INTERPRETAÇÃO QUANTITATIVA DE RESULTADOS EXPERIMENTAIS

L. T. Atalla

INFORMAÇÃO IEA 60 COURP - ARQ 7	MAIO/1978

INF. IEA 60
COURP - ARQ 7

MAIO/1978

INTERPRETAÇÃO QUANTITATIVA DE RESULTADOS EXPERIMENTAIS

L. T. Atalla

CENTRO DE OPERAÇÃO E UTILIZAÇÃO DO REATOR DE PESQUISA
Área de Radioquímica

INSTITUTO DE ENERGIA ATÔMICA
SÃO PAULO – BRASIL

APROVADA PARA PUBLICAÇÃO EM JANEIRO/1978.

INSTITUTO DE ENERGIA ATÔMICA
Caixa Postal 11.049 (Pinheiros)
Cidade Universitária "Armando de Salles Oliveira"
SÃO PAULO — BRASIL

ÍNDICE

CAPÍTULO VII

CAPÍTULO VIII

INTERPRETAÇÃO QUANTITATIVA DE RESULTADOS EXPERIMENTAIS

L. T. Atalla

RESUMO

Um dos pontos fundamentais de qualquer trabalho experimental, seja no campo da química, física, biologia, etc. é a maneira de apresentar e analisar os resultados. Um trabalho cujos resultados não são analisados e interpretados quanto à sua precisão e exatidão, por exemplo, não pode ser considerado completo, porque não fornece ao leitor uma idéia objetiva da credibilidade que pode ser atribuída aos valores apresentados.

Além disso, quando se desenvolve um método experimental, em qualquer campo, é necessário fazer uma comparação com métodos já padronizados e destinados à mesma finalidade para poder salientar o mérito do novo método.

Estas considerações exigem do pesquisador um conhecimento básico da aplicação de testes estatísticos. Deve-se notar também que o conhecimento desses testes permite não só a interpretação dos resultados obtidos, mas ajuda a estabelecer planos de trabalho de modo racional. Os planos previamente organizados podem reduzir o tempo de execução do trabalho que se pretende desenvolver e, ao mesmo tempo, facilitam a interpretação posterior dos resultados obtidos.

A interpretação quantitativa de resultados experimentais não é, em geral, tratada em profundidade nos cursos de graduação. É freqüente encontrar pesquisadores com o simples conhecimento do cálculo da propagação de erros, do cálculo da média e desvio padrão. Deve-se lembrar que os testes estatísticos têm cunhos peculiares a cada campo de aplicação: engenharia, química, geodésia, etc. Eis porque tratados e cursos sobre Matemática Estatística nem sempre respondem objetivamente às necessidades dos trabalhos nos vários campos de aplicação.

Procurou-se, em vista do que foi exposto, condensar neste trabalho os aspectos mais importantes da aplicação da estatística ao trabalho experimental. A finalidade não foi a apresentação de um trabalho com considerações matemáticas extensas, mas o suficiente para dar ao leitor um conhecimento básico para a aplicação correta dos testes estatísticos aos dados colhidos no laboratório.

Compuseram-se e resolveram-se vários exercícios para tornar mais clara a compreensão e aplicação da teoria. Embora estes exercícios sejam todos aplicados a problemas químicos, o raciocínio e os testes usados na sua resolução podem ser extendidos a qualquer tipo de pesquisa que dependa de coleta de dados.

A autora espera, com este trabalho, difundir a importância da interpretação dos resultados experimentais em trabalhos de pesquisa e agradece qualquer informação sobre erros que possam ter passado despercebidos.

CAPÍTULO I

NOÇÕES DE PROBABILIDADE

O conceito de probabilidade é muito antigo e seu primeiro desenvolvimento foi simplesmente intuitivo. Mesmo aplicado intuitivamente, o conceito de probabilidade se baseia em fatores de ordem quantitativa que podem ser usados como argumento para a intensidade da convicção. Por exemplo, se um médico afirma que a probabilidade de recuperação de um doente é grande, ele se baseia em experiências anteriores, mesmo que não possa atribuir um número a essa expectativa.

A probabilidade, vista de um modo quantitativo, pode ser aplicada a duas situações bem diferentes. Por esse motivo, temos a probabilidade clássica e a probabilidade experimental. A primeira, de utilidade limitada, se aplica a situações em que a ocorrência de um certo evento pode ser prevista por uma análise, a priori, do número de eventos que podem ocorrer com a mesma possibilidade. É o caso dos jogos de azar. Por exemplo, no jogo de um dado, cujas faces são numeradas de 1 a 6, a probabilidade da ocorrência do número 1 é igual à probabilidade da ocorrência do número 2 ou de qualquer um dos outros números previamente especificados. Pode-se afirmar então que a probabilidade de dar o número 1 é 1/6. A segunda, chamada probabilidade experimental, se aplica a situações onde o número de eventos igualmente possíveis não pode ser determinado. É o caso de todas as situações da vida real, em que se pode aplicar o conceito de probabilidade. Deve-se notar que, nessas situações, quando se considera um número grande de eventos, observa-se um certo grau de regularidade. Com uma definição, a posteriori, da probabilidade experimental, podem ser feitas previsões de futuros eventos e, além disso, é possível atribuir a essas previsões um certo grau de segurança.

Chamam-se aleatórios os tipos de eventos que podem ser interpretados pela análise estatística e pela teoria da probabilidade; são eventos que não podem ser previstos individualmente mas sim a partir de um todo. Praticamente, todas as disciplinas que dependem de relações quantitativas — economia, sociologia, biologia, química, física etc — estão relacionados com fenômenos aleatórios.

Em particular, a arma de toda a ciência experimental é a medida e praticamente todas as medidas podem ser incluídas na classe dos fenômenos chamados aleatórios. As medidas devem ser assim classificadas por um (ou ambos) dos seguintes motivos:

a) cada medida é uma "amostra" de um grande número de medidas possíveis que diferem mais ou menos entre si. Em outras palavras, o pesquisador é incapaz de controlar todos os fatores que provocam essas diferenças;

b) a grandeza que está sendo medida pode conter, como característica própria, um certo grau de aleatoriedade.

I.1 — Probabilidade Clássica

Num jogo sem vícios, a chance de ganhar pode ser facilmente deduzida. Por exemplo, a probabilidade de tirar um az de copas de um baralho completo (52 cartas) é de 1/52. As condições para esse cálculo simples de probabilidade são:

a) deve-se conhecer o número total de possibilidades;

b) cada resultado posssível deve ter a mesma chance. Chamando-se de n o número total de resultados possíveis e w o número de resultados que conduzem à condição de "ganhar", a probabilidade de ganhar é definida por:

$$p = \frac{w}{n} \tag{I.1}$$

e a probabilidade de perder por:

$$q = \frac{n - w}{n} \tag{I.2}$$

A equação (I.1) constitue a definição clássica de probabilidade. Esta definição não fornece um critério para achar os valores de w e n e há situações em que essa determinação é complexa. Como exemplo, vamos imaginar o jogo de uma moeda por duas vezes sucessivas e considerar o jogo como ganho, se aparecer pelo menos uma cara. Vamos chamar as caras de C e as coroas de K. Um argumento para o cálculo da probabilidade seria considerar o jogo perdido só na condição KK, isto é, coroa nas duas moedas e ganho nas condições CC, KC e CK. Neste caso, $p = 3/4$. Existe outro argumento que afirma que para os eventos CC e CK o jogo já está ganho na primeira jogada e neste caso a segunda não é necessária. De acordo com este argumento, só existem três alternativas possíveis C, KC e KK, sendo que os dois primeiros resultados conduzem à condição de ganhar. Para este caso $p = 2/3$. Qual é o correto? Este paradoxo se resolve se atribuirmos peso 2 ao evento C do segundo argumento, porque C, KC e KK não são igualmente prováveis, pois C tanto pode ser CK como CC. Consequentemente, o primeiro argumento é o correto.

Este é um exemplo simples, mas há condições em que é difícil achar os números w e n corretos.

I.2 – Combinação de Probabilidades

Sejam A, B, C, . . . resultados que conduzem à condição de "ganhar", onde A pode ser cara de uma moeda, B pode ser valete de paus de um baralho, C pode ser carta vermelha de um baralho, etc. Um dos eventos pode ser mais complexo, como por exemplo, os números 5 ou 6 de um dado. Os eventos compostos requerem uma análise cuidadosa para o cálculo da probabilidade. Duas combinações envolvem eventos mutuamente exclusivos ou eventos independentes.

Dois eventos são **mutuamente exclusivos**, se só um deles pode ocorrer numa jogada. Se A e B são dois eventos mutuamente exclusivos temos que:

$$p(A \text{ ou } B) = p(A) + p(B) \qquad (I.3)$$

Por exemplo, no jogo de um dado, a probabilidade de dar os números de 1 a 6 é mutuamente exclusiva. Neste caso, se o jogo é considerado ganho, se derem os números 5 ou 6, temos que:

$$p(5 \text{ ou } 6) = p(5) + p(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Segue-se daqui que a soma das probabilidades de todos os eventos mutuamente exclusivos é 1, isto é:

$$p(A) + p(B) + \ldots + p(N) = 1 \ldots \qquad (I.4)$$

Dois eventos A e B são **independentes** se a ocorrência ou não de A, não afeta a probabilidade da ocorrência de B. A probabilidade da ocorrência de dois ou mais eventos independentes é o produto das probabilidades individuais, isto é:

$$p(A \text{ e } B \text{ e } C \ldots) = p(A) \,.\, p(B) \,.\, p(C) \ldots \qquad (I.5)$$

Por exemplo, a probabilidade de se obter 3 caras em 3 jogadas de uma moeda é $(1/2)^3$. A probabilidade de se obter duas caras e uma coroa na terceira jogada também é $(1/2)^3$, mas a probabilidade de se obter duas caras e uma coroa em três jogadas, sem especificar a seqüência dos eventos é $(1/2)^3 + (1/2)^3 + (1/2)^3 = 3(1/2)^3$. No último caso temos uma combinação de eventos independentes e eventos mutuamente exclusivos.

A respeito da independência dos eventos, deve-se fazer uma observação quanto à popularmente conhecida "lei das médias". Acredita-se erroneamente que após uma seqüência de resultados negativos deve haver uma seqüência de resultados positivos para haver equilíbrio. Se os eventos são independentes, isto nao faz sentido, porque a probabilidade é sempre a mesma. Por exemplo, se uma pessoa faz 1 jogo de Cr$5,00 toda a semana na loteria esportiva (probabilidade de ganhar é $2/3^{13}$), esta pessoa não pode esperar que em $3^{13}/2$ semanas ganhará uma vez, porque um jogo é independente do outro.

I.3 – Eventos Compostos – Teoremas Gerais de Adição

Vamos considerar agora um conjunto de eventos que se sobrepõem, isto é, que têm probabilidades superpostas. Estes eventos não são mutuamente exclusivos por completo. Por exemplo, se jogarmos dois dados um branco e um azul simultaneamente, qual a probabilidade de dar pelo menos um 6? Podemos ter 6 no dado branco, 6 no dado azul e 6 em ambos. A probabilidade procurada não é a soma das probabilidades dos componentes, isto é, $\frac{1}{6} + \frac{1}{6}$, porque esta soma inclue duas vezes a probabilidade de que ambos sejam o número 6. O problema deve ser encarado da seguinte maneira:

Probabilidade de dar 6 no dado azul e não no branco: 1/6 . 5/6
Probabilidade de não dar 6 no dado azul e sim no branco: 5/6 . 1/6
Probabilidade de dar 6 nos dois dados: 1/6. 1/6
A probabilidade será então:

$$p = 2(\frac{5}{36}) + \frac{1}{36} = \frac{11}{36}$$

que é equivalente a:

$$p = \frac{1}{6} + \frac{1}{6} - (\frac{1}{6} \cdot \frac{1}{6})$$

onde o termo entre parenteses corrige a soma das duas probabilidades quanto à ocorrência simultânea (sobreposição) dos dois eventos.

Dados dois eventos A e B não mutuamente exclusivos, temos:

$$\begin{aligned} p\,(\text{nem A, nem B}) &= [1 - p(A)]\,.\,[1 - p(B)] && (1.6)\\ &= 1 - p(A) - p(B) + p(A)\,.\,p(B) \\ p\,(\text{A ou B, não ambos}) &= p(A)\,[1 - p(B)] + p(B)\,[1 - p(A)] && (1.7)\\ &= p(A) + p(B) - 2p(A)\,.\,p(B) \\ p\,(\text{A ou B ou ambos}) &= p(A)\,.\,p(B) + p(\text{A ou B, não ambos}) && (1.8)\\ &= p(A) + p(B) - p(A)\,.\,p(B) \end{aligned}$$

As equações (I.6), (I.7) e (I.8) são conhecidas como as equações gerais de adição.

I.4 – Probabilidade Condicional – Teorema da Multiplicação

Dados os eventos A e B, se o evento B só pode ocorrer após a ocorrência de A, então a probabilidade de acontecer B deve incluir a probabilidade de ocorrer A. Para este caso, temos:

$$p(A \text{ e } B) = p(A)\, p(B/A) \tag{I.9}$$

Como exemplo, vamos supor duas caixas idênticas: a primeira com uma bola branca, duas azuis e duas pretas e a segunda com quatro bolas brancas, uma azul e uma preta. Se uma pessoa escolhe uma caixa ao acaso e retira uma bola sem olhar, qual a probabilidade de que a bola seja branca?

A probabilidade de que a primeira caixa seja escolhida é 1/2 e a probabilidada de que dela seja retirada uma bola branca é 1/5. Então a probabilidade de que seja retirada uma bola branca da primeira caixa será: 1/2 . 1/5 = 1/10. Do mesmo modo, conclui-se que a probabilidade de tirar uma bola branca da segunda caixa é: 1/2 . 4/6 = 1/3. A resposta do problema será então a soma das duas probabilidades, isto é: 1/3 + 1/10 = 4/30. Outro exemplo: uma pessoa que possue $\underline{a}$ bolas brancas (B) e $\underline{a}$ pretas (P) pega duas ao acaso e coloca numa caixa. Pergunta-se qual a probabilidade de retirar uma bola branca desta caixa?

Podemos presumir que existem três possibilidades quanto às cores das duas bolas: BB (branca e branca), BP e PP. Assim temos:

$$p(BB) = p(BP) = p(PP) = 1/3$$

A probabilidade procurada é a soma das probabilidades de retirar uma bola branca de três caixas hipotéticas (BB, BP e PP), multiplicada por 1/3. Temos então que:

$$p(B) = 1/3\,(1 + \frac{1}{2} + 0) = \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{2}$$

Deve-se salientar que, quando foi atribuída a mesma probabilidade para as três possibilidades de cores das bolas, isto foi feito porque não havia nenhuma informação adicional que pudesse orientar para um julgamento melhor. Neste exemplo, como foi visto anteriormente, pode-se admitir também que as possibilidades sejam quatro, isto é:

$$p(BB) = p(BP) = p(PB) = p(PP) = 1/4$$

e assim temos:

$$p(B) = 1/4\,(1 + \frac{1}{2} + \frac{1}{2} + 0) = \frac{1}{4}\,(2) = \frac{1}{2}$$

que é o mesmo resultado obtido para o caso da três possibilidades.

Ainda no mesmo exemplo, suponhamos que foi tirada uma bola da caixa ao acaso, e verificou-se ser branca. A bola foi reposta na caixa, as duas bolas foram misturadas e depois tirou-se novamente uma bola. Qual a probabilidade de que seja branca?

Como sabemos agora que, pelo menos uma bola é branca, temos só duas possibilidades: (B,B) e (B,P). Vamos chamar de p_o as probabilidades anteriores e p_1 as probabilidades depois de saber que uma bola é branca. Vamos ter:

$$p_1(BB) = \frac{p_o(BB)\ p_1(B/BB)}{p_o(BB)\ p_1(B/BB) + p_o(BP) \cdot p_1(B/BP)}$$

$$p_1(BP) = \frac{p_o(BP) \cdot p_1(B/BP)}{p_o(BB)\ p_1(B/BB) + p_o(BP)\ p_1(B/BP)}$$

onde $p_1(B/BP)$ é a probabilidade de dar bola branca, condicionada à probabilidade de haver uma bola branca e uma preta, visto que a primeira extração deu bola branca.

$$p_1(BB) = \frac{1/3\,.\,1}{1/3\,.\,1 + 1/3\,.\,1/2} = 2/3$$

e

$$p_1(BP) = \frac{1/3\,.\,1/2}{1/3\,.\,1 + 1/3\,.\,1/2} = 1/3$$

Assim, a probabilidade que se tire uma bola branca pela segunda vez será

$$p_2(B) = p_1(BB)\,.\,p(B/BB) + p_1(BP)\ p(B/BP) = 2/3\,.\,1 + 1/3\,.\,1/2 = 5/6$$

Vamos supor que a bola foi reposta na caixa, mais uma vez foi retirada uma bola ao acaso e esta foi branca. A probabilidade de que ambas sejam brancas torna-se cada vez maior e neste caso temos:

$$p_2(BB) = \frac{p_o(BB).\ p_2(B/BB)}{p_o(BB)\,.\,p_2(B/BB) + p_o(BP)\,.\,p_2(B/BP)}$$

$$p_2(BB) = \frac{1/3\,.\,1}{1/3\,.\,1 + 1/3\,.\,1/4} = 4/5$$

$$p_2(BP) = \frac{p_o(BP)\ .\ p_2(B/BP)}{p_o(BP)\,.\,p_2(B/BP) + p_o(BB)\,.\,p_2(B/BB)}$$

$$p_2(BP) = \frac{1/3\ .\ 1/4}{1/3\,.\,1 + 1/3\,.\,1/4} = \frac{1}{5}$$

e a probabilidade de dar bola branca pela terceira vez será:

$$p_3(B) = \frac{4}{5} \cdot 1 + \frac{1}{5} \cdot \frac{1}{2} = \frac{9}{10}$$

Os resultados de todas as extrações são assumidos como sendo independentes. Para a $n^{ésima}$ extração, se as $(n - 1)$ anteriores deram bola branca, se obtém a fórmula:

$$p_n(B) = \frac{1}{1 + (\frac{1}{2})^{n-1}} \cdot 1 + \frac{(\frac{1}{2})^{n-1}}{1 + (\frac{1}{2})^{n-1}} \cdot \frac{1}{2} = \frac{2^n+1}{2^n+2}$$

para qualquer n inteiro > 1 (porque na primeira extração tínhamos 3 hipóteses em vez de duas). Depois de observar bola branca em n extrações $(n > 0)$ a probabilidade para a hipótese BB torna-se:

$$p_n(BB) = \frac{2n}{2n + 1}$$

que se aproxima rapidamente de 1, à medida que n aumenta, mas nunca se torna 1, porque a $(n+1)^{ésima}$ extração pode dar bola preta e a probabilidade de serem duas bolas brancas cai imediatamente para zero. Este exemplo de probabilidade condicional é uma ilustração do problema básico de inferência de conhecimento que se aplica na ciência.

Vimos que depois de n extrações é quase certa a probabilidade da existência de duas bolas brancas. O mesmo acontece para qualquer tipo de conhecimento baseado num número limitado de observações.

Todo o conhecimento numa ciência experimental é obtido por inferência feita a partir dos resultados obtidos. Geralmente, porém, a evidência em que uma generalização científica se apoia é tão complexa que saõ necessários vários resultados não esperados para destruí-la completamente. Uma teoria bem fundamentada, quando depara com um resultado não esperado, em vez de ser abandonada, deve poder ser alterada de maneira a incluir a nova informação como "esperada".

Assim progride o conhecimento por indução e constitue a base do cálculo de toda a probabilidade experimental, como será visto mais tarde.

EXERCÍCIO 1

Duas cartas são tiradas ao mesmo tempo de um baralho. Qual a probabilidade de que:

a) pelo menos uma delas seja espadas?
b) ambas sejam vermelhas?
c) uma seja ás e outra seja 10 preto?

Respostas:

a) Probabilidade de ambas serem espadas:

$$\frac{13}{52} \cdot \frac{12}{51} = \frac{156}{2652} = \frac{12}{204}$$

Probabilidade de uma só ser espadas:

$$\frac{39}{52} \cdot \frac{13}{51} + \frac{13}{52} \cdot \frac{39}{51} = \frac{2 \times 13 \times 39}{52 \times 51} = \frac{78}{204}$$

$$\text{Total: } \frac{78 + 12}{204} = \frac{90}{204} = \frac{15}{34}$$

b) $\frac{26}{52} \cdot \frac{25}{51} = \frac{25}{102}$

c) $\frac{4}{52} \cdot \frac{2}{51} + \frac{2}{52} \cdot \frac{4}{51} = 2 \cdot \frac{1}{12} \cdot \frac{1}{51} = \frac{4}{663}$

EXERCÍCIO 2

Uma caixa contém 6 bolas brancas e 9 bolas vermelhas.

a) Retira-se uma bola sem ver a cor e ela não é reposta. Qual a probabilidade de retirar uma bola branca na próxima extração?

b) Se a primeira bola extraída for branca e não for reposta, qual a probabilidade da segunda ser branca?

c) Se a primeira bola for branca, se não se conhece a cor da segunda bola extraída e se nenhuma das duas foi reposta, qual a probabilidade de retirar uma bola branca na terceira extração?

Respostas:

a) Probabilidade da primeira bola ser branca: 6/15
Probabilidade da segunda bola ser branca, se a primeira for branca: 5/14
Probabilidade da primeira bola ser vermelha: 9/15
Probabilidade da segunda bola ser branca, se a primeira for vermelha: 6/14

$$\text{Total: } \frac{6}{15} \cdot \frac{5}{14} + \frac{9}{15} \cdot \frac{6}{14} = \frac{30 + 54}{210} = \frac{2}{5}$$

b) $\frac{5}{14}$

c) Depois de extraída a primeira bola branca, ficam 5 bolas brancas e 9 vermelhas e o problema é igual a a)

$$\frac{5}{14} \cdot \frac{4}{13} + \frac{9}{14} \cdot \frac{5}{13} = \frac{20 + 45}{182} = \frac{65}{182}$$

EXERCÍCIO 3

As 15 bolas do problema anterior foram distribuídas em duas caixas: a primeira com 4 bolas brancas e 1 vermelha e a segunda com 2 brancas e 8 vermelhas.

a) Qual a probabilidade de tirar uma bola branca de uma caixa escolhida ao acaso?

b) Se não se conhece a cor da primeira bola e ela não é reposta, qual a probabilidade de, numa segunda extração, se obter bola branca?

c) Se a primeira bola for branca e não for reposta, qual a probabilidade de que a segunda bola seja branca?

Respostas:

a) Se a primeira saiu da primeira caixa, a probabilidade de ser branca é:

$$\frac{1}{2} \cdot \frac{4}{5} = \frac{4}{10}$$

Se saiu da segunda caixa é:

$$\frac{1}{2} \cdot \frac{2}{10} = \frac{1}{10}$$

$$\text{Total: } \frac{4}{10} + \frac{1}{10} = \frac{1}{2}$$

b) Probabilidade da primeira bola ser branca = 1/2 = Probabilidade de ser vermelha (Resposta de a)). Esta probabilidade está assim distribuída:

Primeira bola branca e da primeira caixa = 1/2 . 4/5 = 4/10
Primeira bola branca e da segunda caixa = 1/2 . 1/5 = 1/10
Primeira bola vermelha e da primeira caixa = 1/2 . 1/5 = 1/10
Primeira bola vermelha e da segunda caixa = 1/2 . 4/5 = 4/10

Quanto à extração da segunda bola, devemos considerar:

Probabilidade da segunda bola ser branca e da 1a. caixa (primeira bola branca):

$$\frac{1}{2} \cdot \frac{4}{10} \left(\frac{3}{4} + \frac{2}{10}\right) = \frac{1}{5} \left(\frac{19}{20}\right)$$

Probabilidade da segunda bola ser branca e da segunda caixa (primeira bola branca)

$$\frac{1}{2} \cdot \frac{1}{10} \left(\frac{4}{5} + \frac{1}{9}\right) = \frac{1}{20} \left(\frac{41}{45}\right)$$

Probabilidade da segunda bola ser branca e da primeira caixa (primeira bola vermelha)

$$\frac{1}{2} \cdot \frac{1}{10}(1 + \frac{2}{10}) = \frac{1}{20}(\frac{6}{5})$$

Probabilidade da segunda bola ser branca e da segunda caixa (primeira bola vermelha)

$$\frac{1}{2} \cdot \frac{4}{10}(\frac{4}{5} + \frac{2}{9}) = \frac{1}{5}(\frac{46}{45})$$

$$\text{Total: } \frac{19}{100} + \frac{41}{900} + \frac{6}{1100} + \frac{92}{450} = \frac{25}{100} + \frac{225}{900} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

c) Probabilidade da primeira bola (branca) ser da primeira caixa = 4/6 e da segunda caixa = 2/6

Antes de tirar a segunda bola, temos duas possibilidades de distribuição:

1a.) primeira caixa : 3B+1V e segunda caixa : 2B+8V
2a.) primeira caixa : 4B+1V e segunda caixa : 1B+8V

1a.) Possibilidade:

$$\frac{4}{6}(\frac{3}{4} + \frac{2}{10})\frac{1}{2} = \frac{1}{3}(\frac{19}{20})$$

2a.) Possibilidade:

$$\frac{2}{6}(\frac{4}{5} + \frac{1}{9})\frac{1}{2} = \frac{1}{6}(\frac{41}{45})$$

$$\text{Total: } \frac{1}{6}(\frac{19}{10} + \frac{41}{45}) = \frac{1}{6}(\frac{855 + 410}{450}) = \frac{1}{6} \cdot (\frac{1265}{450}) = 0{,}4685$$

I.5 — Análise Combinatória

a) Arranjos — O número total de arranjos de n objetos diferentes k a k é dado pela fórmula:

$${}_nA_k = n(n-1)(n-2)\ldots(n-1+1) = \frac{n!}{(n-k)!}$$

onde k pode ser um número inteiro qualquer tal que $0 \leq k \leq n$. O caso particular em que k=n é chamado permutação. O cálculo do fatorial de um número Z inteiro pode ser feito pela fórmula de Sterling que é:

$$Z! = \sqrt{2\pi Z}\ (\frac{Z}{e})^Z\ (1 + \frac{1}{12Z} + \frac{1}{288Z^2} - \frac{139}{51840Z^3} + \ldots)$$

Para $Z \geqslant 9$, o primeiro termo desta expressão já dá uma boa aproximação, com um erro menor que 1%.

O número das diferentes maneiras em que k objetos podem ser distribuídas em n posições diferentes é chamado amostragem sem reposição. Assim $_nA_k$ é o número de amostras diferentes de tamanho k que podem ser tiradas de uma população de n elementos. Neste caso, "amostra" é um arranjo particular dos k objetos. A probabilidade de se obter um determinado arranjo de k objetos de um total de n é $1/_nA_k$, assumindo-se aqui que cada arranjo possível tenha a mesma probabilidade.

Se ao selecionarmos um objeto de um conjunto de n objetos, o resultado for anotado e o objeto for devolvido ao conjunto antès da próxima seleção, temos um tipo de amostragem com reposição. O mesmo objeto pode ser sorteado mais que uma vez, isto é, são permitidas repetições. Neste caso, o número total de escolhas possíveis de k objetos é n^k. Não se faz nenhuma restrição ao valor de k, pois pode ser maior ou menor que n.

As jogadas de uma moeda são exemplos de amostragem com reposição. Numa moeda, a população n é 2 (cara e coroa) e o tamanho da amostra k é o número de jogadas.

Quando se faz uma amostragem aleatória com reposição, assume-se, a priori, que a probabilidade de qualquer uma das amostras possíveis é a mesma. Assim para uma amostra de tamanho k, incluindo as repetições, a probabilidade é $1/(n^k)$. Por exemplo, no jogo de uma moeda, se especificarmos a ocorrência de 6 caras seguidas, a probabilidade será $(1/2)^6$, porque são 6 eventos independentes. Alguns problemas interessantes envolvem, de um certo modo, os dois tipos de amostragem.

Por exemplo, qual a probabilidade de se ter um número de 5 algarismos, todos eles diferentes?

Neste problema, apesar de não haver reposição, assume-se que há um número infinito de cada tipo de algarismos e o problema é realmente de amostragem com reposição. Como n=10 e k=5, há 10^5 possíveis arranjos, incluindo as repetições assumindo-se todos eles igualmente prováveis. Há $_{10}A_5$ arranjos sem repetição. Assim a resposta à questão é:

$$p = \frac{_{10}A_5}{10^5} \cong 0{,}3024$$

b) Combinações

Consideremos agora k elementos de um total de n elementos agrupados desordenadamente. Este tipo de agrupamento é chamado combinação. Por exemplo, tomando as primeiras 4 letras do alfabeto, agrupadas duas a duas, o grupo ab é o mesmo do grupo ba, cd e o mesmo que dc, etc. O número total de combinações que se podem fazer com n elementos tomados em grupos de k elementos se representa por n^Ck ou por $\binom{n}{k}$. Neste, caso $0 \leqslant k \leqslant n$.

