)\ (N - 1)^{-\frac{1}{2}}$. Daí a prática habitual de interpretar a estatística $t = (r - \varrho)/\sigma_r$ mediante a tabela de áreas da curva normal.

Um estudo extensivo da distribuição de r e de sua normalização, foi realizado cooperativamente por SOPER e outros,[3] mediante o cálculo dos coeficientes β_1 e β_2. As conclusões são que as condições de normalidade, $\beta_1 = 0$ e $\beta_2 = 3$, não são satisfeitas para amostras de 25, e mesmo 50 elementos, qualquer que seja o valor de ϱ. Para amostras de 100, a aproximação é aceitável para baixos valores de ϱ, isto é, $\varrho < .5$, mas

[2] FISHER, R. A., "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population", *Biometrika*, vol. 10 (1915), pág. 507.

[3] SOPER, H. E., and OTHERS, "On the distribution of the correlation coefficient in small samples", *Biometrika*, vol. 11 (1915-17), pág. 328.

inválida para valores superiores. Para amostras de 400 elementos, a normalização melhora, mas há ainda desvios sensíveis para $\varrho \geqslant .8$. Tais resultados evidenciam os perigos da interpretação de r mediante a fórmula habitual de seu êrro padrão, referido à tabela da curva normal.

No estudo citado, figura um conjunto de tabelas das ordenadas da distribuição de r para valores de ϱ de 0 a 1. Outras tabelas, e a integral da função de distribuição, foram calculadas por DAVID[4] para $\varrho = .1, .2, \ldots, .9$, e $n = 3, 4, \ldots, 25, 50, 100, 200$ e 400.

R. A. FISHER[5] mostrou que a transformação da tangente hiperbólica

$$z' = \frac{1}{2} \log_e \left(\frac{1 + r}{1 - r} \right), \quad \xi = \frac{1}{2} \log_e \left(\frac{1 + \rho}{1 - \rho} \right), \tag{9.3}$$

produz uma distribuição que, mesmo para amostras com $N=20$, é aproximadamente normal, com média ξ e desvio padrão $\sigma_{z'} = (N - 3)^{-1/2}$, independente pois de ϱ.

Dêste modo, para verificar se um dado valor de r difere significantemente do valor hipotético da população ϱ, calcula-se a estatística

$$t = (z' - \xi)(N - 3)^{1/2}, \tag{9.4}$$

que é referida à tabela da curva normal.

Se se trata de um coeficiente de correlação parcial, a mesma transformação é utilizada, mas o seu desvio padrão é $\sigma_{z'} = (N - m - 3)^{-1/2}$ onde m o número de variáveis eliminadas; isto é, para o coeficiente $r_{12 \cdot 34 \ldots k}$, adota-se $t = (z' - \xi)\ (N - k - 1)^{1/2}$.

O processo estende-se ao caso da verificação da significância da diferença entre dois coeficientes de correlação. Se r_1 e r_2 são calculados sôbre amostras de tamanhos N_1 e N_2 respectivamente, obtém-se os valores transformados z'_1 e z'_2. O desvio padrão da diferença $d = z'_1 - z'_2$ será $\sigma_{z'_1 - z'_2} = \left(\sigma^2_{z'_1} + \sigma^2_{z'_2} \right)^{1/2} = \left(\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3} \right)^{1/2}$, e a entrada na tabela da curva normal faz-se com a estatística

$$t = (z'_1 - z'_2) \left(\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3} \right)^{-1/2} \tag{9.5}$$

Notemos que os testes da significância de r foram estabelecidos sem aplicação da correção de SHEPPARD, e portanto estas não devem ser utilizadas no cálculo do coeficiente de correlação, para o fim de verificar a sua significância, pois elas tendem a aumentar o valor de r.

Observemos também que êstes testes se baseiam na normalidade da população donde derivam as amostras. Há, porém, estudos experimentais evidenciando que êles prevalecem, dentro de limites práticos razoáveis, desde que as distribuições marginais de uma ou ambas as variáveis não sejam de *talho-J* ou *U*. Mas, nesses casos, é a própria validade do uso de r como medida de associação que se torna questionável.

---

[4] DAVID, F. N., *Tables of the Distribution of the Correlation Coefficient* (Londres, *Biometrika Office*, 1938).