O número total de permutações numa combinação de k elementos é k! e o número total de arranjos em $\binom{n}{k}$ combinações é $\binom{n}{k}k!$. Assim

$$_nA_k = \binom{n}{k} k!$$

e então:

$$\binom{n}{k} = \frac{_nA_k}{k!} = \frac{n!}{k!(n-k)!}$$

Por esta expressão, vemos que k! e (n-k)! podem ser permutados, isto é:

$$\binom{n}{k} = \binom{n}{n-k}$$

O símbolo $\binom{n}{k}$ é chamado coeficiente binomial porque aparece na expansão do binômio de Newton

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \ldots + \binom{n}{k} a^k b^{n-k} + \ldots + \binom{n}{n} b^n$$

EXERCÍCIO 4

Qual a probabilidade de que as 10 primeiras cartas tiradas de um baralho sejam figuras? Temos: 12 figuras em 52 cartas, portanto:

$$p = \frac{12}{51} \cdot \frac{11}{51} \cdot \frac{10}{50} \cdots \frac{3}{43} = \frac{\frac{12!}{2!}}{\frac{52!}{42!}} = \frac{\frac{12!}{(12-10)!}}{\frac{52!}{(52-10)!}} = \frac{\frac{12!}{10!(12-10)!}}{\frac{52!}{10!(52-10)!}} = \frac{\binom{12}{10}}{\binom{52}{10}} = 41 \times 10^{-10}$$

EXERCÍCIO 5

Uma urna contém $\underline{a}$ bolas e extrai-se um grupo delas. Qual a probabilidade de extrair um número par de bolas?

$$p_{(par)} = \frac{\binom{a}{2} + \binom{a}{4} + \binom{a}{6} + \ldots}{\binom{a}{1} + \binom{a}{2} + \binom{a}{3} + \ldots}$$

1º Caso: $\underline{a}$ é par

Sabemos que $(1+1)^a = \binom{a}{0} + \binom{a}{1} + \binom{a}{2} + \ldots + \binom{a}{a}$

$$e\ (1-1)^a = \binom{a}{0} - \binom{a}{1} + \binom{a}{2} - \ldots + \binom{a}{a}$$

$$\text{Somando temos: } 2^a = 2\left[\binom{a}{0} + \binom{a}{2} + \ldots + \binom{a}{a}\right]$$

Temos também que:

$$\binom{a}{1} + \binom{a}{2} + \ldots + \binom{a}{a} = 2^a - \binom{a}{0} = 2^a - 1$$

Como, no caso, a é par, o numerador fica

$$\binom{a}{2} + \binom{a}{4} + \ldots + \binom{a}{a} = \frac{2^a}{2} - \binom{a}{0} = 2^{a-1} - 1$$

$$\text{Conclusão: } P_{par} = \frac{2^{a-1} - 1}{2^a - 1}$$

2º Caso: a é impar

Neste caso, quando somarmos $(1+1)^a$ com $(1-1)^a$, o termo $\binom{a}{a}$ será cancelado e a expressão não se alterará.

O processo de toda a ciência experimental baseia-se na repetição constante de três etapas:

1a) criação de um modelo a partir de uma concepção do comportamento da natureza do melhor modo que a possamos compreender;

2a) cálculo das probabilidades a priori, baseado na concepção feita;

3a) comparação das probabilidades a priori com medidas reais, isto é, com as probabilidades a posteriori ou experimentais. A segurança de que a concepção feita a priori está certa cresce, quando há uma boa aproximação entre a teoria e a prática; a concepção feita porém deve ser modificada, enquanto houver discordância com os resultados experimentais.

CAPÍTULO II

INTRODUÇÃO À ESTATÍSTICA

A aplicação de testes estatísticos na interpretação de resultados foi provocada principalmente pelo grande número de novos métodos experimentais desenvolvidos nas últimas décadas. Como a introdução desses métodos foi muito mais rápida que sua padronização, surgiram problemas cuja solução seria difícil sem a aplicação de testes estatísticos.

Os resultados de qualquer conjunto de medidas estão sujeitos à interação de um grande número de fatores que escapam ao controle do pesquisador. Os testes estatísticos, aplicados aos resultados fornecidos por um método, permitem condensar esses resultados e fornecer uma informação praticamente completa sobre a exatidão e precisão das medidas que o método pode oferecer.

De um modo geral, o número de resultados é pequeno e esses resultados devem ser interpretados como sendo uma amostra aleatória de um conjunto infinito (população) que é o modelo da distribuição matemática de onde foram tirados alguns valores.

A informação que se deseja a partir dos resultados obtidos baseia-se, de um ponto de vista matemático, no fato de que os parâmetros (média e variância, por exemplo) calculados a partir da amostra são estimativas dos parâmetros *reais*, mas desconhecidos, da função de distribuição do conjunto todo.

Os parâmetros do conjunto, calculados a partir da amostra, variam dentro de um certo intervalo, cujos limites dependem do número de elementos de que é constituída a amostra e do grau de confiança com que esses parâmetros podem ser aceitos. Os limites dos intervalos de variação para os parâmetros são obtidos por intermédio de valores calculados a partir da teoria das probabilidades associada à função de distribuição da população. Há algum tempo, alguns pesquisadores não aceitavam a matemática estatística aplicada a amostras com poucos elementos. Entretanto, esse ponto de vista está hoje ultrapassado e a matemática estatística fornece elementos para determinação e cálculo dos parâmetros relativos a uma medida com grau de segurança desejado, conforme será visto posteriormente, mesmo se as amostras são muito pequenas. É claro, porém, que, quanto menor a amostra, maior será o intervalo de variação dos parâmetros, isto é, menos precisa será a sua estimativa.

II.1 – Classificação dos Erros

Para saber como a teoria da probabilidade pode ser aplicada na interpretação de resultados experimentais, é preciso conhecer os tipos de erros que podem aparecer nesses resultados. Existem três tipos de erros: grosseiros ou acidentais, sistemáticos e aleatórios.

Os erros grosseiros são conseqüências de enganos ou de acidentes que podem ocorrer sem o conhecimento do pesquisador. Num conjunto de resultados, se um ou mais são afetados por erros grosseiros, existe a possibilidade de identificá-los seja por meio de testes específicos, ou seja pelo sexto sentido ou intuição do pesquisador, para serem eliminados do conjunto.

Na prática, então, existem os erros sistemáticos ou de método que prejudicam a exatidão dos resultados e os erros aleatórios ou de reprodutibilidade dos quais depende a precisão dos resultados.

Os erros sistemáticos são causados por um ou mais fatores decorrentes, por exemplo, de um defeito constante de um aparelho de medida, ou do uso de uma droga química não suficientemente pura que contém o próprio elemento a ser analisado. Podem resultar também de uma leitura sempre errada, decorrente de um vício de leitura causada por um par de óculos defeituoso do operador, etc. Para determinar a presença e o valor desses erros seria preciso conhecer aquelas causas. Nessas condições, um erro sistemático, uma vez que se conheça a causa, pode ser considerado como uma correção da medida.

Se os resultados decorrentes da repetição de uma mesma medida assumem valores que dependem unicamente do acaso e, se a esse conjunto de valores se pode associar uma função de distribuição, então esses resultados podem ser considerados variáveis aleatórias, isto é, são afetados de erros aleatórios. Uma condição para que sejam obtidos resultados aleatórios é o desconhecimento das leis que causam as variações responsáveis pela falta de reprodutibilidade de um número quando se faz a mesma medida várias vezes. Outra condição é que, se essas leis forem conhecidas, elas devem depender de muitos fatores o que torna difícil, ou mesmo impossível, na prática, um cálculo rigoroso e exato.

É importante observar que a classificação dos erros como sistemáticos ou aleatórios é muito relativa; depende do conjunto de resultados a ser considerado e do problema apresentado ao experimentador.

Por exemplo, suponhamos uma série de resultados obtidos sem levar em conta as possíveis variações de temperatura ocorridas durante o experimento. Neste caso, o erro introduzido pelas variações de temperatura é uma variável aleatória. Se, porém, os resultados forem obtidos em temperaturas bem definidas e houver uma lei de variação dos resultados com a temperatura, o erro introduzido no conjunto deve ser considerado como sistemático, isto é, como uma correção que deve ser feita nas medidas.

II.2 – Variáveis Aleatórias

Podemos dizer que conhecemos uma variável aleatória x se soubermos qual é a sua **função de distribuição**. Por exemplo,

$$F(x_a) = P(-\infty < x < x_a) \qquad (II.1)$$

indica a probabilidade da variável aleatória x assumir valores entre $-\infty$ e x_a. Se x assumir somente valores discretos $x_1, x_2 \ldots x_n$, a probabilidade da ocorrência de cada valor, isto é, $P(x_1), P(x_2), \ldots P(x_n)$ é definida de maneira que

$$\sum_i P(x_i) = 1 \qquad (II.2)$$

No caso de uma variável aleatória discreta, a função de distribuição é definida por:

$$F(x_a) = \sum_{i=-\infty}^{a} P(x_i) \qquad (II.3)$$

e o gráfico que representa esta função é do tipo apresentado na Figura 2.1. Se a variável aleatória for contínua, sua função de distribuição é

$$F(x_a) = \int_{-\infty}^{a} \varphi(x)\, dx \qquad (II.4)$$

Pela definição de probabilidade temos que

$$\int_{-\infty}^{+\infty} \varphi(x)\, dx = 1 \qquad (II.5)$$

A função $\varphi(x)$ é chamada **densidade de probabilidade** ou **diferencial da lei de distribuição**. Na Figura 2.2 a ordenada no ponto x_a dá a probabilidade dos valores experimentais serem menores que x_a. A diferença das ordenadas que correspondem aos pontos x_b e x_a dá a probabilidade dos valores estarem no intervalo $x_a - x_b$

$$P(x_a < x < x_b) = P(x_b) - P(x_a) \tag{II.6}$$

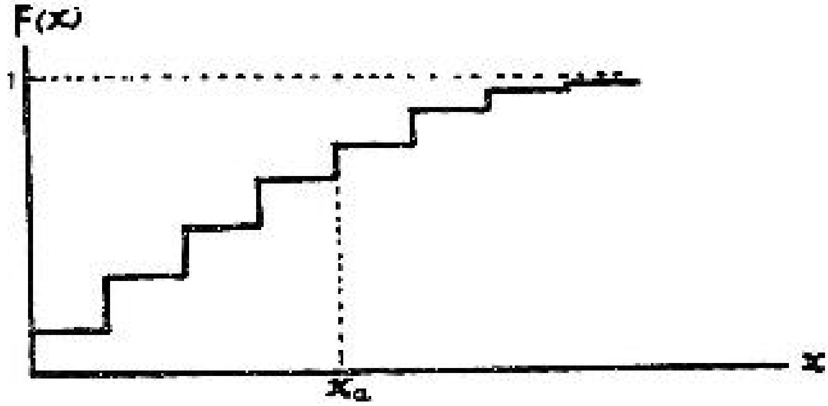

Figura 2.1 — Integral da função de distribuição de uma variável aleatória discreta.

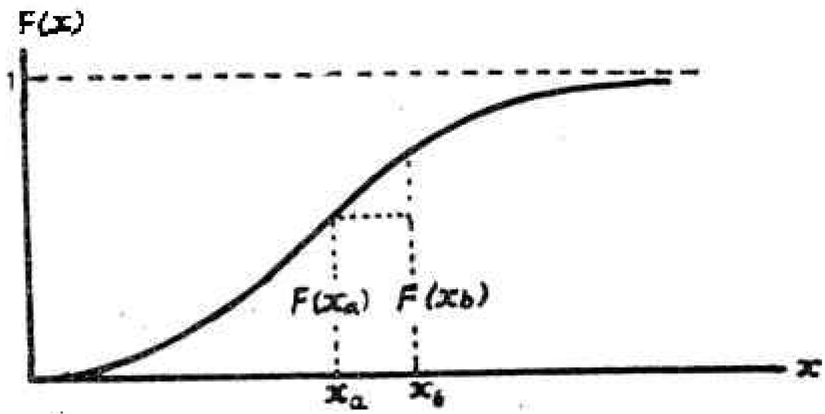

Figura 2.2 — Integral da função de distribuição de uma variável aleatória contínua.

A curva diferencial $y = \varphi(x)$ e a abcissa limitam uma área igual à unidade (Figura 2.3). A área sob a curva, à esquerda da ordenada $y_a = \varphi(x_a)$ determina a possibilidade de x ser menor que x_a. As ordenadas que passam por x_a e x_b limitam uma área que corresponde à probabilidade da ocorrência de x entre x_a e x_b,

$$P(x_a < x < x_b) = \int_{x_a}^{x_b} \varphi(x)\, dx \tag{II.7}$$

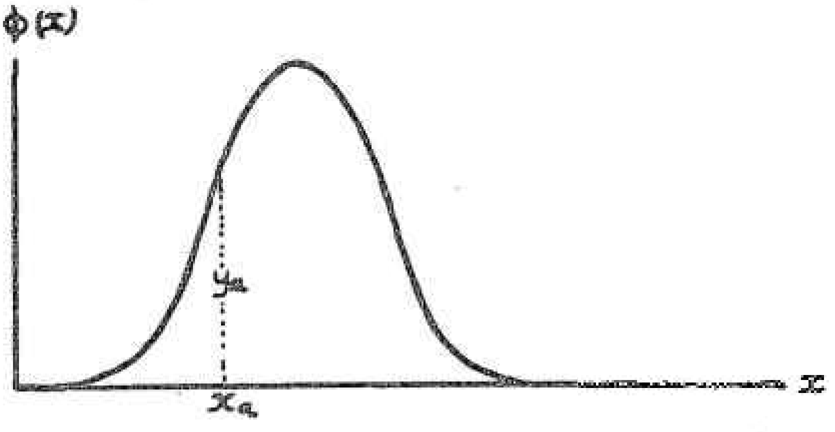

Figura 2.3 – Diferencial da função de distribuição de uma variável aleatória contínua.

A probabilidade de ocorrência de uma variável aleatória contínua x só pode ser definida num intervalo Δx, pelo produto $\varphi(x)\Delta x$. Isto não quer dizer que não seja possível obter um valor $x = x_a$, mas sim que x_a está compreendido num intervalo Δx, e para $\Delta x \rightarrow 0$, $P(x_a) = 0$.

A função de distribuição é um modelo matemático abstrato por meio do qual podem ser descritos os valores experimentais observados.

Uma das finalidades do procedimento estatístico consiste em achar uma função de distribuição que, por um lado, descreva suficientemente bem a distribuição dos valores experimentais observados e, por outro lado, seja de forma tal a permitir análises estatísticas posteriores. As expressões analíticas de uma função de distribuição contém uma ou mais constantes chamadas **parâmetros da distribuição**. Assim, por exemplo, a distribuição normal ou de Gauss tem dois parâmetros: a média ou expectância matemática e a variância. Se conhecermos a lei de distribuição, a variável aleatória pode ser completamente caracterizada pelos valores numéricos dos parâmetros.

II.3 – Média ou Expectância Matemática

O valor da média de uma variável aleatória contínua é definido por:

$$\mu = \frac{\int_{-\infty}^{+\infty} x\,\varphi(x)\,dx}{\int_{-\infty}^{+\infty} \varphi(x)\,dx} = \int_{-\infty}^{+\infty} x\,\varphi(x)\,dx \qquad (II.8)$$

e para uma distribuição discreta:

$$\mu = \frac{\sum_{-\infty}^{+\infty} x_i\,P(x_i)}{\sum_{-\infty}^{+\infty} P(x_i)} = \sum_{-\infty}^{+\infty} x_i\,P(x_i) \qquad (II.9)$$

onde $P(x_i)$ é a probabilidade de ocorrência do valor x_i.

A média de n valores experimentais x_1 $x_2, \dots x_n$ é definida por:

$$\overline{x} = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{II.10}$$

$\overline{x}$ é a média da amostra, que não é a média μ da população (conjunto de n valores quando $n \to \infty$). A média x da amostra é uma estimativa da média μ da população. Também se exprime a média por M $\{x\}$ ou E $\{x\}$

II.3.1 – Propriedades da Média

Vamos supor $\underline{x}$ uma variável aleatória discreta. Para estender o estudo a uma distribuição contínua de $\underline{x}$ basta substituir o sinal de somatório pelo sinal de integral.

a) A média de uma constante é a própria constante

$$E\{K\} = \sum_i K\,P(x_i) = K \sum_i P(x_i) = K \tag{II.11}$$

b) A média de uma variável aleatória multiplicada por uma constante é a média multiplicada por essa mesma constante.

$$E\{Kx\} = \sum_i Kx_i\,P(x_i) = K \sum_i x_i\,P(x_i) = K\,E\{x\} = K\mu \tag{II.12}$$

c) A média da soma de duas variáveis aleatórias x e y é a soma das médias

$$E\{x+y\} = \sum_i \sum_j (x_i + y_j)\,P(x_i,y_j)$$

$$= \sum_i \sum_j x_i P(x_i,y_j) + \sum_i \sum_j y_j P(x_i,y_j)$$

$$= \sum_i x_i \sum_j P(x_i,y_j) + \sum_j y_j \sum_i P(x_i,y_j)$$

Como

$$\sum_j P(x_i,y_j) = P(x_i)$$

e

$$\sum_i P(x_i,y_j) = P(y_j)$$ temos:

$$E\{x+y\} = \sum_i x_i P(x_i) + \sum_j y_j P(y_j) = E\{x\} + E\{y\} = \mu_x + \mu_y \quad (II.13)$$

d) A média da diferença de duas variáveis aleatórias é a diferença das médias. Demonstra-se como no caso da soma.

e) Somando ou subtraindo uma constante de uma variável aleatória, a sua média fica somada ou subtraída da mesma constante,

$$E\{x \pm K\} = E\{x\} \pm K \quad (II.14)$$

Resulta das propriedades c e d fazendo $y = k$

f) A média da distribuição $x - \mu(x)$, cuja variável é chamada "variável centrada", é igual a zero. Basta aplicar a propriedade d.

$$E\{x - \mu(x)\} = E\{x\} - \mu(x) = 0 \quad (II.15)$$

g) A média do produto de duas variáveis aleatórias independentes é o produto das médias

$$E\{xy\} = \sum_i \sum_j x_i y_j P(x_i, x_j) \quad (II.16)$$

Como, por hipótese, x e y são independentes pode-se escrever:

$$P(x_i, y_j) = P(x_i) \cdot P(y_j)$$

Substituindo

$$E\{xy\} = \sum_i \sum_j x_i y_j P(x_i) \cdot P(y_j)$$

$$= \sum_i x_i P(x_i) \cdot \sum_j y_j P(y_j) = E\{x\} \cdot E\{y\} \quad (II.17)$$

EXERCÍCIO 6

Cálculo de u'a Média

Suponhamos que tenham sido feitas seis determinações de urânio num minério padrão que contém 0,318% de U e que os resultados, em porcentagem, são: 0,318; 0,314; 0,323; 0,300; 0,314 e 0,327. Calcular a estimativa x da média verdadeira μ. O cálculo da média é feito pela fórmula (II.10):

$$\bar{x} = \frac{0{,}318 + 0{,}314 + 0{,}323 + 0{,}300 + 0{,}314 + 0{,}327}{6} = 0{,}316$$

$\overline{x} = 0{,}316$ é uma estimativa de $\mu = 0{,}318$

O cálculo da média pode ser simplificado fazendo uma transformação linear de variáveis do tipo:

$$z_i = ax_i + b$$

Para se obterem números inteiros e pequenos, fazemos:

$$\overline{z} = \sum_i \frac{z_i}{n} = \sum_i^n \frac{ax_i + b}{n} = a\overline{x} + b$$

Assim

$$\overline{x} = \frac{\overline{z} - b}{a}$$

Fazendo no exemplo anterior $z_i = (x_i - 0{,}300)\ 1000 = 1000\ x_i - 300$ vamos ter:

$$\overline{z} = \frac{18 + 14 + 23 + 0 + 14 + 27}{6} = 16$$

Entaõ

$$\overline{x} = \frac{16 + 300}{1000} = 0{,}316$$

II.4 — Variância e Desvio Padrão

Chamamos de **momento centrado de ordem k** ao momento em relação à média

$$\mu_k(x) = E\{(x - \mu(x))^k\} \tag{II.18}$$

A **variância** é o momento centrado de segunda ordem e designa-se por $\sigma^2(x)$, ou σ^2_x ou simplesmente σ^2. A variância de uma variável aleatória x com distribuição contínua é

$$\sigma^2(x) = \int_{-\infty}^{+\infty} \{x - \mu(x)\}^2\ \varphi(x)\, dx \tag{II.19}$$

e para distribuição discreta

$$\sigma^2(x) = \sum_i \{x_i - \mu(x)\}^2 \ . \ P(x_i) \qquad \text{(II.20)}$$

Para um número limitado de valores $x_1, x_2 \ldots x_n$, a variância é definida por $s^2(x)$, s_x^2 ou simplesmente s^2.

$$s_x^2 = \frac{\sum_i (x_i - \overline{x})^2}{n-1} \qquad \text{(II.21)}$$

onde s_x^2 é uma estimativa de σ^2, variância da população. No denominador aparece o valor n — 1 porque, depois de calculada a média, temos n — 1 valores independentes, isto é, perdemos um grau de liberdade. Se a média verdadeira μ for conhecida, (caso de uma amostra padrão, cujo valor da grandeza em estudo é, supostamente, o valor real), a estimativa da variância é dada por:

$$s_x^2 = \frac{\sum_i (x_i - \mu)^2}{n} \qquad \text{(II.22)}$$

porque, neste caso, não foi calculada a média $\overline{x}$ e portanto não se perdeu um grau de liberdade.

O valor positivo da raiz quadrada da variância é o desvio padrão da população ou s_x da amostra (número finito de elementos) que é uma estimativa de σ.

O desvio padrão expresso como porcentagem da média é o **coeficiente de variação** V_x, V(x) ou simplesmente V.

$$V(x) = \frac{s_x}{\overline{x}} 100 \qquad \text{(II.23)}$$

II.4.1 — Propriedades da Variância

a) Muitas vezes no cálculo da variância é mais conveniente usar a fórmula que se obtém expandindo o quadrado e aplicando propriedades da soma e diferença da média.

$$\sigma^2 = E\{(x-\mu)^2\} = E\{x^2 - 2\mu x + \mu^2\} = E\{x^2\} - 2\mu E\{x\} + \mu^2$$

Como $E\{x\} = \mu$,

$$\sigma^2 = E\{x^2\} - \mu^2 \qquad \text{(II.24)}$$

b) A variância de uma constante é zero

$$\sigma^2(k) = E\{k-\mu(k)\}^2 = E\{k-k\}^2 = 0 \qquad (II.25)$$

c) Multiplicando-se uma variável aleatória por uma constante, a sua variância fica multiplicada pelo quadrado da constante

$$\sigma^2(kx) = E\{(kx - E(kx))^2\} = E\{[k(x-\mu(x)]^2\}$$

$$= E\{k^2(x-\mu(x))^2\} = k^2\, E\{(x-\mu(x))^2\} = k^2\, \sigma^2(x)$$

d) A variância da soma ou diferença de duas variáveis aleatórias independentes é igual à soma das respectivas variâncias.

$$\sigma^2(x \pm y) = E\{[(x \pm y) - (\mu_x \pm \mu_y)]^2\}$$

$$= E\{[(x-\mu_x) \pm (y-\mu_y)]^2\}$$

$$= E\{(x-\mu_x)^2\} \pm 2E\{(x-\mu_x).(y-\mu_y)\} + E\{(y-\mu_y)^2\}$$

A expressão $E\{(x-\mu_x).(y-\mu_y)\}$ é a co-variância de x e y. Se x e y são independentes,

$$E\{(x-\mu_x).(y-\mu_y)\} = E\{x-\mu_x\} . E\{y-\mu_y\} = 0$$

Então

$$\sigma^2(x \pm y) = \sigma^2(x) + \sigma^2(y) \qquad (II.27)$$

e) Se adicionarmos ou subtrairmos uma constante a uma variável aleatória, a sua variância não se altera. Substituindo y na propriedade anterior por k temos

$$\sigma^2(x+k) = \sigma_2^2(x) + \sigma^2(k) = \sigma^2(x)$$

porque $\sigma^2(k) = 0$

Na prática, o uso da propriedade a) é mais vantajosa, porque evita problemas de aproximação numérica. Considerando a soma dos quadrados

$$SQ_x = \sum_i (x_i - \bar{x})^2$$

temos:

$$\sum (x_i - \bar{x})^2 = \sum_i (x_i^2 - 2\bar{x}\, x_i + \bar{x}^2)$$

$$= \sum_i x_i^2 - 2\,\bar{x} \sum_i x_i + n\,\bar{x}^2$$

$$= \sum_i x_i^2 - 2\left(\frac{\sum_i x_i}{n}\right) \sum_i x_i + n \left(\frac{\sum_i x_i}{n}\right)^2$$

$$= \sum x_i^2 - \frac{(\sum_i x_i)^2}{n}$$

Então

$$s_x^2 = \frac{\sum_i x_i^2 - (\sum x_i)^2/n}{n-1} \qquad \text{(II.28)}$$

EXERCÍCIO 7

Usando os mesmos valores do exercício 1 com a transformação linear de variáveis

$$z_i = ax_i + b$$

temos:

$$s_z^2 = \sum_i \frac{(z_i - \bar{z})^2}{n-1} = \sum_i \left[(ax_i - b) - (a\bar{x}_i - b)\right]^2 \frac{1}{n-1}$$

$$= \sum_i \frac{a^2 (x_i - \bar{x})^2}{n-1} = a^2 s_x^2$$

$$s_x^2 = \frac{1}{a^2} s_z^2$$

O valor de b não aparece nos resultados, em virtude da propriedade e). No nosso exemplo temos então:

$$s_x^2 = \frac{2^2 + 2^2 + 7^2 + 16^2 + 2^2 + 11^2}{1000 \times 5} = 0{,}0000876$$

$$s_x = 0{,}0099$$

Como a média verdadeira é conhecida ($\mu = 0{,}318$) é mais correto calcular a variância e desvio padrão em relação a μ. Assim:

$$s_x^2 = \frac{0^2 + 4^2 + 5^2 + 18^2 + 4^2 + 9^2}{1000^2 \times 6} = 0{,}000077$$

$$s_x = 0{,}0088$$

EXERCÍCIO 8

Oito resultados de uma medida foram os seguintes:

4,23	4,15
4,12	4,21
4,18	4,08
4,09	4,16

Cálculo da média:

Para simplificar, fazemos:

$$z_i = (x_i - 4{,}00)\ 100.$$

Então:

$$\bar{z} = \frac{\sum_1^8 z_i}{8} = \frac{(\sum_1^8 x_i - 4{,}00)\ 100}{8}$$

de onde:

$$\bar{x} = 4{,}00 + \frac{\sum_1^8 z_i}{8 \times 100}$$

$$\bar{x} = 4{,}00 + \frac{23 + 12 + 18 + 9 + 15 + 21 + 8 + 16}{100 \times 8} = 4{,}00 + 0{,}1525$$

$$\bar{x} \cong 4{,}15$$

Fazendo agora a transformação $z_i = (x_i - 4{,}15)\ 100$, vamos obter pela fórmula (II.21):

$$s_z^2 = \frac{8^2 + 3^2 + 3^2 + 6^2 + 0 + 6^2 + 7^2 + 1^2}{7} = \frac{204}{7} = 29{,}14$$

$$s_x^2 = \frac{29{,}14}{100^2} = 0{,}002914$$

e pela fórmula (II.28) se obtém:

$$s_z^2 = \frac{(64 + 9 + 3 + 36 + 0 + 36 + 49 + 1) - 2^2/8}{7} = \frac{204 - 4/8}{7} = \frac{203{,}5}{7} = 29{,}07$$

$$s_x^2 = \frac{s_z^2}{100^2} = 0{,}002907$$

O segundo caso é mais correto porque não foi usada a aproximação da média.

II.4.2 – Variância de Resultados de Rotina

Tem muita aplicação na prospecção de minérios, em química analítica, no controle de produção, etc., onde é necessário fazer numerosas análises de amostras de composição semelhante, durante um intervalo de tempo mais ou menos prolongado.

Um modo de se obter o erro de reprodutibilidade do método adotado, no laboratório onde está sendo aplicado, é fazer a análise estatística de todos os resultados.

No decorrer do trabalho, como o tempo exigido é longo, a distribuição das médias pode variar de maneira incontrolável, mesmo que a amostra analisada seja sempre a mesma. Muitos fatores podem concorrer para essas flutuações, tais como: impurezas dos reagentes, variações de temperatura, pressão, umidade, iluminação, etc. Por esse motivo, é mais conveniente calcular o erro de reprodutibilidade (variância) a partir de todos os resultados.

Suponhamos que temos m amostras para análises e que tenham sido feitas n determinações paralelas de cada uma. Os resultados devem ser apresentados segundo a Tabela II.1.