[5] FISHER, R. A., "On the "probable error" of a coefficient of correlation deduced from a small sample", *Metron*, vol. 1 (1921), pág. 1.

Exemplo. Num estudo biométrico, PEARL determinou os coeficientes de correlação entre o pêso do cérebro e a estatura de mulheres suécas e tchécas, obtendo respectivamente +. 345 numa amostra de 253 indivíduos da primeira nacionalidade, e + . 216 numa amostra de 128 da segunda. Verificar se a diferença entre os dois valores de r é significante.

Temos, no primeiro caso, $z'_1 = .3598$, e no segundo, $z' = .2195$, donde a diferença $.1403$. O êrro padrão dessa diferença é $\sqrt{\frac{1}{250} + \frac{1}{125}} = .1095$. Daí $t = .1403/.1095 = 1.281$, que não é significante. Desvios numèricamente maiores obter-se-iam ao acaso em 20 vêzes sôbre 100.

9.4 Combinação de estimativas homogêneas do coeficiente de correlação.

A transformação z' vai-nos permitir combinar diversos coeficientes de correlação, calculados sôbre amostras extraídas de populações caracterizadas pelo mesmo coeficiente de correlação ϱ de modo a se obter uma estimativa melhor dêsse parâmetro.

Sejam $r_1, r_2, \ldots r_k$ os k coeficientes procedentes de amostras de tamanhos $N_1, N_2, \ldots N_k$ respectivamente. Feita a transformação $z'_i = \frac{1}{2} \log_e \frac{1 + r_i}{1 - r_i}$, $i = 1, 2, \ldots k$, toma-se a média ponderada dos z'_i, sendo os pesos as respectivas variâncias; logo, o valor médio dos z'_i será $z' = \frac{\Sigma (N_i - 3) z'_i}{\Sigma (N_i - 3)}$. Dêsse valor retorna-se à estimativa do coeficiente de correlação mediante a transformação inversa

$$r = \left(e^{2z'} - 1\right) / \left(e^{2z'} + 1\right) = \tanh z' .$$

Uma precaução, contudo, deve ser observada antes de se combinarem os coeficientes; urge verificar se êles realmente provêm de populações homogêneas relativamente ao coeficiente ϱ. Cada um dos r_i é uma estimativa de um coeficiente populacional ϱ_i, e a hipótese a ser verificada é que os $\varrho_i = \varrho$, ou seja que $z'_i = \xi_i = \xi$, sendo ϱ e ξ constantes $(i = 1, 2, \ldots k)$.

Sob essa hipótese, temos k grandezas independentes z'_i, distribuídas de modo aproximadamente normal com média ξ e variância $\sigma^2_{z'_i} = \frac{1}{N_i - 3}$

A estimativa da variância comum é dada por

$$(k-1)'s'^2 = \Sigma (N_i - 3)(z'_i - z')^2 = \Sigma (N_i - 3) z'^2_i - \frac{[\Sigma (N_i - 3) z'_i]^2}{\Sigma (N_i - 3)}. \qquad (9.6)$$

Então, a grandeza $\chi^2 = (k-1)'s'^2/\sigma^2$ tem a distribuição χ^2 com $k-1$ graus de liberdade. Como no caso $\sigma^2 = 1$, a expressão (9.6) é o próprio χ^2, o qual permitirá verificar se a variação dos z'_i é devida apenas a flutuações de amostragem ou não. Se o teste fôr negativo, podemos fazer a combinação dos coeficientes.

Exemplo. Em três anos sucessivos (1925-27), o coeficiente de correlação entre o volume de vendas semanais de peras Bartlett em Nova Iorque e o preço foi o seguinte: —.75, —.80 e —.70, tendo as observações abrangido 14, 16 e 13 semanas de estação respectivamente. Verificar se os três resultados diferem significantemente, e, caso não, combiná-los numa estimativa melhorada.

Com êsses elementos, organizamos o quadro seguinte:

z'_i	$N_i - 3$	$(N_i - 3)z'_i$	$(N_i - 3)z'^2_i$
.9730	11	10.703	10.414
1.0986	13	14.282	15.696
.8673	10	8.673	7.522
	34	33.658	33.632

Daí $'s^2 = 33.632 - 1132.99/34 = .309$. Para $n = 2$, temos que χ^2 no nível de 5% é 5.99, pelo que o nosso χ^2 não é significante e as três estimativas podem-se considerar como homogêneas.

A estimativa melhorada será então $z^1 = \{ (11 \times .9730) + (13 \times 1.9086) + (10 \times .8673) \} /34 = .9899$, donde $r = .757$.

9.5 *Aplicação da análise da variância à regressão linear.*

Consideremos uma série de pares de valores da variável dependente Y e da independente X, tais que a regressão de Y sôbre X se possa representar pela equação linear $Y' = a + bX$. Ora, a soma dos quadrados dos desvios de Y em relação à média geral Y pode ser decomposta como segue:

$$\Sigma (Y - \overline{Y})^2 = \Sigma [(Y - Y') + (Y' - \overline{Y})]^2.$$

O têrmo produto vai anular-se, pois que

$$\Sigma (Y - Y')(Y' - \overline{Y}) = \Sigma (Y - a - bX)(a + bX - \overline{Y})$$

$$= (a - \overline{Y}) \Sigma (Y - a - bX) + b \Sigma X (Y - a - bX),$$

e ambos os têrmos são nulos em vista das equações normais sôbre que baseamos o cálculo da regressão. Temos, por conseguinte, que

$$\Sigma (Y - \overline{Y})^2 = \Sigma (Y - Y')^2 + \Sigma (Y' - \overline{Y})^2.$$

A primeira parcela representa os desvios em relação à função de regressão, pois, para cada valor de Y, consideramos o quadrado de sua discrepância para o valor Y' calculado segundo essa função; a segunda parcela representa a função de regressão porque, para cada valor de Y, tomamos o quadrado da discrepância entre o valor estimado Y' e a média geral $\overline{Y}$.

Ao têrmo $\Sigma (Y - Y')^2$ correspondem $N - 2$ graus de liberdade, pois as discrepâncias são calculadas a partir de uma reta, cuja fixação absorve 2 graus de liberdade. Quanto ao segundo têrmo, temos que

$$\Sigma (Y' - \overline{Y})^2 = \Sigma (a + bX - \overline{Y})^2 = b^2 \Sigma (X - \overline{X})^2.$$

Como $\Sigma (X - X)^2$ é independente da correlação, qualquer variação em $\Sigma (Y' - Y)^2$ é devida ùnicamente a b, o que mostra que êsse têrmo, para uma dada distribuição de Y, depende apenas da estatística b e representa 1 grau de liberdade.

Podemos assim organizar o quadro da análise da variância:

NATUR. VAR.	*Soma quadrados*	*G. L.*
Regressão	$b^2 \Sigma (X - \overline{X})^2$	*1*
Residual.....	$\Sigma (Y - Y')^2$	$N - 2$
Total.....	$\Sigma (Y - Y)^2$	$N - 1$

Para verificar a significância da regressão linear usamos então a estatística

$$F = \frac{b^2 \Sigma (X - \overline{X})^2 (N - 2)}{\Sigma (Y - Y')^2}, \qquad (9.7)$$

ou, alternativamente, tomamos a raiz quadrada dessa expressão e reportamo-nos à tabela da distribuição de STUDENT, com $n = N - 2$ graus de liberdade.

A questão pode ser apreciada sob um ponto de vista mais geral. Suponhamos que se quer verificar a hipótese da compatibilidade da equação de regressão obtida com a regressão hipotética da população $Y'_\infty = \alpha + \beta x$, admitindo que a origem das abscissas coincide com a média. A decomposição faz-se agora segundo os têrmos

$$\Sigma (Y - Y'_\infty)^2 = \Sigma \left[(Y - a - bX) + (a - \alpha) + (b - \beta)\, x \right]^2$$

$$= (Y - a - bx)^2 + N (a - \alpha)^2 + (b - \beta)^2 \Sigma x^2,$$

os têrmos produtos anulando-se. Daí o quadro da análise da variância:

NATUR. VAR.	*Soma quadrados*	*G. L.*
Têrmo constante...	$N (a - \alpha)^2$	*1*
Têrmo 1.º gráu.....	$(b - \beta)^2 \Sigma x^2$	*1*
Residual.....	$\Sigma (Y - a - bx)^2$	$N - 2$
Total.....	$\Sigma (Y - \alpha - \beta x)^2$	$N - 1$