Tabela II.1

Apresentação de Resultados de Análises de Amostras Semelhantes

	Amostra 1	Amostra 2	–––	Amostra i	–––	Amostra m
1ª Determ.	x_{11}	x_{21}	–––	x_{i1}	–––	x_{m1}
2ª Determ.	x_{12}	x_{22}	–––	x_{i2}	–––	x_{m2}
.	.	.	–––	–––	–––	–––
.	.	.	–––	–––	–––	–––
.	.	.	–––	–––	–––	–––
jª Determ.	x_{1j}	x_{2j}	–––	x_{ij}	–––	x_{mj}
.	.	.	–––	–––	–––	–––
.	.	.	–––	–––	–––	–––
.	.	.	–––	–––	–––	–––
nª Determ.	x_{1n}	x_{2n}	–––	x_{in}	–––	x_{mn}
Totais	X_1	X_2		X_i		X_m

Para achar a variância que caracteriza a reprodutibilidade do método nesse laboratório, acha-se a variância para cada amostra separadamente e depois determina-se a média das variâncias, de acordo com a expressão:

$$s_x^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \overline{x}_i)^2}{m(n-1)} \qquad (II.29)$$

onde $\overline{x}_i = X_i/n$ é a média de cada coluna. É mais conveniente calcular pela fórmula equivalente em que X_i é a soma dos elementos da coluna i

$$s_x^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{m} X_i^2/n}{m(n-1)} \qquad (II.30)$$

Neste caso, o número de graus de liberdade é m(n — 1), porque foram impostas m relações no conjunto de medidas ao calcular m médias.

No caso em que o número n de determinações em cada coluna não é o mesmo, a variância total deve ser determinada por uma fórmula que se obtém tomando a soma das variâncias ponderadas de acordo com seus graus de liberdade e dividindo a soma pelo número total de graus de liberdade.

O número de graus de liberdade é igual ao número total de resultados $\sum_{i=1}^{m} n_i$ menos o número de relações usadas para a determinação das médias

$$s_x^2 = \frac{\sum_{i=1}^{m} (n_i - 1)\, s_i^2}{\sum_{i=1}^{m} n_i - m} \qquad (II.31)$$

Por uma transformação semelhante à feita anteriormente obtém-se a seguinte expressão conveniente para os cálculos:

$$s_x^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{m} X_i^2/n_i}{\sum_{i=1}^{m} n_i - m} \qquad (II.32)$$

EXERCÍCIO 9

Resultados de determinações de tório em cinco amostras de um mesmo minério, (em porcentagem):

(1)	(2)	(3)	(4)	(5)
2,51	2,93	2,31	2,10	2,84
2,47	2,80	2,40	2,28	2,63
2,38	2,85	2,51	2,15	2,71
2,41	2,70		2,21	2,68
2,49				2,78

Para simplificar os cálculos usamos a transformação

$$z_{ij} = \frac{x_{ij} - b_i}{a}$$

Como a variância independe dos valores b_i, estes podem ser diferentes em cada coluna:

(1) $z_{1j} = (x_{1j} - 2{,}45)100$; valores de z_1 : 6, 2, -7, -4, 4
(2) $z_{2j} = (x_{2j} - 2{,}80)100$; valores de z_2 : 13, 0, 5, -10
(3) $z_{3j} = (x_{3j} - 2{,}40)100$; valores de z_3 : -9, 0, 11
(4) $z_{4j} = (x_{4j} - 2{,}15)100$; valores de z_4 : -5, 13, 0, 6
(5) $z_{5j} = (x_{5j} - 2{,}70)100$; valores de z_5 : 14, -7, 1, -2, 8

Vamos ter:

$$s_z^2 = \frac{36 + 4 + 49 + \ldots + 4 + 64 - \left(\frac{(6+2-7-4+4)^2}{5} + \ldots + \frac{(14-7+1-2+8)^2}{5} \right)}{21 - 5}$$

$$s_z^2 = \frac{1161 - \left(\frac{1^2}{5} + \frac{8^2}{4} + \frac{2^2}{3} + \frac{14^2}{4} + \frac{11^2}{5} \right)}{16} = \frac{1161 - 90{,}7333}{16} = 66{,}8917$$

$$s_x^2 = \frac{66{,}8917}{(100)^2} = 0{,}00668917$$

$$s_x^2 = 0{,}082$$

Para o cálculo da variância de resultados desse tipo, partimos da hipótese que os resultados das m amostras podem ser considerados variáveis aleatórias de m populações. Pela própria natureza do problema, essas populações podem ter médias diferentes, mas a mesma variância. Assim, cada um dos valores $s_1^2, s_2^2, \ldots s_m^2$ podem ser considerados estimativas do mesmo espalhamento total.

Uma combinação de amostras que diferem na sua constituição só é possível dentro de certos limites, isto é, desde que o erro de reprodutibilidade independa da média. Admite-se, na prática, que esta condição é satisfeita quando os valores extremos do componente a ser analisado mantém entre si uma relação até 1:3.

O cálculo do erro de reprodutibilidade de análises em série torna-se mais simples se, em cada amostra, forem feitas duas determinações paralelas conforme será visto mais adiante.

II.4.3 – Lei da Propagação de Erros

Se um resultado y depende de uma medida x, temos:

$$y = f(x)$$

Suponhamos que a medida x seja afetada de um erro Δx. Como y é função de x, o resultado y será afetado de um erro Δy. Na prática, os erros são geralmente pequenos, quando comparados com x e y. Com boa aproximação, podemos admitir que a relação $\Delta y/\Delta x$ se aproxima da derivada num ponto. (x,y) genérico. Temos:

$$\frac{\Delta y}{\Delta x} \cong \frac{dy}{dx}$$

Para um ponto (x_i, y_i) pode-se admitir que:

$$dy_i = \left(\frac{dy}{dx}\right) dx_i$$

onde dy_i e dx_i são os erros de y e x, respectivamente, no ponto considerado. Generalizando, vamos considerar y como função de variáveis independentes:

$$y = f(x_1, x_2, \ldots x_n)$$

$$dy = \left(\frac{\partial f}{\partial x_1}\right) dx_1 + \left(\frac{\partial f}{\partial x_2}\right) dx_2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right) dx_n$$

Elevando ao quadrado, vamos ter:

$$dy^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 dx_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 dx_2^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right)^2 dx_n^2$$

Os termos do tipo

$$\left(\frac{\partial f}{\partial x_i}\right) \left(\frac{\partial f}{\partial x_j}\right) dx_i \; dx_j \qquad (\text{para } i \neq j)$$

se anulam, porque os erros em relação à média são afetados de sinais positivos ou negativos com igual probabilidade. Em termos de variância, vamos ter:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2 \qquad \text{(II.33)}$$

a) **Variância de uma Soma ou Diferença**

$$y = x_1 \pm x_2$$

$$\sigma_y^2 = (1)^2\,\sigma_{x_1}^2 + (\pm 1)^2\,\sigma_{x_2}^2$$

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \qquad (II.34)$$

Este resultado corresponde à propriedade d) da variância

b) **Variância de um Produto**

$$y = k\,x_1\,x_2$$

$$\sigma_y^2 = (kx_1)^2\,\sigma_{x_2}^2 + (kx_2)^2\,\sigma_{x_1}^2$$

$$\sigma_y^2 = k^2\,(x_1^2\,\sigma_{x_2}^2 + x_2^2\,\sigma_{x_1}^2) \qquad (II.35)$$

c) **Variância de um Quociente**

$$y = k\,\frac{x_1}{x_2}$$

$$\sigma_y^2 = \left(\frac{k}{x_2}\right)^2 \sigma_{x_1}^2 + \left(-k\,\frac{x_1}{x_2^2}\right)^2 \sigma_{x_2}^2$$

$$\sigma_y^2 = k^2\left(\frac{x_1}{x_2}\right)^2 \left[\frac{\sigma^2 x_1}{x_1^2} + \frac{\sigma^2 x_2}{x_2^2}\right] \qquad (II.36)$$

d) **Variância de Potências e Raizes**

$$y = b + k\,x^n$$

$$\sigma_y^2 = n^2\,k^2\,(x^{n-1})^2\,\sigma_x^2 \qquad (II.37)$$

e) **Variância de Logaritmos**

$$y = k \log x$$

$$\sigma_y^2 = \frac{k^2}{x^2}\,\sigma_x^2$$

f) **Variância de uma Média**

Sejam $x_1, x_2, \ldots x_n$ medidas de mesma grandeza feitas com a mesma precisão. Vimos que:

$$\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n}$$

$$\sigma^2_{\bar{x}} = \left(\frac{\partial \bar{x}}{\partial x_1}\right)^2 \sigma^2_{x_1} + \left(\frac{\partial \bar{x}}{\partial x_2}\right)^2 \sigma^2_{x_2} + \ldots + \left(\frac{\partial \bar{x}}{\partial x_n}\right)^2 \sigma^2_{x_n}$$

$$\sigma^2_{\bar{x}} = \left(\frac{1}{n}\right)^2 (\sigma^2_{x_1} + \sigma^2_{x_2} + \ldots + \sigma^2_{x_n})$$

Como por hipótese os $\sigma^2_{x_i}$ são todos iguais, temos:

$$\sigma^2_{\bar{x}} = \frac{1}{n^2} (n\,\sigma^2_x)$$

$$\sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n} \qquad \text{(II.39)}$$

II.4.4 – Aplicações da Lei de Propagação dos Erros

Vamos considerar os seguintes exemplos:

a) Pela equação (II.39), pode-se ver que o aumento do número de determinações paralelas para aumentar a precisão de uma média só tem sentido até um certo limite. Se temos um valor de $\sigma_{\bar{x}}$ para $n = 3$ e quisermos aumentar a precisão duas vezes, temos que fazer $3 \times 2^2 = 12$ determinações. Para ter uma precisão 10 vezes maior são necessárias $3 \times 10^2 = 300$ determinações o que, do ponto de vista prático, não tem sentido.

b) Ainda pela equação (II.39), devemos considerar o seguinte: se temos, por exemplo, dois métodos experimentais, o primeiro com coeficiente de variação de 5% e o segundo com coeficiente de variação de 10% e este permite obter quatro resultados num tempo t em que o primeiro só permite obter um, pode-se dizer que os dois métodos são igualmente precisos, se levarmos em conta o fator tempo. Em análises de rotina, em que o fator tempo é importante, somos forçados a admitir que os métodos mais rápidos, muitas vezes, são mais precisos que os métodos clássicos, geralmente lentos.

c) Outra aplicação importante da lei de propagação de erros é a obtenção de congruência nos resultados. Se um resultado depende de várias medidas e uma delas só pode ser obtida com precisão relativamente baixa, é inútil obter as demais com boa precisão, porque, no resultado, vai prevalecer o erro da primeira. Como exemplo, vamos considerar a determinação do peso molecular de uma substância pelo método de Beckman. Para soluções diluídas é válida a relação:

$$M = K \frac{w}{D_W}$$

onde:

M = peso molecular do soluto

w = massa em gramas do soluto

W = massa em gramas do solvente

D = diferença (abaixamento) do ponto de congelamento do solvente produzido por w gramas de soluto em W gramas de solvente. K é uma constante teórica e vamos considerá-la livre de erros.

Temos:

$$\sigma^2_M = (\frac{K}{DW})^2 \sigma^2_w + \frac{K^2 w^2}{D^4 W^2} \sigma^2_D + \frac{K^2 w^2}{D^2 W^4} \sigma^2_W$$

$$\sigma^2_M = \frac{K^2 w^2}{D^2 W^2} (\frac{\sigma^2_w}{w^2} + \frac{\sigma^2_D}{D^2} + \frac{\sigma^2_W}{W^2})$$

$$\frac{\sigma^2_M}{M} = \frac{\sigma^2_w}{w^2} + \frac{\sigma^2_D}{D^2} + \frac{\sigma^2_W}{W^2}$$

Multiplicando por 100^2, temos em termos de coeficiente de variação:

$$V^2_M = V^2_w + V^2_D + V^2_W$$

Os valores geralmente usados são:

w = 0,5000 g ± 0,0002 g

W = 25,0000 g ± 0,0002 g

para produzir um abaixamento de temperatura de 0,5°C. Suponhamos que queremos para o peso molecular uma precisão de ± 0,1%. Com que precisão deverá ser lida a temperatura?

$$V_D = \sqrt{V^2_M - V^2_w - V^2_W}$$

$$V_D = \sqrt{(0,1)^2 - (0,04)^2 - (0,0008)^2}$$

$$V_D \cong 0,09\%$$

Assim o erro na leitura da temperatura deveria ser no máximo de 0,0000045°C o que só poderia ser obtido com aparelhos especiais de altíssima precisão. Com aparelhos considerados de boa precisão se consegue observar uma variação de ± 0,002°C o que dá um coeficiente de variação de ± 0,4%. Admitindo agora que a leitura da temperatura tenha sido feita com um aparelho deste tipo vamos calcular com que precisão se obtém o valor do peso molecular. Temos:

$$V_M = \sqrt{(0{,}04)^2 + (0{,}4)^2 + (0{,}0008)^2}$$

$$V_M \cong 0{,}4\%$$

Isto mostra que a grande precisão na pesagem do solvente não tem sentido. Bastaria usar uma balança com precisão de ± 0,01 g, o que permitiria uma pesagem mais rápida, sem afetar a precisão do resultado final.

d) Pelo uso da fórmula (II.34), pode-se obter um meio prático para o cálculo do erro de reprodutibilidade de um método usado para análises de rotina, quando são feitas duas determinações paralelas. Sejam x e y os resultados de duas determinações paralelas:

$$s^2_{(x-y)} = s^2_x + s^2_y = d^2_i$$

onde $d_i = x - y$. Se temos n determinações paralelas:

$$s^2_{(x-y)} = \sum_{i=1}^{n} \frac{d^2_i}{n}$$

Aqui, o número de graus de liberdade é n porque os d_i são independentes (não houve cálculo da média). Como o método usado é sempre o mesmo e as condições também:

$$s_x = s_y = s$$

e temos:

$$s = \sqrt{\frac{\sum_{i=1}^{n} d^2_i}{2n}} \qquad \text{(II.40)}$$

e) Quando usamos as fórmulas da lei de propagação de erros, devemos observar um detalhe muito importante: as variáveis devem ser independentes. No caso da fórmula (II.34), por exemplo, se houver uma dependência funcional entre as medidas, a variança será:

$$s^2_{(x \pm y)} = s^2_x \pm 2\, r\, s_x\, s_y + s^2_y \qquad \text{(II.41)}$$

onde r_{xy} é o coeficiente de correlação entre as variáveis. O coeficiente de correlação é uma medida da relação linear entre as duas variáveis aleatórias e seu valor varia entre +1 e −1. Se não houver correlação entre as variáveis, $r_{xy} = 0$. O valor positivo do coeficiente de correlação indica que as duas variáveis variam no mesmo sentido, ao passo que, um valor negativo indica que se uma variável aumenta, a outra diminue.

CAPÍTULO III

DISTRIBUIÇÃO NORMAL E BINOMIAL

III.1 — Distribuição de Resultados e de Erros

Suponhamos um número muito grande de resultados de uma medida. Se os resultados obtidos obedecem a um certo modelo de distribuição, segue-se que os erros que afetam esses resultados se distribuem segundo esse mesmo modelo. Vamos considerar um exemplo teórico, (Tabela III.1), como ilustração, em que x_i são os resultados, f_i as freqüências (número de resultados com o valor x_i) e d_i os desvios, (erros), em relação à média $\overline{x}$, dos valores x_i

Tabela III.1

Exemplo Teórico de Distribuição de Resultados e Erros

x_i	f_i	$d_i \times 10$
9,7	10	-3
9,8	20	-2
9,9	40	-1
10,0	70	0
10,1	40	1
10,2	20	2
10,3	10	3

As Figuras 3.1 e 3.2 ilustram a distribuição dos resultados e dos erros.

Se existe uma equação matemática que descreve a primeira distribuição, uma simples mudança de escala fará com que a mesma equação matemática seja válida para a segunda. A equação deste tipo de distribuição foi deduzida a partir de considerações matemáticas de probabilidade.

III.2 — Distribuição Binomial

Vamos tomar a expressão $(p+q)^n$ onde p é a probabilidade da ocorrência de um fato e q a probabilidade da não ocorrência desse mesmo fato. No caso particular de uma moeda, seja p a probabilidade de dar cara e q a probabilidade de dar coroa. A soma de p+q é uma certeza e vale 1.

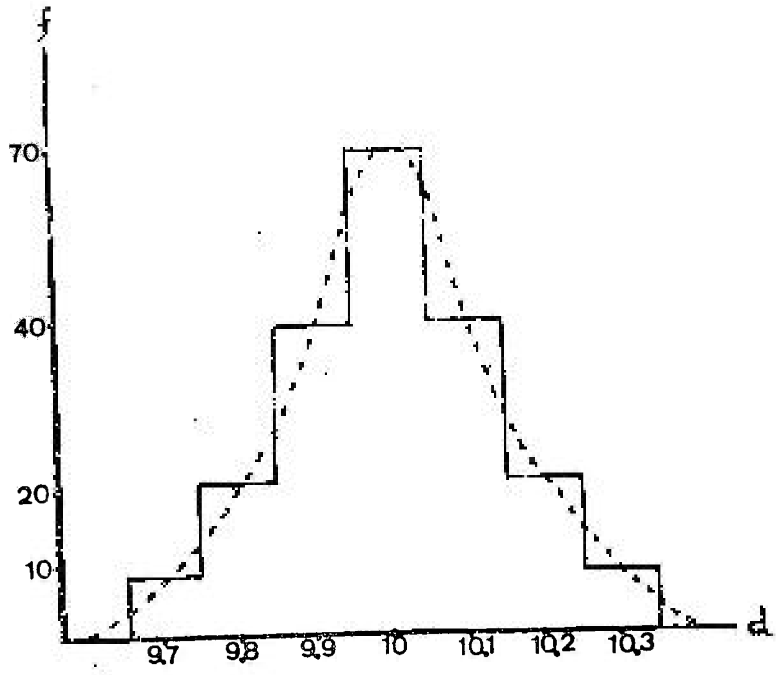

Figura 3.1 – Distribuição dos resultados.

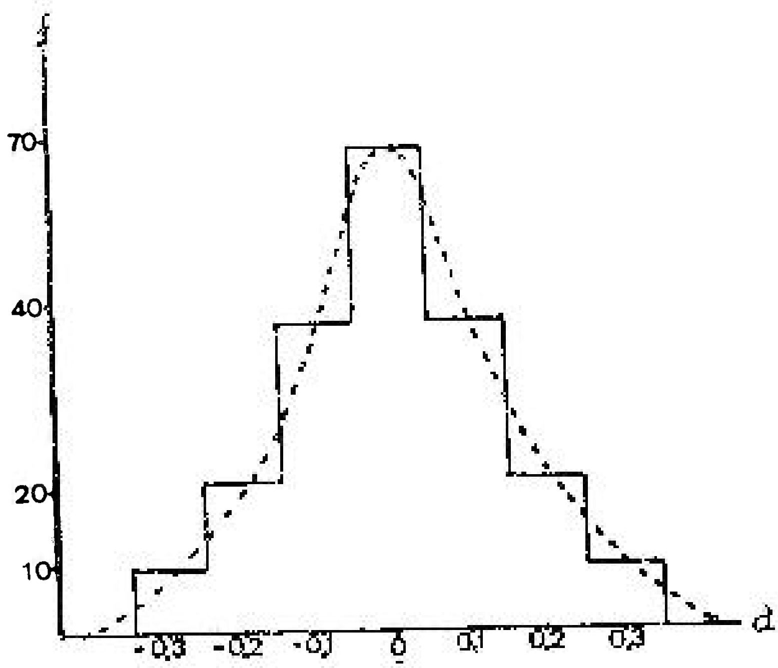

Figura 3.2 – Distribuição dos erros.

Suponhamos que se joguem ao mesmo tempo n moedas. Expandindo $(p+q)^n$, obtemos um tabela de freqüências prováveis. Vamos ter:

$$(p+q)^n = p^n + {}_nC_1\, p^{n-1} + {}_nC_2\, p^{n-2} q^2 + \ldots + {}_nC_{n-2}\, p^2 q^{n-2} + {}_nC_{n-1}\, p\, q^{n-1} + q^n \quad \text{(III.1)}$$

onde p^n representa a freqüência provável das n moedas darem cara, ${}_nC_1\, p^{n-1} q$, a freqüência provável de aparecer uma vez coroa em uma (qualquer) das n moedas, etc.

Pela fórmula da combinação de n elementos r a r sabemos que:

$${}_nC_r = \frac{n!}{(n-r)!\, r!}$$

e a série se torna:

$$(p+q)^n = p^n + np^{n-1} q + \frac{n(n-1)}{2!} p^{n-2} q^2 + \ldots + \frac{n(n-1)}{2!} p^2 q^{n-2} + npq^{n-1} + q^n \quad \text{(III.2)}$$

III.2.1 – Média da Distribuição

Sabemos que para a ocorrência de r caras em n moedas, temos a probabilidade:

$$P(r) = \frac{n!}{(n-r)!\, r!} p^r q^{n-r}$$

A média é definida como:

$$\mu_r = \sum_{r=0}^{n} r\, P(r) \quad \text{(III.3)}$$

Vamos considerar o desenvolvimento de $(px+q)^n$, onde $p+q = 1$ e x é um valor qualquer:

$$(px+q)^n = \sum_{r=0}^{n} \frac{n!}{(n-r)!\, r!} p^r x^r q^{n-r} = \sum_{r=0}^{n} x^r P(r)$$

Diferenciando em relação a x, vamos ter:

$$np(px+q)^{n-1} = \sum_{r=0}^{n} rx^{r-1} P(r) \quad \text{(III.4)}$$

Em particular, para $x = 1$, a expressão se torna

$$np = \sum_{r=0}^{n} r\,P(r)$$

O 2º membro desta equação é a média μ_r. Então:

$$\mu_r = np \qquad \text{(III.5)}$$

III.2.2 — Variância e Desvio Padrão

Se diferenciarmos mais uma vez, em relação a x, a expressão (III.4), vamos ter:

$$n(n-1)\,p^2\,(px+q)^{n-2} = \sum_{r=0}^{n} r^2\,P(r) - \sum_{r=0}^{n} r\,P(r)$$

onde o último termo é a média μ_r e o penúltimo é a média dos quadrados $\overline{r^2}$, isto é:

$$n(n-1)\,p^2 = \overline{r^2} - \mu_r \qquad \text{(III.6)}$$

Sabemos que a variância é a diferença entre a média dos quadrados e o quadrado da média, isto é:

$$\sigma_r^2 = \overline{r^2} - \mu_r^2 \qquad \text{(III.7)}$$

Combinando (III.6) e (III.7) vamos ter:

$$\sigma_r^2 = n(n-1)\,p^2 + \mu_r - \mu_r^2$$

Como achamos que $\mu_r = np$, substituindo vamos ter:

$$\sigma_r^2 = n(n-1)\,p^2 + np - n^2\,p^2$$

$$\sigma_r^2 = np - np^2 = np(1-p)$$

$$\sigma_r^2 = n\,p\,q \qquad \text{(III.8)}$$

$$\sigma_r = \sqrt{npq} \qquad \text{(III.9)}$$

III.3 — Curva de Erros — Distribuição Normal ou de Gauss

Com a obtenção do desenvolvimento da expressão $(p+q)^n$, é possível obter uma expressão geral para a distribuição normal dos coeficientes de freqüência. Pode-se definir um **erro** como a discrepância entre um valor observado e o valor verdadeiro. Dificilmente temos o valor verdadeiro, mas somente uma estimativa que seria a média de um número finito de observações. Os erros das medidas resultam de uma série de fatores não perfeitamente controlados e que atuam sobre os resultados de modo praticamente independente uns dos outros.

Colocando num gráfico (Figura 3.3) os termos da expressão $(p+q)^{2n}$, se n é um número pequeno, vamos ter uma linha poligonal e não uma curva contínua.

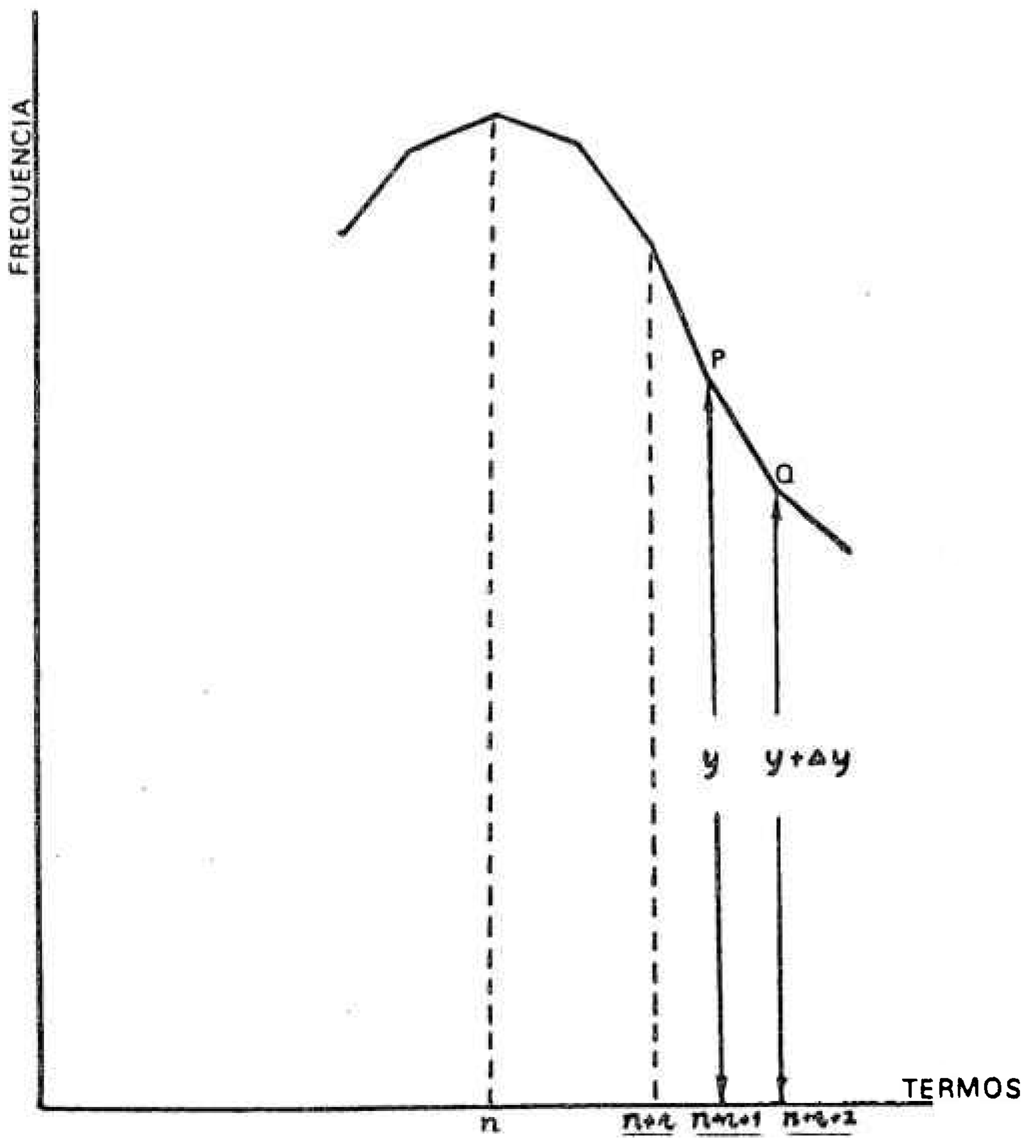

Figura 3.3 — Representação de termos do desenvolvimento da expressão $(p + q)^{2n}$.

Vamos considerar a expansão de $(1/2 + 1/2)^{2n}$. Escolhemos **2n** como expoente por conveniência, porque o número de termos da expansão é sempre o expoente mais um e assim teríamos 2n+1 termos o que assegura um termo médio. Aplicando (III.9) ao caso considerado temos:

$$\sigma = \sqrt{2n\ 1/2\ .\ 1/2} = \sqrt{n/2\ .\ 1\ .\ 1}$$

ou

$$2\sigma^2 = n(\Delta x)^2 \qquad \text{(III.10)}$$

Seja n um número finito qualquer e vamos considerar dois termos consecutivos da série expandida: o termo r^{esimo} e $(r+1)^{esimo}$ a partir do centro. (Figura 3.3). Em P temos:

$$y = {}_{2n}C_{n+r}\left(\frac{1}{2}\right)^{2n} = \left(\frac{1}{2}\right)^{2n}\frac{2n!}{(n-1)!\,(n+r)!} \qquad \text{(III.11)}$$

Em Q temos:

$$y + \Delta y = {}_{2n}C_{n+r+1}\left(\frac{1}{2}\right)^{2n} = \left(\frac{1}{2}\right)^{2n}\frac{2n!}{(n-r-1)!\,(n+r+1)!} \qquad \text{(III.12)}$$

Dividindo (III.12) por (III.11) temos:

$$\frac{y+\Delta y}{y} = \frac{n-r}{n+r+1}$$

$$\Delta y = -\frac{2r+1}{n+r+1}\,y \qquad \text{(III.13)}$$

Dividindo os dois membros de (III.13) por Δx, vamos obter:

$$\frac{\Delta y}{\Delta x} = -\frac{2r+1}{n+r+1}\,\frac{y}{\Delta x}$$

Multiplicando o 2º Membro por Δx/Δx se obtém:

$$\frac{\Delta y}{\Delta x} = -\frac{2\,r\Delta x + \Delta x}{n(\Delta x)^2 + r(\Delta x)^2 + (\Delta x)^2}\,y \qquad \text{(III.14)}$$

O valor de x em P, considerando o termo central n+1 = 0, é: $x = r\Delta x$. Substituindo este valor e o valor de σ^2 de (III.10) em (III.14) temos:

$$\frac{\Delta y}{\Delta x} = -\frac{2x + \Delta x}{2\sigma^2 + x\Delta x + (\Delta x)^2}\,y$$

Fazendo $n \to \infty$ e $x \to 0$, no limite teremos:

$$\frac{dy}{dx} = -\frac{2xy}{2\sigma^2} = \frac{xy}{\sigma^2}$$

ou

$$\frac{dy}{y} = -\frac{xdx}{\sigma^2}$$

cuja integral dá:

$$\ln y = -\frac{x^2}{2\sigma^2} + \ln C$$

ou

$$y = C\, e^{-\frac{x^2}{2\sigma^2}} \qquad \text{(III.15)}$$

III.3.1 — Propriedades da Curva de Erros

1) Como x no expoente está elevado ao quadrado, para x=a ou x=-a, o valor de y é o mesmo, isto é a curva é simétrica em relação a x=0 (centro).