Calculamos as estatísticas F, para verificar a significância do têrmo constante e do 1.º grau isoladamente, ou então, tomando-os em conjunto,

$$F = \frac{\left[N (a - \alpha)^2 + (b - \beta)^2 \Sigma x^2 \right] (N - 2)}{2 \Sigma (Y - a - bX)^2}, \qquad (9.8)$$

a significância da regressão observada em relação à regressão hipotética da população. Notemos que neste caso, $n_1 = 2$, $n_2 = N - 2$, e não podemos pois usar o teste t.

9.6 Aplicação à correlação curva ou múltipla.

O processo exposto generaliza-se fàcilmente para os casos de regressão curva ou múltipla. Consideremos o caso de uma regressão parabólica. Quando verificamos a regressão linear, decompusemos a variação total em uma parcela devida a esta regressão e outra residual, a qual inclui, além dos efeitos fortuitos, possìvelmente regressões de ordem superior. Seja R_1, êsse primeiro resíduo, isto é, $R_1 = \Sigma(Y - Y')^2$. A análise da variância corresponde a

NATUR. VAR.	*Soma quadrados*	*G. L.*
Regressão linear....	$\Sigma(Y' - \overline{Y})^2$	1
Resíduo.....	$R_1 = \Sigma(Y - Y')^2$	$N - 2$
Total.....	$R_0 = \Sigma(Y - \overline{Y})^2$	$N - 1$

Se introduzirmos um segundo têrmo na função de regressão, isto é, tomando-a como $Y'' = a_2 + b_2 X + c_2 X^2$, êsse resíduo R_1, se decomporá em dois outros têrmos: um representando a variação da parábola em tôrno da reta de regressão, o outro a segunda variação residual R_2 ; e a análise será

NATUR. VAR.	*Soma quadrados*	*G. L.*
Regressão paraból..	$\Sigma(Y'' - Y')^2$	1
Resíduo.....	R_2	$N - 3$
Total.....	R_1	$N - 2$

Essa decomposição continua indefinidamente, permitindo a cada passo verificar a significância do têrmo adicional introduzido. Tanto que o resultado do teste é positivo, o têrmo introduzido representa uma peculiaridade real do fenômeno. Evidentemente, para-se quando o novo têrmo não atinge à significância, ficando então englobado com os resíduos devidos a causas aleatórias.

No caso de regressão múltipla, $Y' = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k$, a decomposição da variação total compreendendo um têrmo representativo da função de regressão, e os resíduos em tôrno da mesma. Tem-se, em virtude de propriedades conhecidas, representando por R^2 o coeficiente de correlação múltipla, que

NATUR. VAR.	*Soma quadrados*	*G. L.*
Regressão.....	$\Sigma(Y' - \overline{Y})^2 = R^2 \Sigma(Y - \overline{Y})^2$	k
Resíduos.....	$\Sigma(Y - Y')^2 = (1 - R^2)\Sigma(Y - \overline{Y})^2$	$N - k - 1$
Total.....	$\Sigma(Y - \overline{Y})^2$	$N - 1$

O teste de significância da regressão múltipla importa, pois, em verificar a significância do coeficiente R. Para tal fim, temos a estatística

$$F = \frac{R^2}{(1 - R^2)} \cdot \frac{N - k - 1}{k}, \qquad (9.9)$$

entrando-se na tabela de F com os graus de liberdade $n_1 = k$ e $n_2 = N - k - 1$.

9.7 Significância da razão de correlação.

A razão de correlação η_{yx}, definida como $\eta^2_{yx} = \frac{\Sigma_j\, n_j\,(Y_j - \overline{Y})^2}{\Sigma_j \Sigma_i (Y_{ij} - \overline{Y})^2}$.

mede a variabilidade total das médias das colunas, independentemente da natureza linear ou não da relação entre as variáveis. Temos que $\Sigma\Sigma(Y_{ij} - \overline{Y})^2 = \Sigma\, n_j\,(\overline{Y}_j - \overline{Y})^2 + \Sigma\Sigma(Y_{ij} - \overline{Y}_j)^2$, e daí o quadro da análise da variância:

NATUR. VAR.	*Soma quadrados*	*G. L.*
Inter-colunar.....	$\Sigma\, n_j(\overline{Y}_j - \overline{Y})^2 = \eta^2 \Sigma\Sigma(Y - \overline{Y})^2$	$k - 1$
Intra-colunar.....	$\Sigma\Sigma\,(Y_{ij} - \overline{Y}_j)^2 = (1 - \eta^2)\,\Sigma\Sigma\,(Y - \overline{Y})^2$	$N - k$
Total.....	$\Sigma\Sigma\,(Y - \overline{Y})^2$	$N - 1$

Para verificar a existência da correlação, ou, o que é o mesmo, se o η^2 difere significantemente de zero, usamos a estatística

$$F = \frac{\eta^2}{1 - \eta^2} \cdot \frac{N - k}{k - 1} \qquad (9.10)$$

9.8 Teste de linearidade da regressão.

O teste de lineridade da regressão equivale a verificar se η^2 difere significantemente de r^2. Para tal fim, subdividimos a variação intercolunar em duas parcelas, correspondendo a primeira à soma dos quadrados dos desvios das médias das colunas relativamente à regressão linear, a segunda à soma dos quadrados dos desvios devidos a essa última regressão. Temos assim

$$\Sigma\, n_j(\overline{Y}_j - \overline{Y})^2 = \Sigma\, n_j(\overline{Y}_j - Y')^2 + \Sigma\, n_j(Y' - \overline{Y})^2.$$

Se as médias das colunas caem sôbre a linha de regressão, $\Sigma\, n_j\,(Y_j - Y')^2$ anula-se, ao passo que o seu valor aumenta à medida que a linha de tendência das médias colunares se afasta da regressão linear. Tomando a variação intracolunar, devida a flutuações casuais, como têrmo de comparação, podemos verificar a linearidade da regressão.

O quadro da análise da variância toma a forma:

NATUR. VAR.	*Soma quadrados*	*G. L.*
Média col./regressão	$\Sigma n_j(\overline{Y}_j - Y')^2 = (\eta^2 - r^2)\,\Sigma\Sigma\,(Y_{ij} - \overline{Y})^2$	$k - 2$
Regressão.....	$\Sigma n_j(Y' - \overline{Y})^2 = r^2 \quad \Sigma\Sigma\,(Y_{ij} - \overline{Y})^2$	1
Resíduo/média col...	$\Sigma\Sigma\,(Y_{ij} - \overline{Y}_j)^2 = (1 - \eta^2)\,\Sigma\Sigma\,(Y_{ij} - \overline{Y})^2$	$N - k$
Total.....	$\Sigma\Sigma\,(Y_{ij} - \overline{Y})^2$	$N - 1$

Daí a estatística

$$F = \frac{\eta^2 - r^2}{1 - \eta^2}\,\frac{N - k}{k - 2}. \qquad (9.11)$$

Se não tiver sido calculado o coeficiente η^2, podemos verificar a linearidade da regressão diretamente, usando $F = \dfrac{\Sigma n_j(Y_j - Y')^2}{\Sigma\Sigma\,(Y_{ij} - \overline{Y}_j)^2}\,\dfrac{N - k}{k - 2}$

Para calcular as diversas somas de quadrados, partimos da identidade $\Sigma n_j(\overline{Y}_j - \overline{Y})^2 = \Sigma\left(\dfrac{T_j}{n_j}\right)^2 - \dfrac{T^2}{N}$. Em seguida calcula-se $\Sigma n_j\,(Y' - Y)^2$, e por diferença obtém-se o numerador. Quanto ao denominador, resulta por diferença entre $\Sigma\Sigma(Y - \overline{Y})^2$ e $\Sigma n_j\,(Y_j - \overline{Y})^2$.

Êste teste, que é exato, substitui o habitual teste de linearidade de BLACKEMAN, que é errôneamente utilizado, pois a distribuição de $(\eta^2 - r^2)$ nem para grandes amostras tende para a normalidade.

Exemplo. Num estudo sôbre a relação entre o rendimento unitário do trigo e o custo de produção em 216 fazendas nos Estados Unidos, os resultados foram grupados em 7 classes de rendimento, e obtiveram-se o coeficiente de correlação linear $r = -.642$ e a razão de correlação $\eta = .7574$. Pede-se verificar a linearidade da regressão.