2) Para x =0:

$$y = C$$

3) Examinando a forma da curva, observam-se pontos de inflexão em relação a x=0. Para localizá-los igualamos a segunda derivada a zero

$$\frac{dy}{dx} = C\, e^{-\frac{x^2}{2\sigma^2}} \left(-\frac{x}{\sigma^2}\right) = -\frac{Cx}{\sigma^2}\, e^{-\frac{x^2}{2\sigma^2}}$$

$$\frac{d^2y}{dx^2} = -\frac{C}{\sigma^2}\left[x e^{-\frac{x^2}{2\sigma^2}} \left(-\frac{x}{2}\right) + e^{-\frac{x^2}{2\sigma^2}}\right]$$

$$\frac{d^2y}{dx^2} = \frac{C}{\sigma^2}\, e^{-\frac{x^2}{2\sigma^2}} \left(\frac{x^2}{\sigma^2} - 1\right)$$

Para que a segunda derivada se anule, temos duas possibilidades:

$$e^{-\frac{x^2}{2\sigma^2}} = 0,$$

que é satisfeita para $x = \pm \infty$ ou então:

$$\frac{x^2}{\sigma^2} - 1 = 0$$

cuja solução é:

$$x = \pm \sigma \qquad \text{(III.16)}$$

isto é, os pontos de inflexão estão localizados nos pontos da curva que caracterizam a dispersão. De acordo com a dedução de (III.15), um erro de valor x tem como coeficiente de freqüência o valor da ordenada correspondente, y. Pode-se concluir então que um erro x, compreendido num intervalo infinitesimal dx, tem como probabilidade a área do retângulo de altura y e largura dx ou:

$$p_1 = y_1 \, dx$$

Como a ocorrência de um valor x no intervalo $-\infty \rightarrow +\infty$ é certa, isto é, a probabilidade é 1, temos que:

$$\int_{-\infty}^{+\infty} y \, dx = C \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx = 1$$

Sabe-se que

$$\int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{\pi} \cdot$$

No nosso caso $z = \dfrac{x}{\sigma\sqrt{2}}$ e $dz = \dfrac{dx}{\sigma\sqrt{2}}$

Então:

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{dx}{\sigma\sqrt{2}} = \sqrt{\pi}$$

$$\therefore \int_{-\infty}^{+\infty} e^{\frac{x^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi}$$

Como $C\sigma\sqrt{2\pi} = 1$, segue-se que:

$$C = \frac{1}{\sigma\sqrt{2\pi}} \tag{III.17}$$

e temos a equação final:

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}} \tag{III.18}$$

Observa-se pela equação (III.17) que o valor de C que dá o valor máximo para y, ordenada de $x = 0$, é inversamente proporcional a σ, isto é, quanto maior a dispersão dos erros, mais achatada é a curva.

III.3.2 – Demonstração pelo Método dos Mínimos Quadrados que a Média Aritmética é o Valor mais Provável

Suponhamos uma série de medidas $x_1, x_2, \ldots x_n$ obtidas com a mesma precisão e seja z o valor mais provável. Os desvios das medidas serão $(x_1 - z), (x_2 - z) \ldots (x_n - z)$ respectivamente. Admitindo esses desvios como sendo erros, a probabilidade de um desvio $(x_i - z)$ é:

$$y_i = C\, e^{-\frac{(x_i - z)^2}{2\sigma^2}}$$

Como os erros são independentes, a probabilidade de ter todos os erros é o produto das probabilidades.

$$P = C\, e^{-\frac{(x_1 - z)^2}{2\sigma^2}}\; C\, e^{-\frac{(x_2 - z)^2}{2\sigma^2}} \ldots C\, e^{-\frac{(x_n - z)^2}{2\sigma^2}}$$

$$P = C^n\, e^{-\frac{1}{2\sigma^2}\left[(x_i - z)^2 + \ldots + (x_n - z)^2\right]}$$

A probabilidade será máxima, quando o expoente de $\underline{e}$ for *mínimo*, isto é,

$$Q = (x_1 - z)^2 + (x_2 - z)^2 + \ldots + (x_n - z)^2$$

deve ser mínimo

As condições de mínimos são:

$$\frac{dQ}{dZ} = 0 \quad ; \quad \frac{d^2Q}{dZ^2} > 0$$

Temos:

$$\frac{dQ}{dZ} = -2(x_1 - z) - 2(x_2 - z) - \ldots - 2(x_n - z) = 0$$

$$\frac{d^2Q}{dZ^2} = 2 + 2 + \ldots + 2 = 2n > 0 \therefore \text{mínimo}$$

Para $\frac{dQ}{dZ}$ ser zero, devemos ter:

$$nZ = \sum_{i=1}^{n} x_i$$

$$Z = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

Isto demonstra que o valor mais provável das medidas é a média aritmética e a soma dos quadrados dos desvios em relação à média aritmética é menor que a soma dos quadrados dos desvios tomados em relação a qualquer outro valor.

III.3.3 – Densidade Normalizada da Distribuição de Gauss

Até aqui vimos a distribuição normal definida pelos parâmetros $\mu = 0$ e σ. A expressão mais geral é então:

$$y = \varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x \ \mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty) \qquad \text{(III.19)}$$

$$\psi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \qquad \text{(III.20)}$$

A probabilidade que a variável x assuma um valor compreendido entre x_1 e x_2 será:

$$\psi(x_2) - \psi(x_1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad \text{(III.21)}$$

Para simplificar, pode-se definir uma nova variável u como sendo:

$$u = \frac{x - \mu}{\sigma} \qquad \text{(III.22)}$$

Neste caso, vamos ter:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{III.23}$$

que é chamada a densidade normalizada da distribuição de Gauss cujos parâmetros são (0,1).

A integral da função normalizada é definida pela expressão:

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-u^2/2}\, du \tag{III.24}$$

que dá a probabilidade da variável assumir qualquer valor no intervalo $-\infty$ - u

EXERCÍCIO 10

Achar a probabilidade que u assuma valores entre -0,90 e + 0,70.

$$\begin{aligned} P(-0{,}90 < u < 0{,}70) &= \phi(0{,}70) - \phi(-0{,}90) \\ &= 0{,}7580 - 0{,}1841 = 0{,}5739 \\ &= 57{,}39\% \end{aligned}$$

Os valores usados são encontrados em Tabelas. Existem também Tabelas que levam em conta a simetria da curva e os valores tabelados correspondem a:

$$2\,\theta(u) = \frac{1}{\sqrt{2\pi}} \int_{0}^{u} e^{-\frac{u^2}{2}}\, du$$

Com os valores tabelados da função da distribuição normal podem ser estabelecidos limites de confiança para resultados experimentais, se conhecermos o desvio padrão do método usado.

EXERCÍCIO 11

Qual a probabilidade de obter um resultado x cujo valor esteja compreendido entre os limites estabelecidos por $\pm 2\sigma$, isto é:

$$x - \mu = \pm 2\sigma$$

Os limites dos valores assumidos são:

$$u_1 = \frac{x_1 - \mu}{\sigma} = -2$$

$$u_2 = \frac{x_2 - \mu}{\sigma} = 2$$

Na tabela, achamos que:

$$P(-2{,}0 < \mu < 2{,}0) = 0{,}9545$$

isto é, a probabilidade é de 95,45%

Em muitas aplicações práticas, o intervalo limitado por $\pm\, 2\, \sigma$ define os valores x_1(mínimo) e x_2(máximo) que podem ser esperados, isto é:

$$|2\sigma| = x_2 - \mu = x_1 - \mu$$

é o erro máximo permissível. O conceito de erro máximo assim definido é válido só no caso de ser conhecido o valor de σ do método usado. Na prática, conhece-se s (estimativa de σ) e seu valor depende do número de resultados paralelos obtidos.

III.3.4 – Propriedades Aditivas da Distribuição Normal

A soma de duas distribuições normais cujos parâmetros são (μ_x, σ_x^2) e (μ_y, σ_y^2) é também uma distribuição normal com os parâmetros ($\mu_{(x+y)}$, $\sigma^2_{(x+y)}$) onde:

$$\mu_{x+y} = \mu_x + \mu_y$$

e

$$\sigma^2 (x + y) = \sigma_x^2 + \sigma_y^2$$

Isto é válido também para o somatório de n variáveis aleatórias de uma distribuição normal. Segue-se que a média de n variáveis aleatórias independentes de uma distribuição normal com parâmetros (μ_x, σ_x^2) se distribue como uma normal com parâmetros (μ, $\sigma_{\bar{x}}^2$) onde $\sigma_{\bar{x}}^2 = \sigma_x^2/n$. Normalizando esta nova variável, temos:

$$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad (III.25)$$

EXERCÍCIO 12

Sabe-se que um certo padrão de urânio contém 0,300% de U_3O_8. Na determinação de urânio neste padrão foi usado um método cujo coeficiente de variação é de 10%. Deseja-se saber qual a probabilidade da média de 25 determinações estar no intervalo 0,297 – 0,303%.

OBSERVAÇÃO – O coeficiente de variação é definido por

$$V = \frac{100\sigma}{\mu} \tag{III.26}$$

Neste caso conhecemos: $\mu = 0{,}300$ e $\sigma_x = 0{,}03$. Então:

$$\sigma_{\bar{x}} = \frac{0{,}03}{\sqrt{25}} = 0{,}006$$

$$u_1 = \frac{\bar{x}_1 - \mu}{\sigma_{\bar{x}}} = \frac{0{,}297 - 0{,}300}{0{,}006} = -0{,}5$$

$$u_2 = \frac{\bar{x}_2 - \mu}{\sigma_{\bar{x}}} = \frac{0{,}303 - 0{,}300}{0{,}006} = 0{,}5$$

$$P(-0{,}5 < u < 0{,}5) = 2\,\theta(0{,}5) = 0{,}3830$$

isto é, a probabilidade é de 38,30%

CAPÍTULO IV

DISTRIBUIÇÕES RELACIONADAS COM A DISTRIBUIÇÃO NORMAL – APLICAÇÕES

Neste Capítulo são analisadas as distribuições que se aplicam aos casos práticos, isto é, quando o número de observações é limitado. Estas distribuições estão relacionadas com a distribuição normal, porque se aplicam a conjuntos de números ou de resultados experimentais que obedecem à distribuição normal.

A distribuição de Poisson que se aplica, entre outros, à interpretação dos fenômenos radioativos, será examinada num Capítulo a parte.

IV.1 – Distribuição t de Student

Geralmente na prática, não se conhece o valor verdadeiro de σ. O número de resultados experimentais de uma medida é quase sempre pequeno e calcula-se s, estimativa de σ, a partir dos resultados obtidos. Se substituirmos σ por s, calculado nessas condições, não se obtém uma distribuição normal, porque o valor verdadeiro de um dos parâmetros da distribuição, isto é, da variância, não é conhecido. Consequentemente, a teoria dos erros, baseada na distribuição normal, não poderia ser aplicada a um número pequeno de resultados. Este problema foi resolvido com a distribuição t de Student e permitiu o desenvolvimento da microestatística ou estatística aplicada a um número pequeno de observações. A estimativa de σ^2 por s^2 é tanto mais válida, quanto maior for o número n de resultados, ou melhor, quanto maior for o número de graus de liberdade usado para o cálculo de s^2. A expressão:

$$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

quando não se conhece $\sigma_{\bar{x}}$, torna-se:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad \text{(IV.1)}$$

O valor de t é calculado pela distribuição de Student cuja densidade de probabilidade é dada pela expressão:

$$\varphi(t) = \frac{1}{\sqrt{\pi f}} \frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})} (1 + \frac{t^2}{f})^{-(\frac{f+1}{2})} \qquad \text{(IV.2)}$$

para valores de t variando de $-\infty$ a $+\infty$

O número de graus de liberdade com que o valor de s^2 foi calculado é indicado por f e $\tau(f)$ é a função gama definida pela integral

$$\Gamma(u + 1) = \int_0^\infty e^{-x} x^u \, dx$$

que pode ser considerada como a generalização do fatorial de um número natural. Se u é inteiro e positivo

$$\Gamma(u + 1) = u!$$

As relações

$$\Gamma(u + 1) = u!$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

permitem calcular a função gama de qualquer valor de u múltiplo de 1/2.

Assim:

$$\Gamma(\frac{3}{2}) = \Gamma(\frac{1}{2} + 1) = \frac{1}{2}\,\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$$

Pela expressão (IV.1) vemos que a distribuição t só depende do número de graus de liberdade com que o valor de s^2 foi calculado. A curva da função t é semelhante a uma gaussiana. Para valores de f pequenos a curva se aproxima da abcissa bem lentamente e para valores de $f \to \infty$ a curva se aproxima à curva da distribuição normal. A probabilidade que a variável t caia num intervalo $t_{p/2} - t_{1-p/2}$ é definida pela expressão:

$$P(t_{p/2} < t < t_{1-p/2}) = 1 - p \qquad \text{(IV.3)}$$

o que pode ser visto pela Figura 4.1.

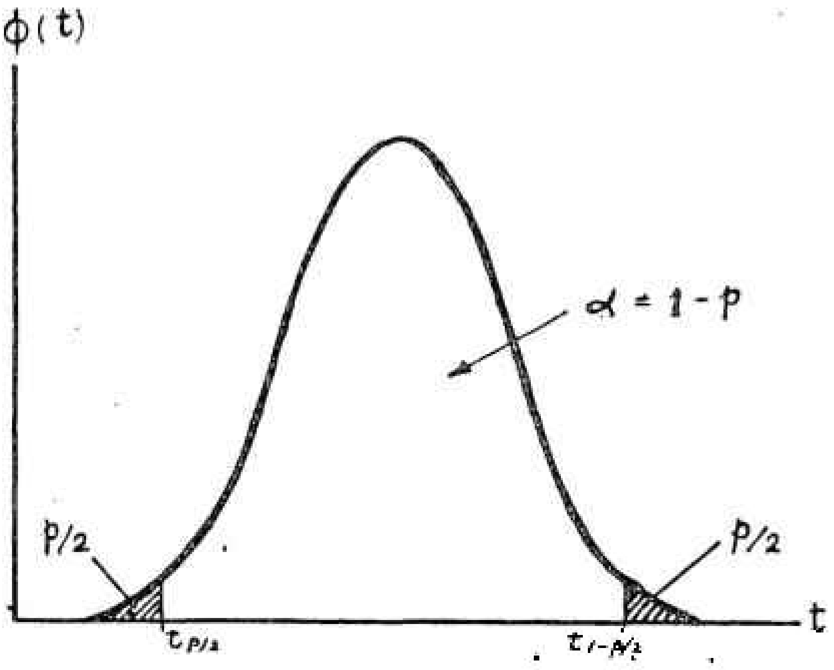

Figura 4.1 – Probabilidade de $t_{p/2} < t < t_{1-p/2}$.
Teste Bicaudal

Pela simetria da curva de distribuição segue-se que:

$$t_{p/2} = -t_{1-p/2}$$

Assim a probabilidade que t esteja fora dos limites do intervalo definido por $t_{p/2}$ e $t_{1-p/2}$ será:

$$p = \int_{-\infty}^{t_{p/2}} \varphi(t)\,dt + \int_{t_{1-p/2}}^{\infty} \varphi(t)\,dt = 2\int_{t_{1-p/2}}^{\infty} \varphi(t)\,dt = 2\int_{-\infty}^{t_{p/2}} \varphi(t)\,dt \qquad \text{(IV.4)}$$

que é o complemento de (IV.3).

Em alguns tipos de problemas, interessa saber a probabilidade de t ser maior ou menor que um certo valor t_p. Esse é um tipo de teste chamado monocaudal, (Figura 4.2) enquanto que, no caso anterior, o teste é bicaudal. Em virtude de existirem esses dois tipos de testes, encontram-se tabelas dos valores de t para o critério monocaudal ou bicaudal. Ao procurar o valor de t nas tabelas, é preciso verificar qual o critério adotado na construção da tabela.

Geralmente usa-se para t a notação $t_{p,f}$ onde f = *graus* de liberdade e p é a probabilidade de não cair no intervalo considerado (nível de significância). Os valores de t em função dos graus de liberdade são tabelados para as probabilidades p = 0,20, 0,10, 0,05 . . . etc. Por exemplo, se p = 0,10 e f = 10 podemos ter dois casos:

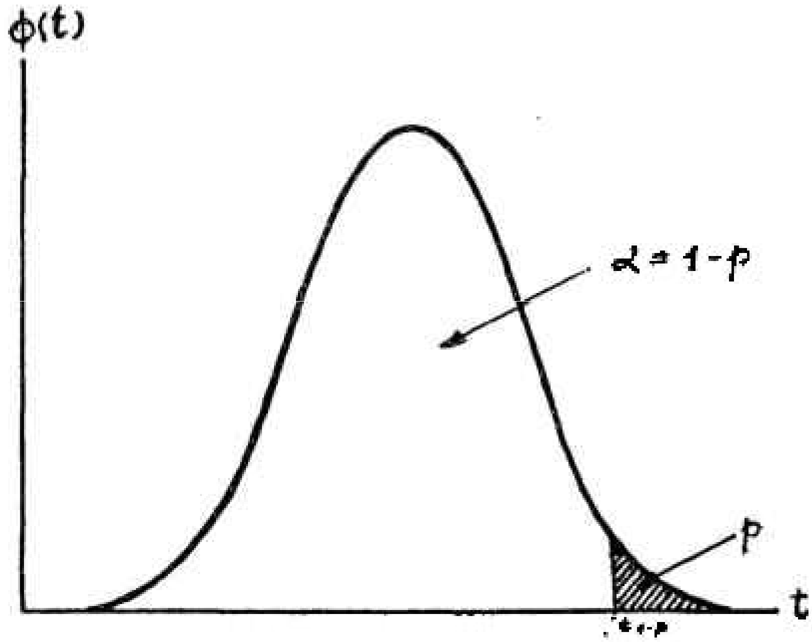

Figura 4.2 — Probabilidade de $t < t_{1-p}$.

Teste Monocaudal

1º) Teste Bicaudal

Procura-se numa tabela construída segundo o critério bicaudal o valor de t para $p = 0,10$ e $f = 10$. Acha-se $t = 1,81$ o que significa que:

$$P(t < 1,81) = 0,95$$
$$P(t < -1,81) = 0,05$$

e

$$P(-1,81 < t < 1,81) = 0,90$$
$$P(|t| > 1,81) = 0,10$$

2º) Teste Monocaudal

Quando dispomos só de tabelas construídas segundo o critério bicaudal, isto é p/2 de cada lado, deve-se procurar o valor de t correspondente a $p = 0,20$ com 10 graus de liberdade. Obviamente o valor de t neste caso será menor que no caso anterior e teremos diretamente.

$$P(t > 1,37) = 0,10$$
$$P(t < 1,37) = 0,90$$

ou se o teste for para o lado dos valores de t negativos teremos:

$$P(t > -1,37) = 0,90$$
$$P(t < -1,37) = 0,10$$

OBSERVAÇÃO

Se a tabela de t consultada for construída segundo o critério monocaudal, para o primeiro caso (teste bicaudal) para $p = 0,10$ deve-se procurar o valor de t na coluna de $p = 0,05$ e no 2º caso diretamente para $p = 0,10$.

A partir de $f = 20$, a distribuição t se aproxima da normal, mas para valores pequenos de f a discrepância é grande.

Pelas equações (IV.1) e (IV.3) temos:

$$P\left(t_{p/2} < \frac{\bar{x} - \mu}{s_{\bar{x}}} < t_{1-p/2}\right) = 1 - p \qquad \text{(IV.5)}$$

de onde se podem tirar os limites de confiança para a média verdadeira μ no nível de significância p.

$$P(\bar{x} - t_{p/2}\, s_{\bar{x}} < \mu < \bar{x} - t_{1-p/2}\, s_{\bar{x}}) = 1 - p \qquad \text{(IV.6)}$$

A probabilidade $\alpha = 1 - p$ é chamado nível de confiança que é o complemento do nível de significância p. O nível de confiança corresponde à probabilidade de encontrar o valor verdadeiro no intervalo calculado e o nível de significância corresponde à probabilidade desse valor cair fora do intervalo.

EXERCÍCIO 13

A análise de um minério de tório deu os seguintes resultados em porcentagem: 2,52, 2,60, 2,43, e 2,41. Calcular os limites de confiança para a média num nível de confiança de 0,95.

Calcula-se a média x, estimativa de μ:

$$\bar{x} = \frac{2{,}52 + 2{,}60 + 2{,}43 + 2{,}41}{4} = 2{,}49$$

Calcula-se o desvio padrão

$$s_x = \sqrt{\frac{\sum_{i=1}^{4} (2{,}49 - x_i)^2}{3}} = 0{,}088$$

Calcula-se o desvio padrão da média

$$s_{\bar{x}} = \frac{0{,}088}{\sqrt{n}} = \frac{0{,}088}{2} = 0{,}44$$

O valor de t para f = 3 e p = 0,05 é $t_{0,05,3} = 3{,}18$ (bicaudal). Então:

$$t\, s_{\bar{x}} = 3{,}18 \times 0{,}044 = 0{,}14$$

Acha-se então que num nível de confiança de 0,95 ou 95%

$$\mu = 2{,}49 \pm 0{,}14$$

ou $$2{,}35 < \mu < 2{,}63$$

EXERCÍCIO 14

Suponhamos agora que o minério de tório analisado é um padrão e sabe-se que contém 2,39% de tório. Deseja-se saber se a divergência entre o valor real e a média achada é estatisticamente significativa num nível de confiança de 0,95.

Deve-se calcular o valor de t e comparar com o valor tabelado

$$t = \frac{|x - \mu|}{s_{\bar{x}}} = \frac{2{,}49 - 2{,}39}{0{,}044} = 2{,}27$$

Na tabela achamos que $t_{0,05,3} = 2,35$. Como o valor calculado é menor que o valor tabelado aceita-se o valor 2,49 como estimativa de $\mu = 2,39$.

Note-se que neste caso o teste foi monocaudal, isto é, testamos se 2,49 > 2,39 nas condições do problema.

EXERCÍCIO 15

O mesmo problema pode ser examinado por um outro aspecto. Vamos admitir que as análises de tório foram feitas por um método bem padronizado cujo coeficiente de variação é de 2%. Neste caso, pode-se supor σ conhecido, isto é, $(2,49 \times 2)/100 = 0,05$ e então devem ser usados os valores da probabilidade da distribuição normal. Deve-se calcular:

$$u = \frac{|\bar{x} - \mu|}{\sigma_x/\sqrt{n}} = \frac{2,49 - 2,39}{0,05/2} = 4,00$$

Na tabela achamos que $P(\mu < 4,0) = 0,999968$ o que corresponde a um nível de significância 0,000032. Nestas condições não se pode aceitar 2,49 como estimativa de $\mu = 2,39$ e deve-se reconhecer a existência de um erro sistemático nas condições em que foram obtidos os resultados experimentais.

Em aplicações práticas, é costume adotar o nível de significância de 5%, mas em casos especiais pode-se aceitar um nível de significância de 1%. Na prática , quando se obtém um resultado para t experimental muito próximo de $t_{0,05,f}$, é conveniente obter mais resultados para evitar erros na aceitação ou rejeição da hipótese formulada. Nesse tipo de problemas, é costume formular a hipótese nula H_0 e a alternativa H_1. No caso do último exercício, por exemplo, as hipóteses são:

$$\left.\begin{array}{l} H_0 : \bar{x} = \mu \\ H_1 : \bar{x} > \mu \end{array}\right\} \qquad (IV.7)$$

A hipótese nula corresponde sempre a admitir que o efeito observado é insignificante, no nível de confiança adotado. As hipóteses de (IV.7) correspondem a um teste monocaudal. No caso de teste bicaudal as hipóteses correspondentes seriam:

$$\left.\begin{array}{l} H_0 : \bar{x} = \mu \\ H_1 : \bar{x} \neq \mu \end{array}\right\} \qquad (IV.8)$$

A hipótese alternativa H_1 é que estabelece o tipo do teste.

IV.2 – Distribuição χ^2

Assim como é preciso estabelecer um intervalo de confiança para μ, usando sua estimativa $\bar{x}$, deve-se também estabelecer um intervalo de confiança para σ, quando é conhecida sua estimativa s. Para o caso da média utiliza-se a distribuição t e para o caso do desvio padrão utiliza-se a distribuição χ^2.

Se temos f variáveis aleatórias $u_1, u_2, \ldots u_f$ que se distribuem segundo a distribuição normal com parâmetros $\mu = 0$ e $\sigma^2 = 1$, o somatório $\sum_{i=1}^{k} u_i^2$ é chamado χ^2 (f graus de liberdade). Este é um caso particular mas, de um modo mais geral, vamos supor que n variáveis aleatórias $x_1, x_2 \ldots x_n$ são independentes e com distribuição normal. Pode-se demonstrar que:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2 = \frac{n-1}{\sigma_x^2} s_x^2 \qquad \text{(IV.9)}$$

com n — 1 graus de liberdade

A densidade de probabilidade da distribuição dos valores de χ^2 é definida pela expressão

$$\varphi(\chi^2) = \frac{1}{2^{f/2}\,\Gamma(f/2)} (\chi^2)^{(f-2)/2} e^{-\chi^2/2} \qquad (0 \leqslant \chi^2 < \infty) \qquad \text{(IV.10)}$$

onde f é o número de graus de liberdade. As curvas da densidade de probabilidade da distribuição χ^2 são assimétricas. A assimetria diminue com o aumento do número de graus de liberdade (Figura 4.3). Por (IV.10) pode-se ver que o valor de χ^2 só depende do número de graus de liberdade e geralmente é representado por $\chi^2_{p,f}$. Por exemplo, $\chi^2_{0,05,6} = 12{,}6$ significa que

$$P(\chi^2_6 > 12{,}6) = 0{,}05$$

Pela relação:

$$P(\chi^2_{p_1} < \chi^2 < \chi^2_{p_2}) = p_1 - p_2 \qquad \text{(IV.11)}$$

podem ser obtidos os limites de confiança para χ^2 com um nível de confiança $\alpha = p_1 - p_2$. Por exemplo, os limites de confiança para χ^2 com 5 graus de liberdade, num nível de confiança de 80%, são determinados procurando na Tabela de χ^2 os valores

$$\chi^2_{0,90,5} = 1{,}61 \quad \text{e} \quad \chi^2_{0,10,5} = 9{,}2$$

$$P(1{,}61 < \chi^2_5 < 9{,}2) = 0{,}90 - 0{,}10 = 0{,}80$$

Por meio das expressões (IV.9) e (IV.11) obtemos os limites de confiança da variância da população

$$P\left(\frac{(n-1)\, s_x^2}{\chi^2_{p_2}} < \sigma_x^2 < \frac{(n-1)\, s_x^2}{\chi^2_{p_1}}\right) = p_1 - p_2 \qquad \text{(IV.12)}$$

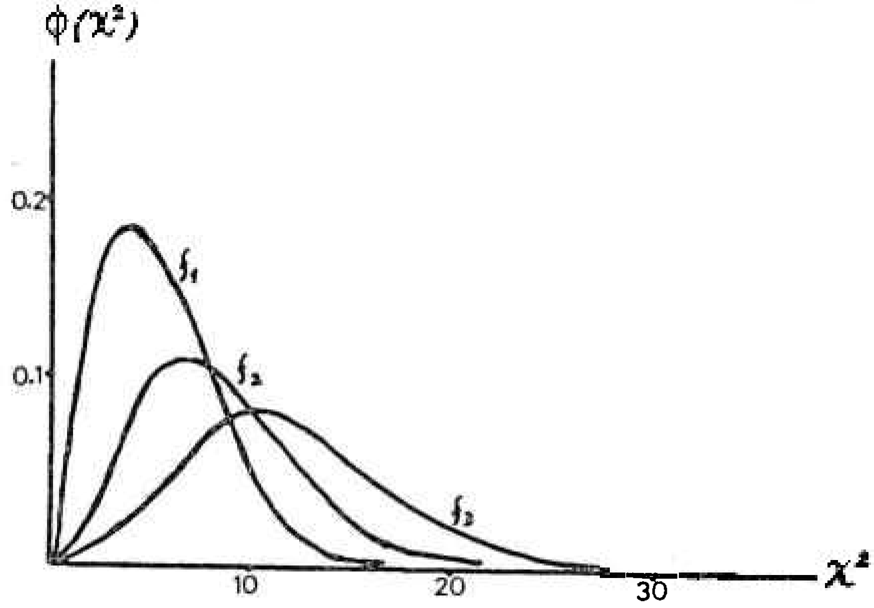

Figura 4.3 — Densidade de probabilidade da distribuição χ^2 para $f_1 < f_2 < f_3$.