Temos $\eta^2 - r^2 = .4988 - .4522 = .0866$, e portanto $F = \dfrac{.0866}{.5012}\,\dfrac{209}{7} = 5.112$.

Para os graus de liberdade $n_1 = 7$ e $n_2 = 200$, o valor de F no nível de 5% é 2.05. Por conseguinte, os dados não satisfazem a condição de linearidade, e não deve ser utilisado o coeficiente r.

9.9 A análise da covariância.

A precisão dos experimentos estatísticos aumenta, se conseguimos igualar as fontes de êrro incidentes sôbre os fatôres principais ou *tratamentos*. Suponhamos uma experiência para determinar a eficiência de dois métodos de ensino; se as turmas são escolhidas de modo que a cada aluno da primeira corresponde outro tendo obtido num teste inicial o mesmo escore, é óbvio que eliminamos as perturbações oriundas das diferenças individuais de aprendizagem anterior. Nem sempre, porém,

é possível obter um emparelhamento dos grupos experimentais. Por exemplo, num experimento agrícola, não se pode eliminar totalmente a diferença de fertilidade entre os diversos lotes de um bloco, ou as diferenças de rendimento oriundas do diverso número de plantas por unidade de área. Devemos, contudo, levar em conta êsses fatôres, que não podem ser igualados satisfatòriamente. É o que se consegue com a *análise da covariância*. O seu escopo é, portanto, verificar a homogeneidade do material em problemas envolvendo duas ou mais variáveis correlacionadas. Atendendo à natureza das diferenças iniciais existentes, ela permite ajustar convenientemente as médias dos grupos que se vão comparar pela análise da variância.

A covariância da população define-se como $E\,(X-\mu_1)\;(Y-\mu_2)$, denotando por μ_1 e μ_2 as médias das duas variáveis. A estimativa da covariância a partir de uma dada amostra basear-se-á na soma de produtos $\Sigma\;(X-\overline{X})\;(Y-\overline{Y})$. O processo focalizado funda-se, em suma, na possibilidade de decompôr essa soma de produtos em componentes tal como decompusemos a soma de quadrados. Obtém-se, dest'arte, estimativas da covariância, e ainda dos coeficientes de correlação e regressão, dos quais se isolam os efeitos representados pelos diversos critérios de classificação adotados.

9.10 A decomposição da covariância.

Consideremos os resultados de observações simultâneas sôbre duas variáveis X, Y, dispostos em k classes com h elementos cada. Contém cada célula um par de valores X_{ij}, Y_{ij}; representemos por $\overline{X}_{ij}, \overline{Y}_{ij}$ as médias de X e Y respectivamente na classe j^{esima}.

CLASSES	*1* ..... *k*
	X_{11}, Y_{11} ..... X_{1k}, Y_{1k}
	.....
	.....
	X_{h1}, Y_{h1} ..... X_{hk}, Y_{hk}
Médias	$\overline{X}_1, \overline{Y}_1$ ..... $\overline{X}_k, \overline{Y}_k$

Podemos então decompôr a soma de produtos como segue:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}(X_{ij}-\overline{X})(Y_{ij}-\overline{Y}) = \Sigma\Sigma\,[(X_{ij}-\overline{X}_j)+(\overline{X}_j-\overline{X})]\,[(Y_{ij}-\overline{Y}_j)+(\overline{Y}_j-\overline{Y})]$$

$$= \Sigma\Sigma\,(X_{ij}-\overline{X}_j)(Y_{ij}-\overline{Y}_j) + h\,\Sigma_j(\overline{X}_j-\overline{X})(\overline{Y}_j-Y), \qquad (9.13)$$

pois que se anulam os produtos da forma $\Sigma\Sigma(\overline{X}_j-\overline{X})\;(Y_{ij}-\overline{Y}_j) = \Sigma_j(\overline{X}_j-\overline{X})\;h\;(\overline{Y}_j-\overline{Y}_j)=0$.

Dêste modo, a soma de produtos das discrepâncias em relação às médias gerais é igual à soma dos produtos das discrepâncias intra-classes mais a soma de produtos das discrepâncias das médias dos grupos, multiplicada pelo número de elementos nas classes. A êsses dois têrmos correspondem os graus de liberdade respectivamente de $k\,(h-1)$ e $k-1$. Daí podermos organizar o quadro da análise da covariância:

NATUR. VAR.	*G. L.*	$\Sigma(X-\bar{X})^2$	$\Sigma(X-\bar{X})(Y-\bar{Y})$	$\Sigma(Y-\bar{Y})^2$	b_{yx}	$b_{yx}\,\Sigma(X-\bar{X})(Y-\bar{Y})$
Blócos	h	A_0	B_0	C_0		
Tratamento	k	A_1	B_1	C_1	$b_1 = B_1/A_1$	$b_1 B_1$
Erro	n	A_2	B_2	C_2	$b_2 = B_2/A_2$	$b_2 B_2$
$T+E$	$n+k$	A_t	B_t	C_t	$b_t = B_t/A_t$	$b_t B_t$

Êsse quadro nos fornece, além da estimativa da regressão baseada na totalidade dos dados, as correspondentes a blocos, tratamentos e erros, mas a que nos interessa é a b_2, que traduz a regressão dos valores observados sôbre os iniciais, após eliminação do efeito dos blocos e tratamentos.

A significância dêsse coeficiente de regressão será verificada pelos processos anteriormente indicados. A soma de quadrados devida à regressão é $b_2 B_2$, e a das discrepâncias relativamente à regressão $C_2 - b_2 B_2$, com os graus de liberdade respectivamente de 1 e $n-1$.

Daí o teste $F = \dfrac{b_2 B_2 (n-1)}{C_2 - b_2 B_2}$.

9.11 Ajustamento de valores na análise da variância.

Podemos agora corrigir os valores observados Y segundo a regressão observada, isto é, tomando os valores ajustados $Y - b\,(X - \bar{X})$. Essas são as melhores estimativas do que se presume que as médias teriam sido, caso não tivesse havido a perturbação devido à variável independente. Como a primitiva variância devida ao êrro dos dados experimentais contém uma componente devida à regressão, a soma de quadrados adequada para verificar a significância dos valores ajustados é a residual, após eliminação da parte devida à regressão, isto é $C_2 - b_2\ B_2$ com $n-1$ graus de liberdade.

Além disso, como o coeficiente de regressão está êle mesmo sujeito a erros de amostragem, os valores ajustados tem precisão variável, o que deve ser levado em conta ao obter a estimativa justa da soma de quadrados devida a tratamentos. Consideremos os valores inscritos na linha $T+E$; a estimativa da soma de quadrados mais êrro é C_t; dessa grandeza devemos abater $b_t\ B_t$, que representa a quantidade com a qual essa soma de quadrados, inclusive a regressão, está inflacionada, devido aos erros do coeficiente de regressão. Deduzindo, pois, de $C_t - b_t\ B_t$ a parte relativa ao êrro, teremos a justa estimativa referente aos tratamentos.

A análise toma assim a forma:

NATUR. VAR.	*G. L.*	*Soma quadrados*	*Variância*
Tratamento	q	$C_1 + b_2 B_2 - b_t B_t$	V_1
Erro	$n-1$	$C_2 - b_2 B_2$	V_2
$T+E$	$n+q-1$	$C_t - b_t B_t$	

O teste de significância da diferença de tratamentos, após ajustamento devido à regressão dos Y sôbre os X, importa, pois, na comparação das variâncias V_1 e V_2.

Se as diferenças entre as médias da variável independente são pequenas, podemos utilizar um êrro padrão comum para tôdas as comparações; caso contrário, deve-se calcular um êrro para cada diferença entre as médias dos valores ajustados, de modo a levar em conta a variabilidade de X. A variância da diferença entre as médias Y_1 e Y_2 será igual a

$$s^2 \left\{ \frac{2}{h} + \frac{(\overline{X}_1 - \overline{X}_2)^2}{\Sigma (X - \overline{X})^2} \right\}, \qquad (9.13)$$

onde s^2 é a variância correspondente ao êrro na tabela da análise da variância, e $\overline{X}_1$, $\overline{X}_2$ são as médias usadas no cálculo de Y_1 e Y_2.

ÍNDICE