EXERCÍCIO 16

Foram obtidos 10 resultados experimentais de uma medida e foi achado para s o valor 0,25. Quais os limites de confiança para σ num nível de confiança $\alpha = 0{,}90$?

Na Tabela achamos: $\chi^2_{0,05,9} = 18{,}3$ e $\chi^2_{0,95,9} = 3{,}94$

Os limites procurados são dados por:

$$\sqrt{\frac{9}{18{,}3}} \cdot 0{,}25 < \sigma < \sqrt{\frac{9}{3{,}94}} \cdot 0{,}25$$

$$0{,}175 < \sigma < 0{,}378$$

Pode-se notar que os limites são assimétricos. Para $n \geqslant 30$ os limites são praticamente simétricos, porque a variável aleatória s_x passa a obedecer à distribuição normal com média $\bar{s}$ e desvio padrão:

$$s_{\bar{s}} = \frac{s_x}{\sqrt{2f}} \qquad \text{(IV.13)}$$

Neste caso, os limites de confiança para o desvio padrão são definidos por:

$$s_x - \frac{u s_x}{\sqrt{2f}} < \sigma < s_x + \frac{u s_x}{\sqrt{2f}} \qquad \text{(IV.14)}$$

EXERCÍCIO 17

Achar os limites de confiança para o desvio padrão, $s_x = 0{,}40$, obtido a partir de 30 resultados, num nível de confiança de 0,90.

Na Tabela da distribuição normal, acha-se que para a condição $2\ \theta(u) = 0{,}90$, o valor correspondente de u é 1,645. Assim os limites de confiança para σ são definidos por

$$0{,}40 - \frac{1{,}645 \times 0{,}40}{\sqrt{2 \times 29}} < \sigma < 0{,}40 + \frac{1{,}645 \times 0{,}40}{\sqrt{2 \times 29}}$$

$$0{,}40 - 0{,}086 < \sigma < 0{,}40 + 0{,}086$$

$$0{,}314 < \sigma < 0{,}486$$

Se usarmos a fórmula (IV.12), o valor achado será:

$$\chi^2_{0{,}95,29} = 17{,}7 \quad \text{e} \quad \chi^2_{0{,}05,29} = 43{,}2$$

$$\sqrt{\frac{29}{43{,}2}} \cdot 0{,}40 < \sigma < \sqrt{\frac{29}{17{,}7}} \cdot 0{,}40$$

$$0{,}328 < \sigma < 0{,}512$$

Vemos então que o uso dos valores de u da distribuição normal dão um resultado que se aproxima razoavelmente do obtido por meio de χ^2. Este resultado será idêntico para $f \to \infty$

IV.3 — Distribuição F

É usada para comparar as variâncias s^2_1 e s^2_2 calculadas a partir de 2 séries de resultados x_1, $x_2, \ldots x_n$ e $y_1, y_2, \ldots y_n$. A comparação entre variâncias póde indicar qual o método mais preciso para um certo tipo de trabalho, se dois operadores trabalham com o mesmo cuidado, etc.

A densidade de probabilidade da distribuição F é definida por:

$$\varphi(F) = \frac{\Gamma(\frac{f_1 + f_2}{2})}{\Gamma(\frac{f_1}{2})\,\Gamma(\frac{f_2}{2})}\, f_1^{(f_1/2)}\, f_2^{(f_2/2)}\, \frac{F^{(f_1 - 2)/2}}{(f_2 + f_1 F)^{(f_1 + f_2)/2}} \qquad (0 < F < \infty)$$

(IV.15)

As curvas são assimétricas e a forma (Figura 4.4) só depende do número de graus de liberdade f_1 e f_2 das duas variâncias. Os valores de F são encontrados em tabelas onde, na primeira linha, estão os valores de f_1 (número de graus de liberdade da variância numericamente maior) e na primeira coluna a esquerda os valores de f_2. Cada tabela corresponde a um nível de significância para o valor de F. Pode-se notar que $F_{(f_1, f_2)} \neq F_{(f_2, f_1)}$. Para aplicar o teste F na comparação de duas variâncias, testa-se somente a relação

$$\frac{s_1^2}{s_2^2} > 1$$

isto é, no numerador coloca-se sempre o valor da variância numericamente maior. Mesmo assim, como para o teste t, existem os testes monocaudal e bicaudal, conforme será ilustrado pelos seguintes exemplos.

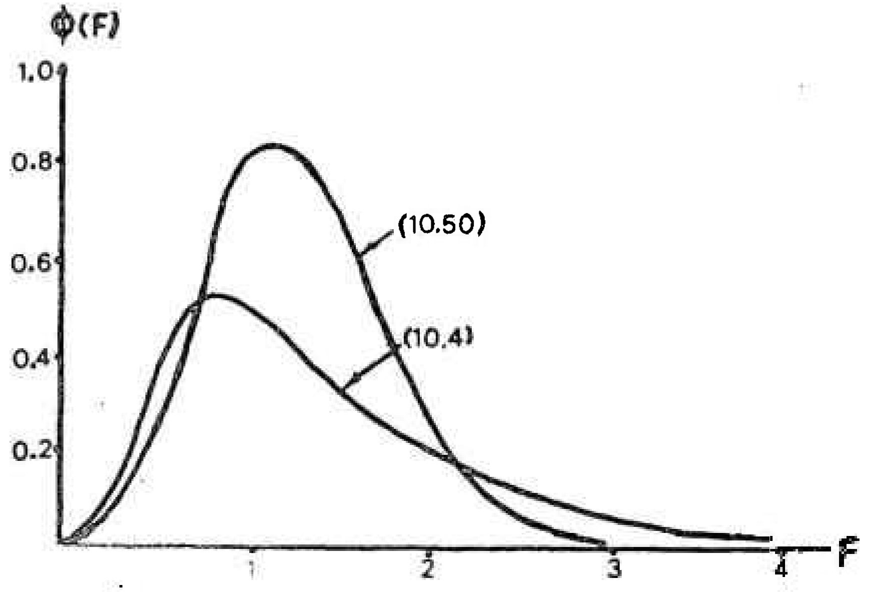

Figura 4.4 — Densidade de probabilidade da distribuição F para dois valores de (f_1, f_2).

EXERCÍCIO 18

Suspeita-se da precisão de um aparelho A que já tem muitos anos de uso. Para testar seu funcionamento, foram feitas n_1 medidas com esse aparelho e n_2 medidas com um aparelho B igual ao primeiro, mas em perfeitas condições de funcionamento. A variância calculada com as $n_1 = 10$ medidas feitas com A deu 0,36 e a variância obtida com as $n_2 = 8$ medidas feitas com B deu 0,21. Deseja-se saber se as variâncias são estatisticamente iguais num nível de significância 0,05.

Neste caso, sabemos "a priori" que a variância das medidas obtidas com o aparelho A só pode ser igual ou maior que a variância das medidas obtidas com o aparelho B. Temos então um teste monocaudal e as hipóteses são:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 > \sigma_B^2$$

A hipótese nula será rejeitada se:

$$\frac{s_A^2}{s_B^2} > F_{p(f_A, f_B)} \qquad \text{(IV.16)}$$

Vamos obter para F o valor:

$$F = \frac{0{,}36}{0{,}21} = 1{,}714$$

Procuramos na Tabela o valor de $F_{0,05(9,7)}$ e achamos o valor 3,7. Como o valor de F tabelado é maior que o valor de F calculado, aceita-se H_0, isto é, as variâncias são estatisticamente iguais e não há razão para suspeitar do aparelho A. É conveniente observar que, num problema deste tipo, em que o funcionamento de um aparelho é de vital importância, deve-se obter o maior número possível de resultados (n_1 e n_2 grandes) ou então usar um nível de significância maior.

EXERCICIO 19

Foram aplicados dois métodos (A e B) na medida de uma grandeza e deseja-se saber se os dois métodos são igualmente precisos. Neste caso, não se sabe nada sobre a precisão dos métodos e as hipóteses formuladas serão:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

Suponhamos ${s^2}_A > {s^2}_B$ numericamente. Neste caso, temos um teste bicaudal e a hipótese nula será rejeitada se:

$$\frac{s_A^2}{s_B^2} > F_{p/2(f_A, f_B)} \qquad \text{(IV.17)}$$

ou

$$\frac{s_A^2}{s_B^2} < F_{1-p/2(f_A,f_B)} \qquad \text{(IV.18)}$$

Da relação:

$$P\left(\frac{s_A^2}{s_B^2} > F_{p/2(f_A,f_B)}\right) = \frac{p}{2}$$

segue-se que:

$$P\left(\frac{s_B^2}{s_A^2} < \frac{1}{F_{p/2(f_A,f_B)}}\right) = \frac{p}{2}$$

e é suficiente testar a desigualdade (IV.17).

Suponhamos, no caso, que foram obtidos 13 resultados com o método A e a variância deu 0,58 e 10 resultados com o método B dando uma variância 0,26. A relação será:

$$F = \frac{0{,}58}{0{,}26} = 2{,}23$$

Se quisermos testar a igualdade das variâncias num *nível de significância* 0,10, procuramos o valor de $F_{0,05(12,9)}$ numa Tabela construída segundo o critério monocaudal ou $F_{0,10(12,9)}$ numa Tabela em que os valores de F seguem o critério bicaudal. Acha-se o valor 3,1. Não podemos então rejeitar H_0, isto é, os dois métodos são igualmente precisos, nesse nível de significância.

Pode-se demonstrar que as distribuições t e χ^2 são casos particulares da distribuição F. Assim:

$$F_{(f_1,\infty)} = \frac{\chi^2_{(f_1)}}{f_1}$$

e (IV.19)

$$F_{(1,f_2)} = t^2_{(f_2)}$$

Exemplos:

$$F_{0,50(4,\infty)} = 2{,}4 \quad \text{e} \quad \frac{\chi^2_{0,05(4)}}{4} = \frac{9{,}5}{4} = 2{,}4\,.$$

$$F_{0,05(1,10)} = 5{,}0 \quad \text{e} \quad t^2_{0,05,10} = (2{,}23)^2 = 5{,}0$$

IV.4 — Critério para Verificar se um Conjunto de Resultados Experimentais Obedece à Distribuição Normal

Podemos encontrar dois tipos de problemas:

1) Verificar se a distribuição dos resultados pode ser considerada uma distribuição normal, no nível de confiança desejado. Se o teste permitir a aceitação da hipótese num nível de significância mais baixo, isto é, se o afastamento da distribuição normal não for muito grande, é preciso investigar as causas para alterar as condições experimentais de tal maneira que essas causas sejam eliminadas. Se o número de resultados obtidos for muito grande, é possível aplicar os testes F, t e χ^2 mesmo que haja um afastamento não muito grande da distribuição normal.

2) Resolver o problema se a distribuição for essencialmente diferente da distribuição normal, o que só é possível por comparação da distribuição obtida experimentalmente com outras distribuições teóricas. Neste caso não é possível uma interpretação estatística a não ser que, por meio de transformações convenientes aplicadas aos resultados obtidos, seja possível obter uma distribuição normal.

É preciso lembrar que, geralmente o número de resultados experimentais é pequeno e aceita-se a hipótese da normalidade não por ser a distribuição que melhor descreva esses resultados, mas por ser o modelo mais conveniente para análises estatísticas.

O critério mais usado para resolver se uma distribuição segue a normal é o critério do χ^2. Suponhamos n resultados $x_i (i = 1,2, \ldots n)$ ordenados de modo não decrescente isto é: $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n$. O conjunto de x_i pode ser distribuído em m intervalos convenientes $t_1, t_2, \ldots t_m$. Em cada intervalo há ν_j observações. Como estamos interessados na distribuição normal, devemos calcular as probabilidades $p_1, p_2, \ldots p_m$ do número teórico $\nu_j' = np_j$ ocorrências em cada intervalo. Pode-se demonstrar que a soma ponderada dos quadrados dos desvios $(\nu_j - \nu_j')$ se distribui aproximadamente como um χ^2, isto é:

$$\frac{(\nu_1 - \nu_1')^2}{\nu_1'} + \frac{(\nu_2 - \nu_2')^2}{\nu_2'} + \ldots + \frac{(\nu_m - \nu_m')^2}{\nu_m'} \cong \chi^2 \qquad \text{(IV.20)}$$

O número de graus de liberdade do χ^2 é $f = m - k - 1$ onde k é o número de conecções impostas aos resultados para poder calcular as freqüências teóricas. Se na Tabela de probabilidades de valores de χ^2 acharmos que o valor de χ^2 experimental corresponde a uma probabilidade menor que o nível de significância tolerado, 0,01 ou 0,05 dependendo do rigor exigido, deve-se rejeitar a hipótese da normalidade.

EXERCÍCIO 20

Foi preparada uma chapa metálica de alumínio-cobalto (0,5% de Co) em condições experimentais bem definidas. Para verificar a homogeneidade da distribuição do cobalto na liga depois de pronta, foram colhidas amostras ao longo da chapa e submetidas à análise por ativação. Os 200 resultados obtidos deram como média 0,520 ± 0,050. Tomando os intervalos com valor de 0,2 s em redor da média isto é:

no 1º intervalo : valores de 0,51 —— 0,53
no 2º intervalo : valores de 0,50 —— 0,51 e 0,53 —— 0,54

no 3º intervalo : valores de 0,49 --- 0,50 e 0,54 --- 0,55

⋮

no 10º intervalo : valores de 0,42 --- 0,43 e 0,61 --- 0,62
no 11º intervalo : valores $<$ 0,42 e $>$ 0,62

foram obtidas as freqüências experimentais ν e teóricas ν' que constam da Tabela IV.1.

Tabela IV.1

Distribuição de Resultados de Co em Liga de Al-Co

Intervalos em função de | s |	Freqüências ν	Freqüências ν'	$\frac{(\nu' - \nu)^2}{\nu'}$
$\leqslant$ 0,2	28	31,72	0,4363
0,2 – 0,4	32	30,44	0,0800
0,4 – 0,6	25	28,12	0,3462
0,6 – 0,8	21	24,96	0,6283
0,8 – 1,0	23	21,28	0,1390
1,0 – 1,2	15	17,44	0,3414
1,2 – 1,4	16	13,736	0,3732
1,4 – 1,6	8	10,384	0,5473
1,6 – 1,8	12	7,548	2,6259
1,8 – 2,0	4	5,272	2,2039
$>$ 2,0	16	9,100	
total	200	200,000	7,7213

A escolha dos intervalos é arbitrária. Nas colunas 2 e 3 estão as freqüências experimentais e as freqüências teóricas de uma distribuição normal. Por exemplo, para o intervalo 0,8 a 1,0 s achamos, na Tabela dos valores da distribuição normal, o valor $p = 0{,}6826 - 0{,}5762 = 0{,}1064$ que multiplicado por 200 vai dar 21,28. No penúltimo intervalo considerado, foi obtido $\nu = 4$. Como uma condição é que a freqüência não pode ser menor que 5, os dois últimos intervalos foram reunidos num único. Na última coluna da Tabela IV.1 estão os valores correspondentes de χ^2. No cálculo dos graus de liberdade $f = m - k - 1$, $k = 2$, porque foi calculada a média e o desvio padrão dos resultados. Temos então $f = 10 - 2 - 1 = 7$. $P(\chi^2_7 \geqslant 7{,}7213) \simeq 0{,}4$. Há então grande evidência de que a distribuição do cobalto na liga é homogênea.

Quanto maior for o número de resultados, tanto mais possível se torna a observação de fatores que provocam erros sistemáticos, mesmo pequenos. Uma vez obtida a informação de que existe um erro sistemático, cabe ao pesquisador descobrir as causas e procurar eliminá-las. Todavia, a experiência mostra que, na presença de fatores que provocam efeitos pequenos que só podem ser detectados por um grande número de observações, é válido admitir que os resultados experimentais se aproximam da distribuição normal e aplicar os critérios que se adotam para essa distribuição.

IV.4.1 — Causas que Provocam Desvios da Distribuição Normal

É muito importante determinar em que condições se pode esperar uma distribuição normal para um conjunto de resultados experimentais.

Pelo "Teorema do Limite Central", se ν for suficientemente grande, a soma de ν variáveis independentes $x_1, x_2, \ldots x_\nu$, tem distribuição normal mesmo quando as variáveis aleatórias independentes x_i tem uma distribuição arbitrária, desde que entre as variáveis consideradas não exista nenhuma cujo valor se compare, em ordem de grandeza, com a soma das demais. Em outras palavras, a variância de cada x_i deve ter pequena influência na variância total.

Em trabalhos experimentais, muitos fatores independentes podem contribuir para os erros aleatórios e sistemáticos. No desenvolvimento de um método experimental, ao examinar o conjunto dos resultados obtidos, é preciso observar quais as causas que podem provocar um afastamento da distribuição normal, de modo que, quando o método for padronizado, não existam fatores de importância dominante.

Levando em consideração a experiência acumulada por trabalhos experimentais, pode-se esperar um afastamento da distribuição normal nas seguintes circunstâncias:

1) quando o método experimental é aplicado em condições tais em que aparecem fatores dominantes cujas variáveis não estão sujeitas à distribuição normal. Por exemplo, no exercício anterior, em que foi considerada uma liga de alumínio-cobalto, a temperatura, o tempo de fusão, etc. podem ser fatores dominantes para causar uma falta de homogeneidade de cobalto na liga e, consequentemente, provocar uma distribuição de resultados que se afasta da normal.

2) em análises químicas, a distribuição normal ocorre, quando se combinam, num só conjunto, análises de amostras em que a concentração do componente a ser determinado, varia no máximo de 3 a 4 vezes. Caso contrário, pode-se obter uma distribuição assimétrica, causada pela correlação entre a média e o desvio padrão que pode ser expressa por:

$$s_x = \cong b\bar{x} + a \qquad \text{(IV.21)}$$

Esta situação se observa particularmente na análise da variância, onde frequentemente é necessário combinar, num conjunto estatístico, resultados correspondentes a um intervalo muito grande de concentrações. Neste caso, é preciso achar uma função de transformação que permita igualar as variâncias.

Suponhamos m populações com distribuição normal cujos parâmetros são $(\mu_1\ \sigma_1)$, $(\mu_2\ \sigma_2)$, . . . $(\mu_m\ \sigma_m)$ e que entre o desvio padrão e a média exista uma relação que pode ser expressa por:

$$\sigma_x = f(\mu) \qquad \text{(IV.22)}$$

Se substituirmos a variável x por um função de transformação g(x), pela lei de propagação de erros vamos ter:

$$\sigma_{g(x)} = \sigma_x\, g'(x) \qquad \text{(IV.23)}$$

Como μ é uma função de x, a expressão (IV.22) pode ser escrita na forma $\sigma_x = f(x)$ e então temos:

$$\sigma_{g(x)} = f(x)\, g'(x) \qquad \text{(IV.24)}$$

Deve-se escolher uma função de transformação g(x) tal que $\sigma_{g(x)}$ seja a mesma para todas as populações, isto é:

$$f(x)\, g'(x) = C \qquad \text{(IV.25)}$$

A função g(x) procurada será então uma integral indefinida expressa por:

$$g(x) = C \int \frac{dx}{f(x)} \qquad \text{(IV.26)}$$

Conclui-se então que, se houver uma correspondência linear entre σ_x e μ conforme (IV.21) a transformação pode ser feita pela função

$$g(x) = \log\,(bx + a)$$

No caso mais simples em que $\sigma_x = b\mu$, a função de transformação é $g(x) = \log(x)$.

3) Em vários campos experimentais, podem ser obtidos conjuntos de medidas que pertencem a populações não homogêneas, formadas pela mistura de duas ou mais populações normais. Distribuições desse tipo podem ser encontradas, quando se estuda a reprodutibilidade de certas medidas durante um período de tempo longo, no qual podem surgir fatores que deslocam o centro de espalhamento (média). Quando se misturam, num só conjunto, resultados obtidos por laboratórios diferentes também podem aparecer distribuição mistas o que indica um descontrole no método de trabalho em alguns dos laboratórios considerados.

4) Na prática, obtemos valores discretos, porque dependem da sensibilidade dos aparelhos usados nas medidas e da aproximação feita nos resultados finais. Como a distribuição normal é contínua, esses dois fatores podem ter grande influência no tipo de distribuição obtida, principalmente quando as flutuações dos resultados são da mesma ordem de grandeza do seu próprio valor. Isto se observa, em química analítica, no caso de análises de traços, quando se encontram desvios de até 9 σ em cerca de 150 determinações. Por considerações matemáticas, nestes casos se obtém populações com distribuição que se aproxima da distribuição de Poisson e as análises são consideradas semi-quantitativas. No caso de análises qualitativas, a distribuição pode ser considerada binomial, porque existem duas possibilidades: presença ou ausência do elemento procurado.

IV.5 – Aplicações dos Testes a Resultados Experimentais

IV.5.1 – Comparação de Duas Médias pelo Critério t

Suponhamos dois conjuntos de resultados $x_1, x_2, \ldots x_n$ e $y_1, y_2, \ldots y_m$ obtidos para a mesma medida por dois métodos diferentes. Deseja-se saber se existe uma diferença estatisticamente significativa entre as respostas dos dois métodos (médias).

Vamos admitir que os dois conjuntos obedecem a distribuições normais com médias μ_x e μ_y e variâncias σ_x^2 e σ_y^2. As estimativas das médias e variâncias são respectivamente: $\overline{x}$, $\overline{y}$ e s_x^2 e s_y^2.

A finalidade do teste é verificar se as médias μ_x e μ_y são iguais e então devem ser formuladas as hipóteses

$$H_0 : \mu_x = \mu_y$$

$$H_1 : \mu_x \neq \mu_y$$

Em primeiro lugar, verifica-se se as variâncias s_x^2 e s_y^2 não diferem significativamente uma da outra, pelo teste F (bicaudal). Devem então ser considerados dois casos:

1º caso: $s_x^2 = s_y^2$

Uma vez verificada a igualdade das variâncias, calcula-se a variância poderada

$$\overline{s}^{\,2} = \frac{(n-1)\, s_x^2 + (m-1)\, s_y^2}{n+m-1} \qquad \text{(IV.27)}$$

e calcula-se o valor de t:

$$t = \frac{\overline{x}-\overline{y}}{\overline{s}} \sqrt{\frac{mn}{m+n}} \quad (f = m+n-2) \qquad \text{(IV.28)}$$

Se o valor de t achado for maior que o t tabelado no nível de significância desejado, rejeita-se a hipótese da igualdade das duas médias. O teste é bicaudal, porque, "a priori", não temos elementos para supor qual das duas médias é maior. No caso de m = n a expressão se simplifica e se torna

$$t = \frac{|\overline{x}-\overline{y}|}{\overline{s}} \sqrt{\frac{n}{2}}$$

EXERCÍCIO 21

Foram obtidos 10 resultados para a medida de uma grandeza, usando um método A e 8 resultados por um método B. As médias e respectivas variâncias foram:

$\overline{x}_A = 2{,}43$ $\qquad$ $s_A^2 = 0{,}0169$

$\overline{x}_B = 2{,}61$ $\qquad$ $s_B^2 = 0{,}0256$

Verificar, num nível de significância de 0,10 se as médias são iguais.

Teste para a igualdade das variâncias:

$$F = \frac{0{,}0256}{0{,}0169} = 1{,}515$$

$$F_{0,10(7,9)} = \cong 3{,}25 \text{ (teste bicaudal)}$$

Como $F_{exp} < F_{tab.}$, conclui-se que as variâncias são iguais.

Calcula-se a variância média:

$$\bar{s}^2 = \frac{0{,}0169 \times 9 + 0{,}0256 \times 7}{16} = 0{,}0207$$

$$\bar{s} = 0{,}1434$$

teste para a igualdade das médias:

$$t = \frac{|2{,}43 - 2{,}61|}{0{,}1439} \sqrt{\frac{10 \times 8}{10 + 8}} = 2{,}64$$

$$t_{0,10,16} = 1{,}75$$

Como t experimental é maior que o t tabelado conclui-se que, num nível de significância de 0,10, as médias são diferentes.

No caso das médias e variâncias serem iguais, as m+n amostras podem ser consideradas como pertencentes à mesma população com média:

$$\bar{z} = \frac{\Sigma x_i + \Sigma y_i}{m + n} \qquad \text{(IV.29)}$$

e variância:

$$s_a^2 = \frac{(n-1)\, s_x^2 + (m-1)\, s_y^2}{m + n - 1} \qquad \text{(IV.30)}$$

2º caso: $s_x^2 \neq s_y^2$

Se, ao aplicarmos o teste F, verificarmos que as variâncias não são estatisticamente iguais no nível de significância desejado, o critério t pode ser aplicado com aproximação. Vamos ter:

$$t = \frac{|\bar{x} - \bar{y}|\, \sqrt{n}}{\sqrt{s_x^2 + s_y^2}} \qquad \text{(IV.31)}$$

onde n é o número tótal de determinações.

O número de graus de liberdade f neste caso será:

$$f = \frac{n-1}{C^2 + (1-C)^2} \qquad \text{(IV.32)}$$

$$\text{onde: } C = \frac{s_x^2}{s_x^2 + s_y^2} \qquad \text{(IV.33)}$$

Fazendo $C^2 + (1 - C)^2 = a$, verifica-se que os valores possíveis de a são: 0,5 a 1. Nestas condições, o número de graus de liberdade f varia de $(n - 1)$ até $2(n - 1)$. Este teste só pode ser aplicado, quando o número de determinações é o mesmo para a obtenção de x e y, isto é, n/2.

EXERCÍCIO 22

Foram aplicados dois métodos para a determinação de tório numa solução e foram obtidos 8 resultados para cada método. Deseja-se saber se as médias obtidas são estatisticamente iguais, num nível de significância de 0,10.

Método A : $\bar{x} = 1,29$; $s_x = \pm 0,05$

Método B : $\bar{y} = 1,02$; $s_y = \pm 0,19$

Aplica-se o teste F para verificar a igualdade das variâncias:

$$F = \frac{(0,19)^2}{(0,05)^2} = 14,44$$

O teste é bicaudal, porque a priori não se sabe se uma variância é maior que a outra. O valor de F com $f_1 = f_2 = 7$ graus de liberdade, no nível de significância 0,10 (bicaudal) é $\sim 3,8$. Conclui-se que $s_x^2 \neq s_y^2$. Aplica-se o teste t conforme (IV.31) e temos:

$$t \cong \frac{(1,29 - 1,02)}{\sqrt{(0,05)^2 + (0,19)^2}} \sqrt{16} = 5,50$$

$$C = \frac{(0,05)^2}{(0,05)^2 + (0,19)^2} = 0,065$$

$$f = \frac{16 - 1}{(0,065)^2 + (1 - 0,065)^2} = 17,08$$

$$t_{0,10,17} = 1,74$$

$$t_{0,001,17} = 3,97$$

Conclui-se que, mesmo num nível de significância de 0,001 não se pode aceitar a igualdade das médias.

IV.5.2 – Comparação de Métodos pelas Diferenças entre Resultados

Esta técnica é usada para comparar dois métodos experimentais ou o trabalho de dois técnicos que aplicam o mesmo método. Usa-se quando se quer comparar resultados de medidas diferentes, mas do mesmo tipo. Aplica-se mais frequentemente, pela própria natureza do problema, em química analítica.

Os resultados devem ser dispostos conforme a Tabela IV.2.

Tabela IV.2

Disposição dos Resultados de 2 Métodos para Comparação de Resultados

Amostras	Resultados		Diferença
	Método A	Método B	
1	x_1	y_1	$d_1 = x_1 - y_1$
2	x_2	y_2	$d_2 = x_2 - y_2$
3	x_3	y_3	$d_3 = x_3 - y_3$
.	.	.	
.	.	.	
.	.	.	
n	x_n	y_n	$d_n = x_n - y_n$

Se os dois métodos forem equivalentes, é óbvio que, para um número grande de resultados $\sum_1^n d_i = 0$, isto é, $\bar{d} = 0$, então as hipóteses são:

$$H_0 \; : \; \bar{d} = 0$$

$$H_1 \; : \; \bar{d} \neq 0$$

É portanto um teste bicaudal e calcula-se a variância das diferenças pela fórmula

$$s_d^2 = \frac{\Sigma d_i^2 - (\Sigma d_i)^2/n}{n-1} \qquad (IV.34)$$

e calcula-se t_{exp} pela expressão:

$$t = \frac{|\bar{d} - 0|\sqrt{n}}{s_d} = \frac{|\bar{d}|\sqrt{n}}{s_d} \qquad \text{(IV.35)}$$

com $(n - 1)$ graus de liberdade. Deste modo podem ser comparados dois métodos, independentemente das variâncias serem iguais ou não. É uma vantagem, no caso das variâncias serem diferentes, porque não é necessário aplicar o teste t, aproximado. Por outro lado é uma desvantagem, porque não permite uma comparação entre as variâncias dos dois métodos.

EXERCÍCIO 23

Foram feitas determinações de ouro em ligas de alumínio-ouro com várias concentrações deste metal por dois métodos diferentes. Os resultados obtidos estão na Tabela IV.3.

Tabela IV.3

Resultados de Determinações de Ouro por 2 Métodos

Amostra	Método I	Método II	Diferenças
1	0,08	0,13	-0,05
2	0,13	0,10	+0,03
3	0,16	0,14	+0,02
4	0,22	0,17	+0,05
5	0,27	0,31	-0,04
6	0,34	0,29	+0,05
7	0,30	0,32	-0,02
8	0,46	0,40	+0,06
Total	1,96	1,86	+0,10

$$\bar{d} = \frac{0,10}{8} = 0,0125$$

$$s_d^2 = 0,001878$$

$$s_d = 0,043$$

$$t = \frac{0,00125\sqrt{8}}{0,043} = 0,82$$

O teste é bicaudal e temos:

$$t_{0,05,7} = 2,37$$

$$t_{0,10,7} = 1,90$$

Então aceita-se a igualdade das médias mesmo num nível de confiança de 0,90 isto é:

$$P(|t| \geqslant 0,82) > 0,10$$

Neste exemplo se faz unicamente a comparação das duas médias. O teste não fornece nenhuma informação sobre as divergências dos resultados, isto é, não se pode saber se são causadas por desvio padrão alto de um ou de ambos os conjuntos. Este problema só pode ser resolvido pela análise da variância.

IV.5.3 – Comparação de Mais de Duas Variâncias

Suponhamos que temos k métodos para medir uma grandeza e queremos uma comparação entre esses métodos quanto à sua precisão. Existem dois casos a considerar que dependem do número de graus de liberdade com que as variâncias foram determinadas.

1º Caso

Quando as variâncias forem calculadas com números de graus de liberdade diferentes, usa-se o critério de Bartlett. Sejam s_1^2, $s_2^2 \ldots s_k^2$ as variâncias obtidas por k métodos, respectivamente com f_1, $f_2 \ldots f_k$ graus de liberdade. A variância da média ponderada é definida por:

$$\bar{s}^2 = \frac{\sum_1^k f_i s_i^2}{f} \qquad (IV.36)$$

onde $f = \sum_1^n f_i$

Calcula-se então o valor de B pela fórmula:

$$B = \frac{2,303}{C} (f \log \bar{s}^2 - \sum_1^k f_i \log s_i^2) \qquad (IV.37)$$

onde

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_1^k \frac{1}{f_i} - \frac{1}{f}\right) \qquad (IV.38)$$

Barlett demonstrou que a distribuição de B é aproximadamente a de um χ^2 com $k - 1$ graus de liberdade, desde que todos os f_i sejam maiores que 2. Se $B > \chi^2_{(k-1)}$, no nível de significância escolhido, rejeita-se a hipótese da homogeneidade das variâncias. Para este teste, as hipóteses são

$$H_0 : s_1^2 = s_2^2 = \ldots = s_k^2$$

H_1 : pelo menos um dos s_i^2 é diferente dos demais.

EXERCÍCIO 24

Temos 4 tipos de misturadores que devem ser testados para saber qual o mais eficiente. No teste é usada uma substância inerte e diuranato de amônio em proporções conhecidas. O tempo T usado para a mistura foi o mesmo em cada aparelho. De cada mistura foi depois tirada uma amostra e feitas várias análises de urânio. Foram obtidos os resultados da Tabela IV.4.

Tabela IV.4

Variâncias de Análises de U e Cálculos para Aplicação do Teste de Bartlett

Aparelho	s_i	f_i	s_i^2	$f_i s_i^2$	$\log s_i^2$	$1/f_i$
A_1	0,012	6	0,000144	0,000864	-3,8416	0,1617
A_2	0,019	7	0,000361	0,002527	-3,4425	0,1429
A_3	0,023	9	0,000529	0,004761	-3,2765	0,1111
A_4	0,015	5	0,000225	0,001125	-3,6478	0,2000
		27		0,009277	-14,2085	0,6206

$$s^2 = \frac{0{,}009277}{27} = 0{,}000344$$

$$C = 1 + \frac{1}{3(4-1)}\left(0{,}6206 - \frac{1}{27}\right)$$

$$C = 1 + \frac{1}{9}(0{,}5836)$$

$$C = 1{,}06484$$

$$B = \frac{2{,}303}{1{,}06484}\left(27 \log 0{,}000344 - (-94{,}8746)\right)$$

$$B = \frac{2{,}303}{1{,}06484}(1{,}3617)$$

$$B = 2{,}94$$

$$\chi^2_{0{,}05(3)} = 7{,}8$$

Como B calculado é menor que $\chi^2_{0,05,3}$ conclui-se pela igualdade das variâncias no nível de significância 0,05. Isto quer dizer que a eficiência dos aparelhos no tempo T é a mesma.

2º Caso

Quando o número de graus de liberdade para todas as variâncias a serem comparadas for o mesmo, em vez de usar o critério aproximado de Bartlett, adota-se o critério de Cochrane. Baseia-se na lei de distribuição da relação entre o valor da maior variância achada com a soma de todas as variâncias. As hipóteses H_0 e H_1 são as mesmas do caso anterior

$$G_{max} = \frac{s^2_{max}}{s_1^2 + s_2^2 + \ldots + s_k^2} \tag{IV.39}$$

Existem tabelas de G_{max} e os valores são encontrados para k, (n − 1) graus de liberdade onde k é o número de variâncias a serem comparadas e n − 1 é o número de graus de liberdade de cada variância. Se o valor de G_{max} calculado for maior que o de G_{max} tabelado, no nível de significância escolhido, rejeita-se a hipótese da igualdade das variâncias.

EXERCÍCIO 25

Deseja-se comparar a precisão de 6 métodos para medir uma grandeza. Foram feitas 10 medidas com cada método e foram obtidos os seguintes valores para os desvios padrões.

$$s_1 = 0{,}23 \;;\; s_2 = 0{,}15 \;;\; s_3 = 0{,}12 \;;\; s_4 = 0{,}28 \;;\; s_5 = 0{,}31 \;;\; s_6 = 0{,}25$$

$$G = \frac{(0{,}31)^2}{(0{,}23)^2 + (0{,}15)^2 + (0{,}12)^2 + (0{,}28)^2 + (0{,}31)^2 + (0{,}25)^2}$$

$$G = 0{,}294$$

Na Tabela, procura-se o valor de G_{max} para $k = 6$ e $n - 1 = 9$. Para o nível de significância 0,05, o valor tabelado é 0,368.

Como G calculado é menor que o G_{max} tabelado, aceita-se a igualdade das variâncias.

IV.5.4 — Homogeneidade de Resultados

Existem critérios para reconhecer resultados afetados de erros acidentais que escaparam à observação do pesquisador, durante a execução do trabalho. Os critérios estatísticos para a estimativa desses erros baseiam-se na distribuição normal que admite a ocorrência desses erros, mesmo que a probabilidade seja muito pequena.

1º Método

Sejam os resultados x_1, x_2, . . . x_n entre os quais suspeitamos de um resultado x_i. Calcula-se x e s_x com os n resultados e aplica-se o critério r a x_i de acordo com a fórmula:

$$r_i = \frac{|x_i - \bar{x}|}{s_x \sqrt{\frac{n-1}{n}}}$$

Os valores r_i estão sujeitos à distribuição r com $f = n - 2$ graus de liberdade, porque existem duas relações entre os r_i que são:

$$\Sigma r_1 = 0$$

$$\Sigma r_i^2 = n$$

Se o valor de r calculado para x_i for menor que o de r tabelado no nível de significância pré-estabelecido, aceita-se x_i como pertencente à população. Para valores grandes de f, a distribuição r se aproxima muito da normal. Por esse motivo, um valor r_i maior que o r tabelado, quando n é grande, não é suficiente para aceitar ou rejeitar x_i. O *critério r só é válido para a medida* x_i considerada.

2º Método

Aplica-se aos valores máximos e mínimo de uma população. Deve-se *considerar:*

$$r_{max} = \frac{x_{max} - \bar{x}}{s_x \sqrt{\frac{n-1}{n}}} \quad \text{e} \quad r_{min} = \frac{|x_{min} - \bar{x}|}{s_x \sqrt{\frac{n-1}{n}}} \qquad \text{(IV.40)}$$

A tabela de r_{max} *ou* r_{min} *só tem* valores para $n = 25$, isto é, para um número de graus de liberdade $f = n - 2 = 23$.

Até esse valor de f, se r_{max} ou r_{min} calculados forem maiores que os tabelados, deve-se rejeitar o resultado. Para $n > 25$, adota-se o critério de $3\,\sigma$ para verificar se os resultados são homogêneos.

EXERCÍCIO 26

Temos 15 resultados de uma medida. A média achada foi $\bar{x} = 1{,}45$ e $s_x = 0{,}07$. Suspeita-se de um resultado cujo valor é 1,28 e deseja-se saber se deve ser rejeitado num nível de significância de 0,05

$$r_{min} = \frac{|1{,}28 - 1{,}45|}{0{,}07\sqrt{\frac{14}{15}}} = \frac{0{,}17}{0{,}0676} = 2{,}51$$

Na Tabela, para 13 graus de liberdade e nível de significância 0,05, acha-se o valor $r_{min} = 2{,}493$. O resultado 1,28 deve ser rejeitado, porque o valor de r_{min} experimental é maior que o tabelado.

Observação

Pode acontecer que dois resultados sejam suspeitos por serem muito altos ou muito baixos. Suponhamos que temos dois resultados aparentemente altos. Rejeita-se arbritariamente o mais alto, calcula-se a média e o desvio padrão com os n − 1 resultados restantes e testa-se o valor em questão. Se for rejeitado, com maior razão deve-se rejeitar o primeiro. Se for aceito, testa-se o que tinha sido rejeitado arbitrariamente, neste caso com os n resultados. A Tabela tem valores para n = 3 até n = 25, porque para n maior que 25, a aceitação ou rejeição de um resultado não altera apreciavelmente a média e o desvio padrão.

IV.5.5 – Intervalo de Confiança para um Resultado x_{n+1}

Se temos n resultados x_i e queremos saber dentro de que intervalo deve-se esperar um resultado x_{n+1}, calcula-se $\bar{x}$ e s_x dos n resultados. Um resultado x_{n+1} deve estar num intervalo definido por:

$$x_{n+1} = \bar{x} \pm t\, s_x \sqrt{\frac{n+1}{n}} \qquad \text{(IV.41)}$$

onde o t tem (n − 1) graus de liberdade e nível de confiança desejado.

EXERCÍCIO 27

Uma série de 10 resultados deu como média 7,70 e o desvio padrão foi 0,32. Dentro de que intervalo deve-se esperar um resultado adicional?

$$x_{11} = 7{,}70 \pm 2{,}26 \times 0{,}32 \sqrt{\frac{11}{10}}$$

$$x_{11} = 7{,}70 \pm 0{,}76$$

IV.6 – Desigualdade de Tschebyscheff

Às vezes é difícil estabelecer qual a lei de distribuição a que obedecem os resultados experimentais que formam um conjunto. Tschebyscheff demonstrou que o desvio padrão σ é uma medida de dispersão da variável aleatória, independentemente do tipo da distribuição.

A fórmula da variância de uma variável aleatória contínua pode ser representada da seguinte maneira:

$$\sigma^2 = \int_{-\infty}^{\mu-k} (x-\mu)^2 f(x)\, dx + \int_{\mu-k}^{\mu+k} (x-\mu)^2 f(x)\, dx + \int_{\mu+k}^{\infty} (x-\mu)^2 f(x)\, dx \qquad \text{(IV.42)}$$

conforme a Figura 4.5.

Se não considerarmos a segunda integral vamos ter:

$$\sigma^2 \geqslant \int_{-\infty}^{\mu-k} (x-\mu)^2 f(x)\, dx + \int_{\mu+k}^{\infty} (x-\mu)^2 f(x)\, dx \qquad \text{(IV.43)}$$

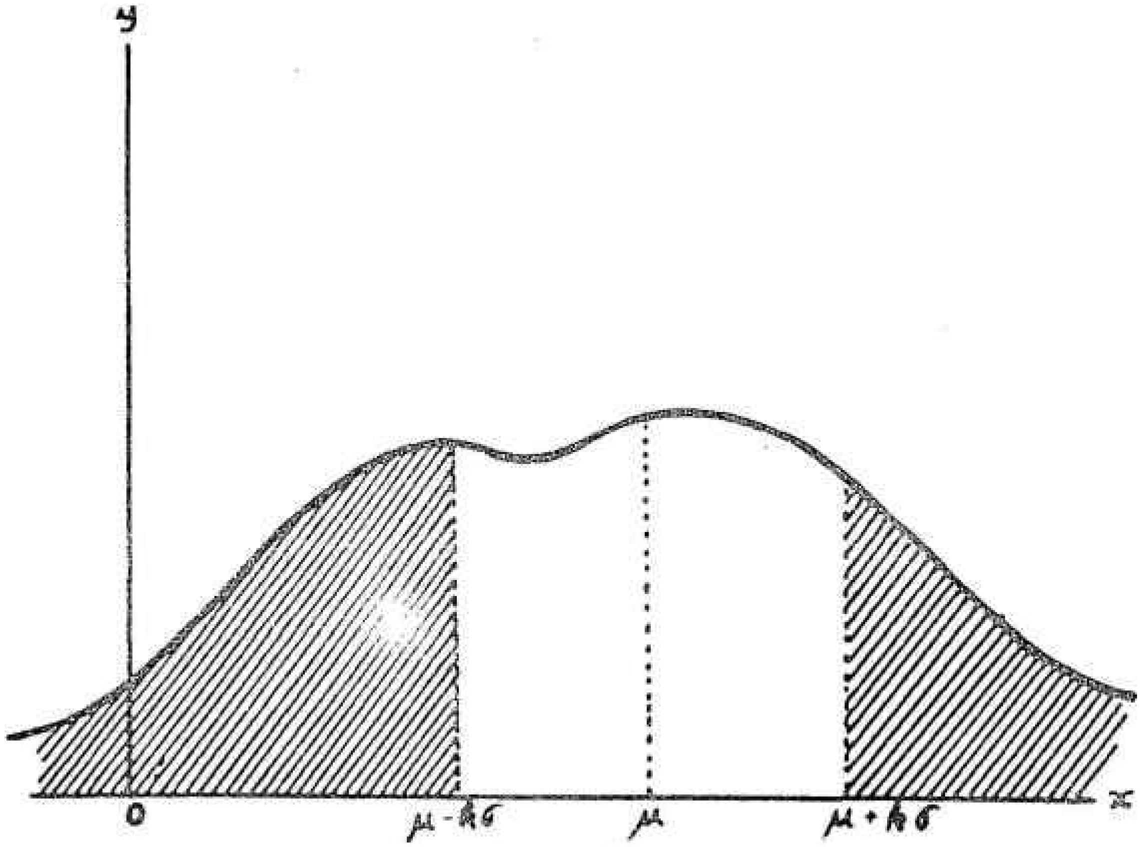

Figura 4.5 — Distribuição contínua qualquer.

Quando $x = \mu \pm k\,\sigma$, o valor de $(x-\mu)^2$ será:

$$(x-\mu)^2 = k^2\,\sigma^2 \qquad \text{(IV.44)}$$

Como os intervalos são $-\infty - \mu - k\sigma$ e $\mu + k\sigma - \infty$, temos $(x-\mu)^2 \geqslant k^2\,\sigma^2$

Então:

$$\int_{-\infty}^{\mu-k\sigma} (x-\mu)^2\, f(x)\, dx \geqslant k^2\,\sigma^2 \int_{-\infty}^{\mu-k\sigma} f(x)\, dx \qquad \text{(IV.45)}$$

O mesmo resultado é válido para a segunda integral de (IV.43). Assim temos:

$$\sigma^2 \geqslant k^2\,\sigma^2 \left[\int_{-\infty}^{\mu-k\sigma} f(x)\, dx + \int_{\mu+k\sigma}^{\infty} f(x)\, dx \right. \qquad \text{(IV.46)}$$

ou

$$\frac{1}{k^2} \geqslant \int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \qquad \text{(IV.47)}$$

O segundo termo da desigualdade (IV.47) é a área sob a curva f(x) nos intervalos $-\infty - \mu - k\sigma$ e $\mu + k\sigma - \infty$. A probabilidade de se assumir um valor fora do intervalo $\mu - k\sigma - \mu + k\sigma$ será então:

$$P(\,|x-\mu| \geqslant k\sigma\,) \leqslant \frac{1}{k^2}$$

e (IV.48)

$$P(\,|x-\mu| < k\sigma\,) > 1 - \frac{1}{k^2}$$

Isto prova que σ mede a dispersão da variável aleatória qualquer que seja o tipo da distribuição desde que tenha média e variância.

Por exemplo, para k = 2

$$P(\,|x-\mu| \geqslant 2\sigma\,) \leqslant 0{,}25$$

No caso da distribuição normal, esta probabilidade é 0,05.

Para a média de uma variável aleatória, temos:

$$P(\,|x-\mu| > \frac{k\sigma}{\sqrt{n}}\,) < \frac{1}{k^2} \qquad \text{(IV.49)}$$

Fazendo:

$$\frac{k\sigma}{\sqrt{n}} = d$$

temos:

$$\frac{1}{k^2} = \frac{\sigma^2}{nd^2} \qquad \text{(IV.50)}$$

EXERCÍCIO 28

Achar o número de resultados necessários para que a diferença entre a média $\overline{x}$ da população e a média verdadeira μ não seja maior que 0,5 σ com uma probabilidade de 0,95. Temos que: d = 0,5. Então:

$$\frac{1}{k^2} = \frac{\sigma^2}{(0{,}5\sigma)^2\, n} = \frac{1}{0{,}25\, n}$$

Pelas condições impostas deve-se ter:

$$\frac{1}{0{,}25\, n} = 0{,}05$$

$$n = \frac{1}{0{,}25 \times 0{,}05} = 80$$

No caso da distribuição normal, acha-se n = 16.

IV.7 – Média Estatisticamente Ponderada

Quando temos séries de medidas de uma mesma amostra, obtidas por métodos ou operadores diferentes, os valores de $\bar{x}$ e $s_{\bar{x}}$ de cada conjunto podem ser combinados para se obter um valor $\bar{\bar{x}}$ mais representativo de μ.

Suponhamos então que temos n médias $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_n$ com as respectivas variâncias $s^2_{\bar{x}_1}, s^2_{\bar{x}_2}, \ldots s^2_{\bar{x}_n}$

Se os valores de $\bar{x}$ seguem a distribuição normal, a expressão:

$$P_i = \frac{1}{s_{\bar{x}_i}\sqrt{2\pi}}\, e^{-\frac{(\bar{x}_i - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_i}}} \qquad \text{(IV.51)}$$

dá a possibilidade de se obter uma média $\bar{x}_i$, quando o valor mais provável é $\bar{\bar{x}}$. Os valores $\bar{x}_i$ são independentes e então a probabilidade de se encontrar os valores $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_n$ é o produto das probabilidades

$$P = \frac{1}{s_{\bar{x}_1}\sqrt{2\pi}}\, e^{-\frac{(\bar{x}_1 - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_1}}} \cdot \frac{1}{s_{\bar{x}_2}\sqrt{2\pi}}\, e^{-\frac{(\bar{x}_2 - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_2}}} \cdots \frac{1}{s_{\bar{x}_n}\sqrt{2\pi}}\, e^{-\frac{(\bar{x}_n - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_n}}} \qquad \text{(IV.52)}$$

Para que a probabilidade P possa ser máxima, a soma dos expoentes deve ser mínima ou:

$$Q = \frac{(\bar{x}_1 - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_1}} + \frac{(\bar{x}_2 - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_2}} + \ldots + \frac{(\bar{x}_n - \bar{\bar{x}})^2}{2 s^2_{\bar{x}_n}} \qquad \text{(IV.53)}$$

deve preencher as condições:

$$\frac{dQ}{d\bar{\bar{x}}} = 0 \quad e \quad \frac{d^2Q}{d\bar{\bar{x}}^2} > 0$$

$$\frac{dQ}{d\bar{\bar{x}}} = -\frac{(\bar{x}_1 - \bar{\bar{x}})}{s^2_{\bar{x}_1}} - \frac{(\bar{x}_2 - \bar{\bar{x}})}{s^2_{\bar{x}_2}} - \ldots - \frac{(\bar{x}_n - \bar{\bar{x}})}{s^2_{\bar{x}_n}} \qquad \text{(IV.54)}$$

Para que (IV.54) seja igual a zero devemos ter:

$$\bar{\bar{x}}\left(\frac{1}{s^2_{\bar{x}_1}} + \frac{1}{s^2_{\bar{x}_2}} + \ldots + \frac{1}{s^2_{\bar{x}_n}}\right) = \frac{\bar{x}_1}{s^2_{\bar{x}_1}} + \frac{\bar{x}_2}{s^2_{\bar{x}_2}} + \ldots + \frac{\bar{x}_n}{s^2_{\bar{x}_n}}$$

ou

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{n} \frac{\bar{x}_i}{s^2_{\bar{x}_i}}}{\sum_{i=1}^{n} \frac{1}{s^2_{\bar{x}_i}}} \qquad \text{(IV.55)}$$

A condição $\frac{d^2Q}{d\bar{\bar{x}}^2} > 0$ é óbvia, se derivarmos a expressão (IV.54). A expressão (IV.55) é semelhante a uma média ponderada e o valor mais provável de $\bar{\bar{x}}$ é a média ponderada dos $\bar{x}_i$ cujos pesos são os inversos das respectivas variâncias.

IV.7.1 — Desvio Padrão da Média Ponderada

Fazendo:

$$d_i = \bar{x}_i - \bar{\bar{x}}$$

a expressão

$$s_{\bar{\bar{x}}} = \sqrt{\frac{\Sigma\, d_i^2 / s^2_{\bar{x}_i}}{\Sigma\, 1/s^2_{\bar{x}_i}}} \qquad \text{(IV.56)}$$

é o desvio padrão experimental da média ponderada. Diferenciando a expressão (IV.55) chega-se a uma fórmula semelhante a saber:

$$s^2_{\bar{\bar{x}}} = \left(\frac{\partial\bar{\bar{x}}}{\partial\bar{x}_1}\right)^2 s^2_{\bar{x}_1} + \left(\frac{\partial\bar{\bar{x}}}{\partial\bar{x}_2}\right)^2 s^2_{\bar{x}_2} + \ldots + \left(\frac{\partial\bar{\bar{x}}}{\partial\bar{x}_n}\right)^2 s^2_{\bar{x}_n}$$

$$= \left(\frac{1/s^2_{\bar{x}_1}}{\Sigma\, 1/s^2_{\bar{x}_i}}\right)^2 s^2_{\bar{x}_1} + \left(\frac{1/s^2_{\bar{x}_2}}{\Sigma\, 1/s^2_{\bar{x}_i}}\right) s^2_{\bar{x}_2} + \ldots + \left(\frac{1/s^2_{\bar{x}_n}}{\Sigma\, 1/s^2_{\bar{x}_i}}\right)^2 s^2_{\bar{x}_n}$$

$$= \frac{\Sigma\, 1/s^2_{\bar{x}_i}}{(\Sigma\, 1/s^2_{\bar{x}_i})^2} = \frac{1}{\Sigma\, 1/s^2_{\bar{x}_i}}$$

$$s_{\bar{\bar{x}}} = \frac{1}{\sqrt{\Sigma\, 1/s^2_{\bar{x}_i}}} \qquad \text{(IV.57)}$$

Este é o desvio padrão teórico da média ponderada. Se as precisões dos vários $\bar{x}_i$ forem todas da mesma ordem de grandeza e não houver erro sistemático, a relação entre $s_{\bar{\bar{x}}}$ experimental e $s_{\bar{\bar{x}}}$ teórico deve ser ~ 1.

Na prática, pode-se esperar valores entre 0,5 e 2. Se a relação for muito maior que 2, o conjunto das médias está afetado de um erro sistemático. Neste caso, deve-se examinar os valores $\bar{x}_i$ e rejeitar aqueles que forem discordantes. Se após a rejeição, o novo cálculo levar a um quociente cujo valor estiver no intervalo 0,5 – 2, pode-se aceitar a média ponderada como o valor mais provável.

EXERCÍCIO 29

Foram usados seis métodos na medida de uma grandeza e foram obtidos os resultados da Tabela (IV.5).

Tabela IV.5

Resultados de uma Medida por 6 Métodos Diferentes

Método	Média (x_i)	Desvio Padrão $s_{\bar{x}_i}$
1	30,1	1,2
2	29,7	0,9
3	30,3	1,1
4	28,9	1,0
5	29,5	0,8
6	29,4	0,6

$$\bar{\bar{x}} = \sqrt{\frac{\left(\frac{30,1}{1,2}\right)^2 + \left(\frac{29,7}{0,9}\right)^2 + \left(\frac{30,3}{1,1}\right)^2 + \left(\frac{28,9}{1,0}\right)^2 + \left(\frac{29,5}{0,8}\right)^2 + \left(\frac{29,4}{0,6}\right)^2}{\left(\frac{1}{1,2}\right)^2 + \left(\frac{1}{0,9}\right)^2 + \left(\frac{1}{1,1}\right)^2 + \left(\frac{1}{1,0}\right)^2 + \left(\frac{1}{0,8}\right)^2 + \left(\frac{1}{0,6}\right)^2}}$$

$$\bar{\bar{x}} = 29,56$$

$$s_{\bar{\bar{x}}_{teórico}} = \frac{1}{\sqrt{\left(\frac{1}{1,2}\right)^2 + \left(\frac{1}{0,9}\right)^2 + \left(\frac{1}{1,1}\right)^2 + \left(\frac{1}{1,0}\right)^2 + \left(\frac{1}{0,8}\right)^2 + \left(\frac{1}{0,6}\right)^2}} = 0,3515$$

$$s_{\bar{\bar{x}}_{exp}} = \sqrt{\frac{\left(\frac{0,54}{1,2}\right)^2 + \left(\frac{0,14}{0,9}\right)^2 + \left(\frac{0,74}{1,1}\right)^2 + \left(\frac{0,66}{1,0}\right)^2 + \left(\frac{0,06}{0,8}\right)^2 + \left(\frac{0,18}{0,6}\right)^2}{\left(\frac{1}{1,2}\right)^2 + \left(\frac{1}{0,9}\right)^2 + \left(\frac{1}{1,1}\right)^2 + \left(\frac{1}{1,0}\right)^2 + \left(\frac{1}{0,8}\right)^2 + \left(\frac{1}{0,6}\right)^2}} = \frac{1,091602}{2,844950}$$

$$s_{\bar{\bar{x}}_{exp}} = 0,3837$$

$$\frac{s_{\bar{\bar{x}}_{exp}}}{s_{\bar{\bar{x}}_{teórico}}} = 1,09$$

Aceita-se então $\bar{\bar{x}} = 29,56$ como o valor mais representativo da média.

CAPÍTULO V

DISTRIBUIÇÃO DE POISSON

Muitos resultados experimentais são obtidos por meio de medidas de quantidades discretas. É o caso, por exemplo, da radioatividade em análises radioquímicas, medidas de raios x, espectroscopia ótica, etc.

Os resultados de análises semi-quantitativas são sempre seqüências de valores discretos independentemente do método usado. Em todos esses casos, os resultados seguem a distribuição de Poisson.

Uma variável aleatória <u>x</u> (valores discretos: x = 0, 1, 2, 3 . . .) segue a distribuição de Poisson se a probabilidade de dar o valor x é definida por:

$$P_{\mu}(x) = \frac{\mu^x e^{-\mu}}{x!} \qquad (V.1)$$

e

$$\sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!} = 1 \qquad (V.2)$$

para qualquer valor de μ.

A distribuição de Poisson só tem um parâmetro que é a média, porque $\mu = \sigma^2$ e se caracteriza pelo seguinte: se num intervalo T, (unidade de medida) a variável aleatória assume um valor N, num intervalo t, a variável aleatória vai assumir um valor n proporcional a t. Então a média de uma distribuição de Poisson pode ser representada por $\mu = kt$ onde

$$k = \frac{N}{T} \qquad (V.3)$$

isto é, k é o valor da variável aleatória por unidade de medida. A lei de Poisson pode ser então representada por:

$$P_{\mu}(x) = \frac{(kt)^x \ e^{-kt}}{x!} \qquad (V.4)$$

onde t pode representar intervalos de tempo iguais em que são feitas as medidas, seqüências de intervalos de igual comprimento, área ou volume em que a variável aleatória ocorre, etc.

V.1 — Aproximação da Distribuição de Poisson da Distribuição Normal

Para fins práticos, a distribuição de Poisson segue razoavelmente bem a distribuição normal a partir de $\mu = 9$, como se pode ver pela Figura V.1. Para $x = 0$ temos:

$$P_9(0) = \frac{9^0 \ e^{-9}}{0!} = 0{,}000123$$

Os outros valores de P(x) estão na Tabela V.1.

A possibilidade de substituir a distribuição de Poisson pela Normal facilita muito o cálculo estatístico.

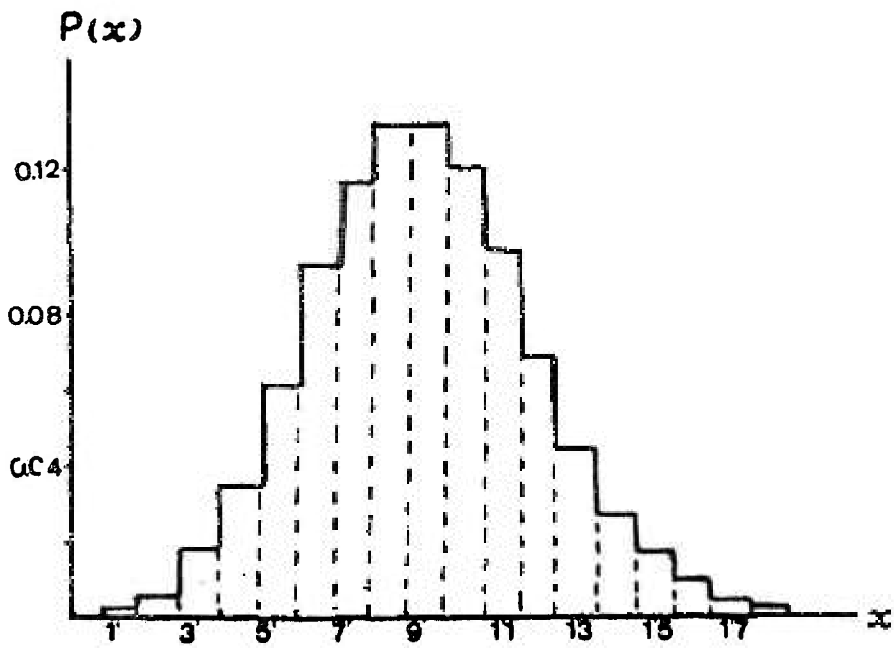

Figura 5.1 — Distribuição de Poisson (μ = 9).

Tabela V.1

Valores de P(x) quando $\mu = 9$

P (1) = 0,001111	P(12) = 0,072765
P (2) = 0,004998	P(13) = 0,050376
P (3) = 0,014994	P(14) = 0,032384
P (4) = 0,033737	P(15) = 0,019431
P (5) = 0,060727	P(16) = 0,010930
P (6) = 0,091090	P(17) = 0,005786
P (7) = 0,117116	P(18) = 0,002893
P (8) = 0,131755	P(19) = 0,001370
P (9) = 0,131756	P(20) = 0,000617
P(10) = 0,118580	P(21) = 0,000264
P(11) = 0,097020	P(22) = 0,000108

EXERCÍCIO 30

Achar a probabilidade de x cair no intervalo limitado por $x_1 = 8$ e $x_2 = 22$, quando a média é $\mu = 16$.

Como a distribuição não é contínua, isto é, temos áreas limitadas por retângulos, podemos fazer uma aproximação com a curva introduzindo uma correção que consiste em substituir os limites x_1 e x_2 pelos limites de integração $x_1 - 1/2$ e $x_2 + 1/2$

$$P(x_1 \leqslant x \leqslant x_2) = \sum_{x=x_1}^{x=x_2} P_\mu(x) \cong \frac{1}{\sqrt{2\pi\mu}} \int_{x_1-1/2}^{x_2+1/2} e^{-(x-\mu)^2} dx$$

$$= \varphi \frac{(x_2 + 1/2 - \mu)}{\mu} - \phi \frac{(x_1 - 1/2 - \mu)}{\mu}$$

Para $x_1 = 8$ e $x_2 = 22$ com $\mu = 16$ temos:

$$u_2 = \frac{22 + 1/2 - 16}{16} = \frac{6,5}{4} = 1,625$$

$$u_1 = \frac{8 - 1/2 - 16}{16} = \frac{8,5}{4} = 2,125$$

Pela distribuição normal acha-se que para:

$u_2 = 1,625$ —— $P(x \leqslant 22) = 0,9479$

$u_1 = 2{,}125$ —— $P(x < 8) = 0{,}0168$

$P(8 \leqslant x \leqslant 22) = 0{,}9479 - 0{,}0168 = \mathbf{0{,}9311}$

A partir da fórmula de Poisson temos:

$P(x \leqslant 22) = P_{16}(0) + P_{16}(1) + \ldots + P_{16}(22) = 0{,}9454$

$P(x < 8) = P_{16}(0) + P_{16}(1) + \ldots + P_{16}(7) = 0{,}0218$

$P(8 \leqslant x \leqslant 22) = 0{,}9454 - 0{,}0218 = 0{,}9236$

o que prova que a aproximação pela Normal é satisfatória.

Para certos tipos de cálculos, como em particular a comparação de várias médias pela análise de variância, o desvio padrão não pode ser uma função da média, como já foi visto no capítulo anterior. No caso da distribuição de Poisson, temos a dependência: $\sigma_x = \sqrt{x}$. Neste caso, deve-se usar uma função de transformação

$$y = g(x) = c \int \frac{dx}{f(x)} \qquad \text{(V.5)}$$

Como $\sigma_x = f(x) = \sqrt{x}$, temos que:

$$g(x) = c \int \frac{dx}{\sqrt{x}} \qquad \text{(V.6)}$$

$$y = g(x) = 2\sqrt{x} \quad \text{(a menos da constante C)} \qquad \text{(V.7)}$$

Se a variável aleatória x se distribui normalmente com média μ e variância $\sigma^2 = \mu$, a nova variável $y = 2\sqrt{x}$ também se distribui normalmente com média $2\sqrt{\mu}$ e variância $\sigma^2 = 1$, porque

$$\sigma_{g(x)} = f(x)\, g'(x) = \frac{\sqrt{x}}{\sqrt{x}} = 1 \qquad \text{(V.8)}$$

Com esta transformação, pode-se escrever uma outra equação aproximada para a ocorrência da variável x no intervalo $x_1 \leqslant x \leqslant x_2$

$$P(x_1 \leqslant x \leqslant x_2) = \phi(2\sqrt{x_2 + 1/2} - 2\sqrt{\mu}) - \phi(2\sqrt{x_1 - 1/2} - \sqrt{2\mu}) \qquad \text{(V.9)}$$

Substituindo x_1, x_2 e μ pelos valores numéricos do exercício temos:

$u_2 = 2(\sqrt{22+1/2} - \sqrt{16}) = 1{,}487$

$u_1 = 2(\sqrt{8-1/2} - \sqrt{16} = 2{,}523$

$P(x \leqslant 1{,}487) = 0{,}9312$

$P(x \leqslant -2{,}523) = 0{,}0058$

Aplicando a distribuição Normal à nova variável acha-se que:

$$P(8 \leqslant x \leqslant 22) = 0{,}9312 - 0{,}0058 = 0{,}9254$$

que também é uma boa aproximação de 0,9236 achado pela distribuição de Poisson.

V.2 — Limites de Confiança para a Média

A distribuição de Poisson só tem valores discretos, e, para uma distribuição discreta, existe a relação

$$P(x > x_a + 1) = 1 - P(x \leqslant x_a)$$

conforme a representação gráfica da Figura 5.2.

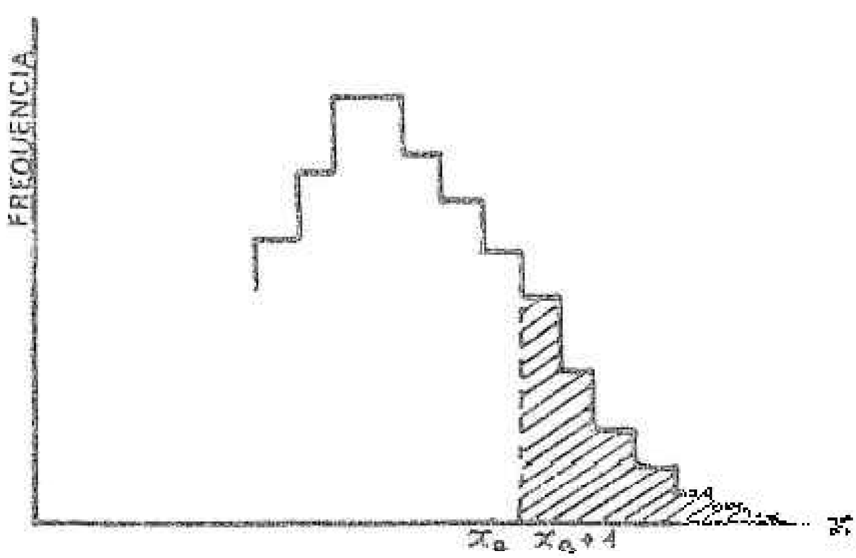

Figura 5.2 — Representação da $P(x \geqslant x_a + 1)$.

Então os limites de confiança x_1 e x_2 para a média μ podem ser obtidos pelas seguintes expressões.

$$u_{p/2} = 2\,(\sqrt{x'+1} - \sqrt{kx_2})$$

$$u_{1-p/2} = 2\sqrt{x'} - \sqrt{kx_1}) \qquad \text{(V.10)}$$

onde: $x' = k\bar{x}$

sendo k o número de unidades de medida considerado e $u_{p/2}$ e $u_{1-p/2}$ são os valores da distribuição normal.

EXERCÍCIO 31

Numa contagem de "background" foram obtidas 80 contagens em 20 minutos, isto é $\mu = 4$ cpm. Calcular os limites de confiança para a média num nível de confiança de 0,95.

$$(1) \text{ ——— } 2\,(\sqrt{80+1} - \sqrt{20\,x_2}) = 1{,}96$$

$$(2) \text{ ——— } 2\,(\sqrt{80} - \sqrt{20\,x_1}) = 1{,}96$$

De (1) : $-20x_2 = -9 - 0{,}98 = -9{,}98 \quad \therefore x_2 = 4{,}98$

De (2) : $-20x_1 = 0{,}98 - 8{,}94 = 7{,}96 \quad \therefore x_1 = 3{,}16$

Temos assim que, num nível de confiança de 0,95, o intervalo de confiança para μ é: 3,16 – 4,98. O intervalo não é simétrico.

EXERCÍCIO 32

Suponhamos agora que uma amostra radioativa deu 10000 contagens em 10 minutos, isto é, $\mu = 1000$ cpm. O intervalo de confiança para a média, num nível de confiança de 0,95 será:

$$2(\sqrt{10000+1} - \sqrt{10\,x_2}) = -1{,}96$$

$$2(\sqrt{10000} - \sqrt{10\,x_1}) = 1{,}96$$

Resolvendo essas equações para x_1 e x_2 obtemos o intervalo

980 – 1020. Neste caso existe simetria, porque a média é suficientemente grande.

V.3 – Hipótese da Igualdade de Duas Médias

Suponhamos duas médias: x_1 obtida por n_1 determinações e x_2 obtida por n_2 determinações e seja $x_1 < x_2$ numericamente.

Para comparar as duas médias, calcula-se a média geral $\overline{x}$ fazendo:

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2} \qquad (V.11)$$

Calculam-se os valores correspondentes da distribuição normal:

$$\mu_1 = 2(\sqrt{n_1 x_1 + 1} - \sqrt{n_1 \overline{x}})$$
$$\mu_2 = 2(\sqrt{n_2 x_2} - \sqrt{n_2 \overline{x}}) \qquad (V.12)$$

A diferença $u_2 - u_1$ se distribui aproximadamente como a diferença de duas variáveis que se distribuem segundo a Normal. Deve-se então considerar que as duas médias diferem entre si num nível de confiança 0,05 ou 0,01 se $u_2 - u_1$ for maior que:

$$1{,}96\sqrt{2} = 2{,}77 \quad \text{(nível 0,05)}$$

$$2{,}58\sqrt{2} = 3{,}64 \quad \text{(nível 0,01)}$$

EXERCÍCIO 33

Foram tiradas duas alíquotas de uma solução radioativa e colocadas em tubos de contagem. A primeira alíquota foi de 0,2 mililitros, a contagem foi de 50 minutos e deu um total de 5210 contagens. A segunda alíquota foi de 0,5 mililitros, a contagem foi de 20 minutos e deu 5142 contagens. Verificar se a pipeta de 0,2 mililitros está certa, sabendo-se que a de 0,5 mililitros foi aferida.

Neste caso, como os tempos de contagem e volumes usados se compensam, pode-se usar como média $\overline{x}$, a média das duas contagens

Hipóteses : $\mu_1 = \mu_2$

$\mu_1 \neq \mu_2$

$$\overline{x} = \frac{5210 + 5142}{2} = 5176$$

$$u_2 = 2(\sqrt{5210} - \sqrt{5176}) = 0{,}472$$

$$u_1 = 2(\sqrt{5142 + 1} - \sqrt{5176}) = -0{,}459$$

$$u_2 - u_1 = 0{,}931$$

Como $0{,}931 < 2{,}77$, aceita-se a igualdade das contagens num nível de confiança de 0,95 e portanto a pipeta de 0,2 mililitros está certa.

EXERCÍCIO 34

Suspeita-se que um certo mineral seja radioativo. Foi feita uma contagem do "background" durante 200 minutos obtendo-se um total de 11280 contagens. Em seguida foi feita uma contagem de uma massa m do minério durante 150 minutos com um resultado de 9480 contagens. Verificar, num nível de confiança de 0,95, se o mineral é radioativo.

Hipóteses : $\mu_1 = \mu_2$

$\mu_1 \neq \mu_2$

$$\bar{x} = \frac{11280 + 9480}{150 + 200} = 59{,}314$$

$$u_2 = 2\,(\sqrt{9480} - \sqrt{150 \times 59{,}314}) = 6{,}082$$

$$u_1 = 2\,(\sqrt{11280 + 1} - \sqrt{200 \times 59{,}314}) = 5{,}409$$

$$u_2 - u_1 = 11{,}491$$

Como $(u_2 - u_1) > 3{,}64$, rejeita-se H_0 mesmo num nível de confiança de 0,01. Pode-se então afirmar que o minério é radioativo.

V.4 – Método para Verificar se um Conjunto de Medidas se Distribui Segundo Poisson

Na análise estatística de resultados experimentais, é preciso às vezes verificar se esses resultados, pelas suas freqüências, obedecem à distribuição de Poisson, conforme se espera teoricamente. Para esta verificação, adota-se o critério do χ^2 já visto para a distribuição Normal:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\nu_i - \nu_i')^2}{\nu_i'} \qquad (V.13)$$

onde os ν_i são as freqüências experimentais e ν_i' são as freqüências teóricas calculadas pela fórmula (V.1). Como não se conhece a média verdadeira μ, usa-se sua estimativa $\bar{x}$.

A distribuição do somatório de (V.13) é aproximadamente a de um χ^2 com $f = k - m - 1$ graus de liberdade. Neste caso, $m = 1$, porque só o parâmetro $\mu \cong \bar{x}$ deve ser definido para o cálculo das freqüências teóricas.

Quando o número de observações for menor que 50, não se pode usar a fórmula (V.13). Neste caso, usa-se

$$\frac{\chi^2}{f} = \frac{s_x^2}{\sigma_x^2} \qquad (V.14)$$

onde se substitui σ^2_x por $\bar{x}$ que é estimativa de σ^2_x e temos:

$$\frac{\chi^2}{n-1} \cong \sum_{1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)\overline{x}} = \frac{s_x^2}{\overline{x}} \qquad \text{(V.15)}$$

Nesta fórmula, os x_i são os valores observados da variável x.

Se a relação $(s_x^2/\overline{x}) > 1$, verifica-se qual a probabilidade

$$P(\chi_f^2 > \chi_{exp,\,f}^2)$$

Se $(s_x^2/\overline{x}) < 1$, procura-se a probabilidade

$$P(\chi_f^2 < \chi_{exp,\,f}^2)$$

Os valores são encontrados em Tabelas que dão a probabilidade em função dos graus de liberdade. Quando as probabilidades achadas correspondem a um nível de significância menor do que 0,01, deve-se admitir a existência de uma divergência entre o conjunto examinado e a distribuição de Poisson. Os valores da distribuição de χ^2/f podem ser encontrados no "Statistical Tables and Formulas" – A. Hald – (1952) onde se acham valores até f = 1000.

A relação (V.15) pode ser usada também para valores pequenos de μ, isto é, $(1 < \overline{x} < 5)$

Quando μ é da ordem de 1, o número de observações deve ser maior que 15.

EXERCÍCIO 35

Para testar o funcionamento de um aparelho para contagem de baixa atividade, foram feitas 200 contagens de uma amostra, cada uma de 10 minutos e foram obtidos os resultados que se encontram na Tabela com as respectivas freqüências experimentais ν e teóricas ν'.

A média $\overline{x}$ das contagens foi: 7,41

As freqüências teóricas foram calculadas pela fórmula (V.1), usando 7,41 como estimativa de μ. Por exemplo, a freqüência teórica para o valor 5 quando a média é 7,41 é:

$$\nu_5 = \frac{200 \times 7{,}41^5}{5!} e^{-7{,}41} = 22{,}53$$

Encontramos na Tabela que $p(\chi^2 \geqslant 24{,}3) = 0{,}002$.

O resultado achado para χ^2 corresponde então a uma probabilidade de $\sim 0{,}002$ o que nos leva a concluir que o aparelho em questão deve ter algum defeito.

EXERCÍCIO 36

O mesmo aparelho foi testado com outra amostra da qual foram feitas 29 contagens de 20 minutos e cujos resultados são os da Tabela V.3. A média dos resultados foi 12,27.

Tabela V.2

Resultados de 200 Contagens de uma Amostra com Atividade Baixa

Resultados da Contagem	Freqüências ν	Freqüências ν'	$\nu - \nu'$	$\frac{(\nu - \nu')^2}{\nu'}$
0	0	0,12		
1	2	0,90		
2	1	3,32	4,55	1,65
3	5	8,21		
4	10	15,20	5,20	1,78
5	18	22,53	4,53	0,91
6	25	27,83	2,83	0,29
7	40	29,46	10,54	3,77
8	42	27,28	14,72	7,94
9	28	22,47	5,53	1,36
10	14	16,65	2,65	0,42
11	8	11,21	3,21	0,92
12	4	6,92		
13	1	3,95		
14	1	2,09		
15	0	1,03	9,1	5,14
16	1	0,48		
17	0	0,21		
18	0	0,09		
19	0	0,03		
20	0	0,01		
	200	199,99		$^2 = 24,18$

Tabela V.3

Resultados de 29 Contagens de uma Amostra

Resultados das Contagens	Freqüências	$\lvert x_i - \bar{x} \rvert$
8	1	4,27
9	0	3,27
10	4	2,27
11	6	1,27
12	7	0,27
13	5	0,73
14	1	1,73
15	3	2,73
16	1	3,73
17	0	4,73
18	1	5,73

$$\Sigma (x_i - \bar{x})^2 = 123,2838$$

$$\frac{\chi^2}{28} = \frac{123,2838}{28 \times 12,27} = 0,3588$$

$$\chi^2 = 10,05$$

Como $\chi^2/f < 1$, deve-se achar qual a probabilidade de χ^2_{28} ser menor que 10,05

Valores tabelados

$$P(\chi^2_{28} \geqslant 13,6) = 0,99$$

$$\therefore P(\chi^2_{28} < 13,6) = 0,01$$

Como o valor achado ($\chi^2_{exp} = 10,05$) é menor que 13,6 temos que:

$$P(\chi^2_{28} < 10,05) < 0,01$$

Este resultado confirma o do exercício anterior e concluímos que o contador é defeituoso.

V.5 – Avaliação de Resultados Semi-Quantitativos pela Distribuição de Poisson

Existem certos tipos de resultados que dependem das escalas dos aparelhos usados nas medidas. Por exemplo, os resultados semi-quantitativos de análises espectroscópicas são representados, em geral, por números discretos, porque são obtidos por meio de uma escala especialmente construída para esse fim. Se uma amostra com 0,01 % de um componente é analisada de acordo com uma escala que corresponde aos valores de 0,001, 0,003, 0,01, 0,03, 0,10 etc. os resultados só podem ser representados por esta série de números discretos. Pode-se dizer que os resultados são quantizados pela escala de medidas. Quando a quantização, provocada por uma escala de medidas deste tipo, é imposta a fatores sujeitos aos requisitos que obedecem ao Teorema do Limite Central de Liapunov, pode-se esperar a distribuição de Poisson. Os desvios são representados por um código especial que consiste de uma série de números inteiros positivos: 0, 1, 2, 3 . . .

Suponhamos que temos uma amostra com 0,01 % de urânio e usamos uma escala de medidas proporcionais a 1, 3, 10, 30, etc . . .

Numa série de análises podemos obter: 0,01 % 0,03 % 0,003 % etc. Teríamos então que:

1) **0,01%** seria codificado como zero (não há desvio);
2) **0,03% ou 0,003%** – este resultado cai no próximo intervalo (abaixo e acima do valor verdadeiro) e neste caso o código é 1;
3) **0,1% ou 0,001%** – neste caso o código é 2, etc.

Como resultado, achamos uma série de números: 0, 0, 2, 0, 1, 0, 1, 3, 0, 0 . . .

Acha-se o valor médio destes números achados e substitui-se na fórmula da distribuição de Poisson para obter os valores teóricos das freqüências.

EXERCÍCIO 37

Determinação espectrográfica de tungstênio num padrão com 0,01 %de W por um aparelho cuja escala de medidas é proporcional a 1, 3, 10, 30, etc.

Foram obtidos 146 resultados distribuídos conforme a Tabela V.4.

Tabela V.4

Resultados de Determinação Espectrográfica de Tungstênio

Resultado	Freqüência	Código	Freqüências		$\frac{(\nu - \nu')^2}{\nu'}$
			Experimental (ν)	Poisson (ν')	
0,01	92	0	92	99,44	0,557
0,03	23	1	49	38,18	3,066
0,003	26				
0,10	2	2	5	7,33	
0,001	3				
0,30	0	3	0	0,94	1,350
0,0003	0				
1,00	0	4	0	0,09	
0,0001	0				
Média (análise)	0,013%		146	145,9	4,973
Média (código)	0,384				

$$P(0) = \frac{0{,}384^0}{0!} e^{0{,}384}\ 146 = 99{,}44$$

$$P(1) = \frac{0{,}384^1}{1!} e^{0{,}384}\ 146 = 38{,}18$$

$$P(2) = \frac{0{,}384^2}{2!} e^{0{,}384}\ 146 = 7{,}33$$

$$P(3) = \frac{0{,}384^3}{3!} e^{0{,}384}\ 146 = 0{,}94$$

$$P(4) = \frac{0{,}384^4}{4!} e^{-0{,}384}\, 146 = 0{,}09$$

$$P(\chi^2_2 > 4{,}973) \cong 0{,}09$$

O teste indica que os resultados experimentais obedecem à distribuição de Poisson como era esperado o que prova que o método analítico adotado é conveniente para a determinação de W, pelo menos nessa concentração. Como a média neste caso foi 0,384, não há possibilidade de usar a distribuição Normal como aproximação.

Deve-se notar que a distribuição ocorre somente neste caso em que a maneira de codificar os erros é bem definida. Por exemplo, se os números 0, 1, 2, 3, . . . forem usados para codificar os resultados, isto é, atribuir o número zero para um certo intervalo de resultados, o número 1 para intervalos imediatamente superior e inferior, etc. vamos ter uma seqüência de números discretos, mas que não obedecem à distribuição de Poisson.

Suponhamos agora que uma amostra contendo 0,2 ug de U foi analisada por meio de um aparelho que tem uma escala que só dá os resultados 0,4 e 0,01. Nos dois casos o erro é de aproximadamente 100 % isto é:

$$1^{\circ}\ \text{caso} : \frac{0{,}4 - 0{,}2}{0{,}2} \times 100 = 100\%$$

$$2^{\circ}\ \text{caso} : \frac{0{,}2 - 0{,}01}{0{,}2} \times 100 = 95\%$$

e poderíamos esperar a ocorrência desses erros com igual probabilidade.

De acordo com a maneira exposta para codificar erros, a escala deveria ser 0,4 – 0,2 – 0,1 – 0,04 – 0,02 – 0,01, isto é, ao resultado 0,4 deveria ser atribuído o código 1 e ao resultado 0,01 o código 4 e a probabilidade de ocorrência desses dois erros deveria ser essencialmente diferente. No primeiro caso, o resultado e o valor real são da mesma ordem de grandeza, enquanto que no segundo caso eles diferem de mais de uma ordem de grandeza, o que não permite admitir que ambos tenham a mesma probabilidade de ocorrência.

CAPÍTULO VI

COMPARAÇÃO DE VÁRIAS MÉDIAS PELA ANÁLISE DA VARIÂNCIA

Em trabalhos experimentais, muitas vezes é de interesse a comparação de resultados obtidos por vários métodos. O problema se resume na comparação das médias obtidas para verificar se podem ser consideradas estatisticamente iguais no nível de significância desejado. O teste t só permite a comparação das médias duas a duas. Quando quisermos comparar 10 médias, por exemplo, temos que formar $\frac{10 \times 9}{2} = 45$ pares distintos para testes. Se o nível de significância adotado for 0,05, deve-se esperar que, em média, uma em cada vinte decisões seja errada. Neste caso, teríamos pela lei das probabilidades duas respostas erradas o que tornaria a aplicação do teste t, além de trabalhosa, de resultado duvidoso.

A finalidade da análise da variância é separar a variância em duas ou mais partes: uma consequente da técnica experimental adotada (erro aleatório do método) e a outra ou outras produzidas por um ou mais fatores cuja influência se quer estudar. Para ser possível a comparação das médias pelas análise da variância, é necessário que as variâncias dos conjuntos sejam estatisticamente iguais. O primeiro passo deve ser então a verificação da igualdade das variâncias por um dos métodos vistos no Capítulo IV.

VI.1 – Determinação da Variância Produzida por um Único Fator

É freqüente, em trabalhos experimentais, o estudo do comportamento de um sistema em vários níveis de temperatura. Neste caso, não são os resultados de vários métodos que serão comparados, mas o mesmo método aplicado nas mesmas condições, a não ser a variação do fator temperatura. Suponhamos que se deseje examinar m níveis de temperatura. Devem ser obtidos n_1 resultados no nível T_i, n_2 resultados no nível $T_2, \ldots, n_m$ resultados no nível T_m e os resultados são tabelados conforme o modelo da Tabela VI.1.

Tabela VI.1

Disposição dos Resultados Experimentais

	Níveis					
Determinação	1	2	---	i	---	m
1a.	x_{11}	x_{21}	---	x_{i1}	---	x_{m1}
2a.	x_{12}	x_{22}	---	x_{i2}	---	x_{m2}
.	.	.		.		.
.	.	.		.		.
.	.	.	.	.		.
jª	x_{1j}	x_{2j}	---	x_{ij}	---	x_{mj}
	x_{1n_1}	x_{2n_2}		x_{in_i}		x_{mn_m}

As médias parciais de cada coluna são:

$$\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} \qquad (i = 1,2, \ldots m) \qquad \text{(VI.1)}$$

A média geral de todos os resultados é:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{n_i} \sum_{i=1}^{m} x_{ij}}{\sum_1^m n_i} = \frac{\sum_1^m n_i \bar{x}_i}{\sum_1^m n_i} = \frac{\sum_1^m n_i \bar{x}_i}{N} \qquad \text{(VI.2)}$$

As hipóteses formuladas para este caso, devem ser:

H_0 – Todas as amostras provem de uma mesma população normal

$N(\mu,\sigma^2)$, isto é : $\mu_1 = \mu_2 = \ldots = \mu_m = \mu$

H_1 – Pelo menos uma das médias μ_i é diferente das demais

Se a segunda alternativa for certa, pelos menos uma das médias μ_i difere da média μ de um certo valor ϵ_i, isto é:

$$\epsilon_i = \mu - \mu_i \qquad \text{(VI.3)}$$

Com isto, as hipóteses podem ser formuladas da seguinte maneira:

H_0 – $\epsilon_i = 0$ (para qualquer valor de i)

H_1 – $\epsilon_i \neq 0$ (pelo menos para um valor de i)

Se H_0 for verdadeira, $\bar{\bar{x}}$ e todos os $\bar{x}_i$ são estimativas verdadeiras da μ.

Por hipótese, a variância σ^2 é a mesma para os m conjuntos e podemos estimá-la de três maneiras.

1) A partir de todos os resultados.

Se H_0 for verdadeira, os N resultados pertencem à mesma população e temos a seguinte estimativa para a variância σ^2 :

$$S_T^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2}{N - 1} \qquad \text{(VI.4)}$$

ou

$$\frac{(N-1)\,S_T^2}{\sigma^2} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2}{\sigma^2} = \frac{SQT}{\sigma^2}, \qquad \text{(VI.5)}$$

que é um χ^2 com N – 1 graus de liberdade. A abreviação SQT significa "soma dos quadrados totais".

2) Tomando como base o conjunto das médias $\bar{x}_i$ de cada uma das m amostras.

Cada média $\bar{x}_i$ pertence a uma população de médias que pode ser representada por $N(\mu, \sigma^2/n_i)$, se H_0 for verdadeira.

Por outro lado, a variância de uma média genérica $\bar{x}_i$ em relação à média geral $\bar{\bar{x}}$ pode ser expressa por:

$$s^2_{\bar{x}_i} = (\bar{x}_i - \bar{\bar{x}})^2 \qquad \text{(VI.6)}$$

que é uma estimativa justa de σ^2/n_i. Então:

$$n_i\, s^2_{\bar{x}_i} = n_i\,(\bar{x}_i - \bar{\bar{x}})^2 \qquad \text{(VI.7)}$$

é uma estimativa justa de σ^2.

Temos m conjuntos e portanto m estimativas deste tipo cuja média será uma estimativa melhor de σ^2. Como uma destas estimativas não é independente porque $\bar{\bar{x}}$ é a média das $\bar{x}_i$ temos:

$$S^2_E = \frac{\sum_{i=1}^{m} n_i\,(\bar{x}_i - \bar{\bar{x}})^2}{m-1} \qquad \text{(VI.8)}$$

onde S^2_E é a variância entre as amostras. Se H_0 for verdadeira, de (VI.8) temos:

$$\frac{(m-1)\,S^2_E}{\sigma^2} = \frac{\sum_{i=1}^{m} n_i\,(\bar{x}_i - \bar{\bar{x}})^2}{\sigma^2} = \frac{SQE}{\sigma^2} \qquad \text{(VI.9)}$$

que é um χ^2 com (m – 1) graus de liberdade e SQE é abreviação de "soma dos quadrados entre as amostras".

3) Outra maneira de estimar σ^2 é diretamente sobre cada amostra. Para a amostra genérica x_i temos:

$$s^2_i = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1} \qquad \text{(VI.10)}$$

que também é uma estimativa justa de σ^2, porque por hipótese a variância é a mesma para todas as amostras. Como temos m amostras e portanto m estimativas deste tipo, a combinação de todas vai dar:

$$S^2_R = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^{m} m(n_i - 1)} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{N - m} \qquad \text{(VI.11)}$$

e podemos escrever:

$$\frac{(N-m)\,S_R^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_j}(x_{ij}-\bar{x}_i)^2}{\sigma^2} = \frac{SQR}{\sigma^2} \qquad \text{(VI.12)}$$

que é um χ^2 com $(N-m)$ graus de liberdade. A abreviação SQR significa "soma dos quadrados dos resíduos".

A relação básica de análise da variância é:

$$SQT = SQE + SQR \qquad \text{(VI.13)}$$

ou

$$\sum_{i=1}^{m}\sum_{j=1}^{n_j}(x_{ij}-\bar{\bar{x}})^2 = \sum_{i=1}^{m} n_i(\bar{x}_i-\bar{\bar{x}})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_j}(x_{ij}-\bar{x}_i)^2 \qquad \text{(VI.14)}$$

Para simplificar essa demonstração vamos estabelecer que:

a soma dos resultados da amostra i é: $\sum\limits_{j=1}^{n_j} x_{ij} = T_i$

a média da amostra i é: $\bar{x}_i = \dfrac{T_i}{n_i}$

a soma de todos os resultados é: $\sum\limits_{i=1}^{m} T_i = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i} x_{ij} = T$

a média geral é: $\bar{\bar{x}} = T/N$

e indicar nos somatórios só os índices de variação.

Desenvolvendo SQT temos:

$$\sum_i\sum_j (x_{ij}-\bar{\bar{x}})^2 = \sum_i\sum_j (x_{ij}^2 - 2\,x_{ij}\,\bar{\bar{x}} + \bar{\bar{x}}^2)$$

$$= \sum_i\sum_j x_{ij}^2 - 2\bar{\bar{x}}\sum_i\sum_j x_{ij} + N\bar{\bar{x}}^2$$

$$= \sum_i\sum_j x_{ij}^2 - 2\,T^2/N + T^2/N$$

$$= \sum_i \frac{T_i^2}{n} - \frac{T^2}{N}$$

Para SQE, vamos ter:

$$\sum_i n_i (\bar{x}_i - \bar{\bar{x}})^2 = \sum_i n_i \bar{x}_i^2 - 2 \sum_i n_i \bar{x}_i \bar{\bar{x}} + \sum_i n_i \bar{\bar{x}}^2$$

$$= \sum_i n_i \bar{x}_i^2 - 2 \bar{\bar{x}} \sum_i T_i + \bar{\bar{x}}^2 \sum_i n_i$$

$$= \sum_i n_i x_i^2 - 2\, T^2/N + T^2/N$$

$$= \sum_i T_i^2/n_i - T^2/N \qquad \text{(VI.15)}$$

Para SQR, temos:

$$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 = \sum_i \sum_j x_{ij}^2 - 2 \sum_i \sum_j x_{ij} \bar{x}_i + \sum_i \sum_j \bar{x}_i^2$$

$$= \sum_i \sum_j x_{ij}^2 - 2 \sum_i T_i \bar{x}_i + \sum_i n_i \bar{x}_i^2$$

$$= \sum_i \sum_j x_{ij}^2 - 2 \sum_i \frac{T_i^2}{n_i} + \sum_i \frac{T_i^2}{n_i}$$

$$= \sum_i \sum_j x_{ij}^2 - \sum \frac{T_i^2}{n_i} \qquad \text{(VI.16)}$$

Somando (VI.15) e (VI.16) fica demonstrado (VI.12):

$$SQE + SQR = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} + \sum_i \sum_j x_{ij}^2 - \sum_i \frac{T_i^2}{n_i}$$

$$= \sum_i \sum_j x_{ij}^2 - \frac{T^2}{N} = SQT$$

VI.2 – Teorema de Cochrane

Se dividirmos um χ_k^2 em r parcelas, cada uma com distribuição $\chi_{k_i}^2$, de tal modo que $\sum_{i=1}^{r} k_i = k$, as parcelas são independentes. Em outras palavras, se:

$$\chi_k^2 = \chi_{k_1}^2 + \chi_{k_2}^2 + \ldots + \chi_{k_r}^2$$

e

$$\sum_i k_i = k$$

Então os $\chi_{k_i}^2$ são independentes.

No nosso caso temos:

$$\frac{SQT}{\sigma^2} = \frac{SQE}{\sigma^2} + \frac{SQR}{\sigma^2}$$

cada um com distribuição de χ^2 (se H_0 for verdadeira) respectivamente com $(N-1)$, $(m-1)$ e $(N-m)$ graus de liberdade. Como:

$$(N-1) = (m-1) + (N-m)$$

segue-se que os χ^2 são independentes. Em conseqüência o quociente:

$$\frac{SQE/_{(m-1)}}{SQR/_{(N-m)}} = \frac{S_E^2}{S_R^2} \qquad \text{(VI.17)}$$

será um F com $(m-1)$ e $(N-m)$ graus de liberdade e podemos usá-lo para testar H_0. Se H_0 não for verdadeiro:

$$E\{S_R^2\} = \sigma^2$$

mas

$$E\{S_E^2\} = \sigma^2 + \frac{\sum_i n_i \epsilon_i^2}{m-1} \neq \sigma^2 \qquad \text{(VI.18)}$$

O teste $F = SE^2/S_R^2$ é monocaudal, porque a priori sabe-se que, se H_0 não for verdadeira, $S_E^2 > S_R^2$, porque S_R^2 é a variância devida a erros aleatórios e S_E^2 é a variância devida a erros aleatórios e erros sistemáticos, se houverem. Para a análise de variância é conveniente dispor os elementos conforme a Tabela (VI.2).

Se os resultados da Tabela VI.1 forem números muito grandes ou muito pequenos, pode-se subtrair ou somar qualquer número a todos, sem alterar a variância. Se os resultados forem números fracionários, pode-se multiplicar todos por um mesmo número, porque as variâncias ficam multiplicadas pelo quadrado do número, mas o F calculado, sendo o quociente de duas variâncias continua o mesmo.

Por exemplo, se tivermos números como 0,057; 0,068; 0,061 e 0,070, pode-se multiplicar por 1000 e subtrair 57 de todos e trabalhar com o conjunto: 0; 11; 4; 13.

EXERCÍCIO 38

Foi feito um estudo da extração de um elemento M por um solvente X em função da acidez da fase aquosa. Foram examinados vários níveis de pH para determinar a faixa em que a extração é maior, num nível de confiança de 0,90. Os resultados em função dos pH: A, B, ... F estão na Tabela (VI.3).

O teste de Bartlett aplicado às variâncias deu como resultado que são estatisticamente iguais num nível de significância de 0,05. Para simplificar os cálculos os resultados foram submetidos à seguinte transformação:

Tabela VI.2

Disposição dos Elementos para a Análise da Variância

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F
Entre amostras	$SQE = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N}$	$m - 1$	$\frac{SQE}{m-1} = S_E^2$	$F = \frac{S_E^2}{S_R^2}$
Dentro das amostras (resíduo)	SQR (por diferença)	$N - m$	$\frac{SQR}{N-m} = S_R^2$	
Total	$SQT = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{N}$	$N - 1$		

Tabela VI.3

Resultados de % de Extração de M em função do pH

Determinações	pH_1	pH_2	pH_3	pH_4	pH_5	pH_6
1	87,2	80,2	90,8	89,8	90,8	88,8
2	90,0	89,6	91,2	93,6	94,6	90,0
3	89,4	88,2	89,4	92,4	93,2	87,6
4	88,6	89,6	92,6	94,0	95,0	89,2
5	86,8	87,0	91,8	91,8	92,8	88,4
6	87,4	88,8	92,4	93,2	94,0	87,8
7		87,2			91,8	
$\overline{x}$	88,23	88,66	91,37	92,47	93,17	88,63
$\lvert s_x \rvert$	1,30	1,24	1,08	1,53	1,51	0,82

$$Y_{ij} = 5(x_{ij} - 90{,}0)$$

e foi obtido o quadro de números com os respectivos quadrados que estão na Tabela (VI.4).

Tabela VI.4

Resultados da Tabela (IV.3) Depois de Transformados
Para Simplificar os Cálculos

Determinações	pH_1		pH_2		pH_3		pH_4		pH_5		pH_6		
1	−14	196	1	1	4	16	−1	1	4	16	−6	36	
2	0	0	−2	4	6	36	18	324	23	529	0	0	
3	−3	9	−9	81	−3	9	12	144	16	256	−12	144	
4	−7	49	−2	4	13	169	20	400	25	625	−4	16	
5	−16	256	−15	225	9	81	9	81	14	196	−8	64	
6	−13	169	−6	36	12	144	16	256	20	400	−11	121	
7			−14	196					9	81			
T_i	−53		−47		41		74		111		−41		$T = 85$
$\sum_j x_{ij}^2$		679		547		455		1206		2103		381	$\sum_i \sum_j x_{ij} = 5371$
T_i^2/n_i	468,2		315,6		280,2		912,7		1760,1		280,2		$\sum_i T_i^2/n_i = 4016,9$
$\bar{x}$	−8,83		−6,71		6,83		12,33		15,9		−6,83		

$N = 38$

$$\frac{T^2}{N} = 190,13$$

$SQT = 5371 - 190,13 = 5180,87$

$SQE = 4016,9 - 190,13 = 3825,77$

$SQR = 5180,87 - 3825,77 = 1355,10$

Tabela VI.5

Análise da Variância dos Resultados da Tabela (VI.3)

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Entre os Resultados	3825,77	5	765,154	$F = 18,07$
Resíduo	1355,10	32	42,347	$F_{0,05(5,32)} = 2,5$
Total	5180,87	37		

Como F experimental é maior que o F tabelado, concluímos que as médias das colunas da Tabela (VI.3) não são iguais. Para saber quais são as médias estatisticamente diferentes aplica-se o teste de Scheffé que tem a vantagem de usar os próprios elementos da análise da variância. Scheffé demonstrou que se pode adotar como critério a *menor diferença significativa no nível desejado*, pela seguinte fórmula:

$$\Delta_\alpha = \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \sqrt{(m-1)\, F_{\alpha, (m-1), (N-m)}} \qquad \text{(VI.19)}$$

No caso do exercício teríamos para os n_i e n_j os valores 6 e 6, 6 e 7 e 7 e 7, isto é:

Comparação entre colunas com:

$0{,}05 = 42{,}347 \left(\frac{1}{6} + \frac{1}{6}\right) 5 \times 2{,}5 = 13{,}28$ .... 6 e 6 elementos

$0{,}05 = 42{,}347 \left(\frac{1}{6} + \frac{1}{7}\right) 5 \times 2{,}5 = 12{,}80$ .... 6 e 7 elementos

$0{,}05 = 42{,}347 \left(\frac{1}{7} + \frac{1}{7}\right) 5 \times 2{,}5 = 12{,}30$ .... 7 e 7 elementos

Estes resultados significam que a maior diferença entre as médias de duas colunas com seis elementos deve ser 13,28, de sete elementos é 12,30 e entre duas colunas com 6 e 7 elementos respectivamente deve ser 12,30.

Na Tabela VI.4, a diferença entre as colunas B e C, por exemplo, é 6,83 – (-6,71) = 13,54 maior que 12,80 e portanto são diferentes. Fazendo uma comparação entre todas as colunas, duas a duas concluímos que a faixa de pH de C a E é favorável para a extração de M, no nível de significância adotado. A análise da variância é usada não só para estudar as melhores condições para um determinado processo, mas também para controles rotineiros de funcionamento de aparelhos, reagentes, etc. . .

Um exemplo prático de aplicação da análise da variância pode ser encontrado na publicação IEA 205 onde se apresenta um problema de identificação de vidros de interesse às Ciências Forenses.

VI.3 – Condições para que a Análise da Variância possa ser Aplicada

O método da análise da variância é uma generalização do método da comparação de duas médias pelo critério t. A análise da variância bem como a comparação de duas médias exige as seguintes condições para poder ser aplicada:

1) As séries de medidas devem poder ser consideradas como amostras aleatórias de distribuições normais.

2) As variâncias devem ser estatisticamente iguais para todas as medidas.

A primeira condição não é muito rigorosa a não ser que se trate de uma distribuição bem definida e substancialmente diferente da distribuição normal.

Algumas vezes acontece que certas médias diferem muito das demais. Esta situação se encontra, por exemplo, no estudo da reprodutibilidade dos resultados de um mesmo método, obtidos em laboratórios diferentes. Neste caso, essas medidas devem ser excluídas da análise da variância e as condições de trabalho nos laboratórios correspondentes devem ser examinadas.

A segunda condição é mais rigorosa e é necessário fazer uma investigação sobre as variâncias por um critério antes de aplicar o teste. Quando as variâncias não são estatisticamente iguais, é possível recorrer a artifícios conforme os seguintes exemplos:

a) Em química analítica, se reunirmos num só conjunto amostras em que a concentração do componente varia muito, aplica-se uma transformação de variável do tipo $y = \log x$, conforme já foi visto anteriormente.

b) No caso de medidas radioativas, os resultados devem ser transformados com o auxílio da função $y = 2\sqrt{x}$ para obter uma variável aleatória em que a variância não depende da média.

c) No estudo comparativo de vários métodos experimentais, pode-se achar que as precisões diferem consideravelmente. Neste caso, é possível adotar dois artifícios:

1) separar os resultados em grupos de tal maneira que em cada grupo a precisão dos métodos seja a mesma.

2) para os grupos de resultados com menor precisão, obter resultados paralelos e usar as médias como resultados individuais.

VI.4 — Aplicação da Análise da Variância no Estudo da Influência de 2 Fatores

No desenvolvimento de um método experimental, vários fatores podem influir nos resultados. No exercício visto, consideramos a variação do pH, mas a concentração dos reagentes, pressão, temperatura, etc., também podem causar alterações nos resultados.

Quando se estuda a variação causada por dois fatores, os resultados são dispostos numa Tabela semelhante à Tabela (VI.1), mas com entradas horizontais e verticais e as respectivas médias. Para diferenciar os resultados de uma linha dos resultados de uma coluna, adota-se a seguinte nomenclatura:

Total da coluna i : $T'_i = \sum_{j=1}^{n} x_{ij} \qquad (i = 1, 2, \ldots m)$

Média da coluna i : $\bar{x}'_i = T'_i/m$

Total da linha j : $T_j = \sum_{i=1}^{m} x_{ij} \qquad (j = 1, 2, \ldots n)$

Média da linha j : $\bar{x}_j = T_j/n$

Total Geral : $T = \sum_i \sum_j x_{ij} = \sum_i T'_i = \sum_j T_j$

Média Geral : $\bar{x} = \dfrac{T}{mn} = \sum_i x_i = \sum_j \bar{x}_j$

Hipóteses:

$$H_0 \begin{cases} \mu_i = \mu & \text{para qualquer } i \ (i = 1, 2, \ldots m) \\ \mu_j = \mu & \text{para qualquer } j \ (j = 1, 2, \ldots n) \end{cases}$$

H_1 — Pelo menos uma das μ_i ou μ_j é diferente das demais.

Podemos supor um efeito δ_i nos resultados das colunas i e um efeito ϵ_j nas linhas j sobre a média geral μ, isto é:

$$E(x_{ij}) = \mu + \delta_i + \epsilon_j$$

$$\mu_i = \mu + \delta_i$$

$$\mu_j = \mu + \epsilon_j$$

Como μ é a média geral da população, se H_o é verdadeira temos:

$$\Sigma \delta_i = 0$$

$$\Sigma \delta_j = 0$$

Temos então duas hipóteses para testar:

$$H_{01} : \delta_i = 0 \qquad i = 1,2, \ldots m$$

$$H_{11} : \delta_j \neq 0 \qquad \text{pelo menos para um } i$$

e

$$H_{02} : \epsilon_j = 0 \qquad j = 1,2, \ldots n$$

$$H_{12} : \epsilon_j \neq 0 \qquad \text{pelo menos para um } j$$

A equação fundamental da análise da variância no caso da influência de dois fatores é:

$$\sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 = \sum_i n\,(\bar{x}_i - \bar{\bar{x}}) + \sum_j m\,(\bar{x}_j - \bar{\bar{x}}) + \sum_i \sum_j (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2 \qquad \text{(VI.20)}$$

que se obtém pelo desenvolvimento da seguinte igualdade:

$$\sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 = \sum_i \sum_j [(\bar{x}_i - \bar{\bar{x}}) + (x_j - \bar{\bar{x}}) + (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})]^2$$

$$= \sum_i \sum_j (\bar{x}_i - \bar{\bar{x}})^2 + \sum_i \sum_j (\bar{x}_j - \bar{\bar{x}})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$$

$$= \sum_i m (\bar{x}_i - \bar{\bar{x}})^2 + \sum_j n (\bar{x}_j - \bar{\bar{x}})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$$

Os duplos produtos se anulam, porque os sinais podem ser positivos ou negativos com igual probabilidade.

O primeiro membro da equação (VI.19) é a soma dos quadrados totais (SQT) que dividida pelo número de graus de liberdade $(mn - 1)$ é $S^2{}_T$, estimativa de σ^2.

O primeiro termo do segundo membro é a soma dos quadrados correspondentes às linhas (SQL) que dividido pelo número de graus de liberdade $(n - 1)$ da $S^2{}_L$ também estimativa de σ^2. O segundo termo do segundo membro é a soma dos quadrados correspondentes às colunas (SQC) que dividido pelo número de graus de liberdade $(m - 1)$ é também estimativa de σ^2. Finalmente, o último termo é o resíduo (SQR) que dividido pelo número de graus de liberdade $(m - 1)\ (n - 1)$ é estimativa de σ^2 e representa a variância devida a erros aleatórios.

Por diferença temos que:

$$SQR = SQT - SQL - SQC$$

SQR, SQL e SQC formam uma partição de SQT, são independentes e podem portanto ser usados para testar as hipóteses:

$$H_{01} \text{ por } : F_L = \frac{SQL/(n-1)}{SQR/(m-1)\ (n-1)} = \frac{S_L^2}{S_R^2} \qquad \text{(VI.21)}$$

$$H_{02} \text{ por } : F_C = \frac{SQC/(m-1)}{SQR/(m-1)\ (n-1)} \qquad \text{(VI.22)}$$

com as regiões críticas monocaudais.

O quadro da análise da variância neste caso, é o da Tabela VI.6.

Para achar quais as linhas ou colunas são diferentes, caso o teste der significativo, aplica-se o teste de Scheffé como foi feito no exercício anterior.

Tabela VI.6

Disposição dos Elementos da Análise da Variância Devida a Dois Fatores

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F
Entre as Linhas	SQL	$n-1$	$S_L^2 = \dfrac{SQL}{n-1}$	$F_L = \dfrac{S_L^2}{S_R^2}$
Entre as Colunas	SQC	$m-1$	$S_C^2 = \dfrac{SQC}{m-1}$	$F_C = \dfrac{S_C^2}{S_R^2}$
Resíduo	Diferença	$(m-1)(n-1)$	$S_R^2 = \dfrac{SQR}{(m-1)(n-1)}$	
Total	SQT	$mn-1$		

EXERCÍCIO 39

Para investigar a influência do tempo de uso no funcionamento de um conjunto de 5 aparelhos (A, B, . . . E) foram feitas 8 medidas sucessivas de um padrão (valor esperado: $\mu = 10{,}0$) uma vez cada dois meses, durante 16 meses. (T_1, T_2, . . . T_8). Verificar se o tempo influiu igualmente em todos os aparelhos, num nível de confiança 0,95.

Os resultados obtidos estão na Tabela (VI.7).

Tabela VI.7

Resultados de Medidas Obtidas com 5 Aparelhos

	A	B	C	D	E	$\bar{x}$
T_1	10,1	9,99	9,8	10,2	9,8	9,96 ± 0,18
T_2	9,8	10,1	9,4	9,7	10,2	9,84 ± 0,32
T_3	9,9	10,2	9,6	9,9	9,9	9,90 ± 0,21
T_4	10,2	9,8	10,0	9,5	10,1	9,92 ± 0,28
T_5	9,7	10,0	9,3	10,1	10,2	9,86 ± 0,36
T_6	9,8	10,2	9,5	9,4	10,0	9,78 ± 0,33
T_7	10,0	9,9	9,2	9,8	9,6	9,70 ± 0,32
T_8	9,7	9,8	9,6	9,6	9,9	9,72 ± 0,13
$\bar{x}$	9,90 ± 0,18	9,99 ± 0,16	9,55 ± 0,26	9,77 ± 0,28	9,96 ± 0,21	

O critério de Cochrane aplicado às variâncias deu como resultado que são estatisticamente iguais num nível de confiança 0,95. Foi feita uma transformação de variável de acordo com:

$$Y_{ij} = (x_{ij} - 10)\ 10$$

e foi obtido o conjunto da Tabela (VI.8).

Tabela VI.8

Transformação dos Resultados da Tabela (VI.7) e Cálculos

	A	B	C	D	E	T'_j	T_j^2/m
T_1	1(1)	-1(1)	-2(4)	2(4)	-2(4)	-2	0,8
T_2	-2(4)	1(1)	-6(36)	-3(9)	2(4)	-8	12,8
T_3	-1(1)	2(4)	-4(16)	-1(1)	-1(1)	-5	5,0
T_4	2(4)	-2(4)	0(0)	-5(25)	1(1)	-4	3,2
T_5	-3(9)	0(0)	-7(49)	1(1)	2(4)	-7	9,8
T_6	-2(4)	2(4)	-5(25)	-6(36)	0(0)	-11	24,2
T_7	0(0)	-1(1)	-8(64)	-2(4)	-4(16)	-15	45,0
T_8	-3(9)	-2(4)	-4(16)	-4(16)	-1(1)	-14	39,2
T_i	-8	-1	-36	-18	-3	-66	140,0
$\sum_j x_{ij}^2$	(32)	(19)	(210)	(96)	(31)	(388)	
T^2/n	8	0,125	162	40,5	1,125	211,75	

$$\frac{T^2}{N} = \frac{66^2}{40} = 108,9$$

SQT = 388 − 108,9 = 279,1

SQC = 211,75 − 108,9 = 102,75

SQL = 140,0 − 108,9 = 31,1

SQR = 279,1 − 133,85 = 145,25

Valores teóricos de F num nível de significância 0,05:

$$F_{7,28} = 2,4$$

$$F_{4,28} = 2,7$$

Tabela VI.9

Elementos para a Análise da Variância

Fonte de Variação	Soma de Quadrados	Número de Graus de Liberdade	Quadrado Médio	F
Entre os Tempos	23,2	7	3,314	$F_L = \frac{3,314}{3,29}$
Entre os Aparelhos	154,065	4	38,516	$F_L = 1,007$
Resíduo	92,135	28	3,290	$F_C = \frac{38,516}{3,29}$
Total	269,4	39		$F_C = 11,706$

Concluímos que o tempo de uso não afetou os aparelhos, porque

$$F_L = 1,007 < F_{0,05(7,28)} = 2,4$$

Apesar disso, o funcionamento dos aparelhos apresentou uma anomalia porque:

$$F_C = 11,706 > F_{0,05(4,28)} = 2,7$$

Para saber quais os aparelhos afetados de erro, aplica-se o Teste de Scheffé às colunas

$$\Delta_{0,05} = \sqrt{3,29 \times \left(\frac{1}{8} + \frac{1}{8}\right)(5-1) \times 2,7} = \pm 2,98 \cong \pm 3$$

As médias das colunas da Tabela (VI.9) são:

	A	B	C	D	E
$\bar{x}$:	-1	-0,125	-4,5	-2,25	-0,375

Se excluirmos o resultado do aparelho C, todos os demais diferem entre si e também do padrão ($\mu = 0$, pela transformação feita) de um número inferior a ± 3. Pode-se concluir então que o aparelho C tem um defeito de fabricação ou houve algum problema quanto à maneira de ser usado durante o período considerado.

VI.5 – Plano Experimental para Melhorar a Precisão dos Resultados

Muitas vezes um experimento é realizado sem um plano pré-estabelecido. Como conseqüência, o erro experimental pode ser grande ou ter número pequeno de graus de liberdade. Este fato torna o teste pouco sensível, quando se quer interpretar os resultados pela análise da variância. Existem várias maneiras que permitem diminuir o erro experimental, se examinarmos, "a priori", o plano experimental para prever o aspecto que a análise dos resultados obtidos vai tomar. Suponhamos que, numa produção de laboratório, se queira estudar a influência das concentrações A, B e C de um reagente. Decide-se fazer três provas paralelas com cada uma das concentrações para poder ter uma estimativa do erro experimental baseado sobre a variância de cada um dos três grupos. O trabalho envolvido na realização destas provas permite que sejam realizadas só 3 por dia, em seqüência, de modo que poderiam ser distribuídas da seguinte maneira:

1º dia	A	A	A	
2º dia	B	B	B	(1)
3º dia	C	C	C	

Esta disposição não leva em consideração as variações possíveis que podem ocorrer de um dia para outro. Pode acontecer, por exemplo, que, no terceiro dia, a temperatura ambiente seja bem mais alta que nos dias anteriores. Se não houver disponibilidade de um termostato, a influência da temperatura se confunde com a da concentração C e será impossível dizer, no caso de ter havido melhora de rendimento, se esta é causada pela concentração. Em todos os experimentos em que se aplica análise estatística, deve-se respeitar o fator "sorte", isto é, não se deve estabelecer uma ordem. No caso, a disposição das provas poderia ser sorteada e poderíamos ter:

1º dia	B	A	B	
2º dia	C	A	A	(2)
3º dia	B	C	C	

A análise da variância seria feita de acordo com o seguinte quadro

Fonte de Variação	Graus de Liberdade
Entre Concentrações	2
Erro Experimental	6
Total	8

Mesmo neste caso, o experimento não é tão sensível como poderia ser, porque houve repetição de concentrações num mesmo dia e o efeito do dia (temperatura, por exemplo) ainda poderia ser sensível.

VI.5.1 – Método dos Blocos

Foi usado inicialmente em experimentos agrícolas, mas, de um modo geral, chamam-se blocos todos os agrupamentos de provas impostos pela necessidade de aumentar a precisão experimental. No nosso caso, os dias constituem 3 blocos e pode-se imaginar que, se cada concentração for testada em três dias diferentes, a influência do dia pode ser eliminada.

Mesmo assim, não se deve obedecer à seqüência A B C em cada dia, porque existe neste caso uma ordem sistemática. Para haver aleatoriedade, as seqüências deveriam ser sorteadas e poderíamos ter:

1º dia	A	C	B	
2º dia	B	C	A	(3)
3º dia	B	A	C	

e a análise da variância teria o seguinte quadro

Fonte de Variação	Graus de Liberdade	Variância
Entre Concentrações	2	s_3^2
Entre Dias	2	s_2^2
Erro Experimental	4	s_1^2
Total	8	

Na organização (3) nota-se que há uma repetição nas colunas. Na primeira coluna, há duas vezes a concentração B, isto é, nas primeiras horas de trabalho de dois dias se testa a mesma concentração o que também poderia ser uma fonte de erro.

Comparando o plano (2) com o plano (3) nota-se que, neste, o erro experimental perde dois graus de liberdade o que dá idéia de uma perda de sensibilidade. Por outro lado, o teste $F = s_2^2/s_1^2$ vai permitir decidir se a variância entre os dias é ou não significativa. Se o resultado não for significativo, as variâncias s_1^2 e s_2^2 podem ser englobadas e o erro experimental terá novamente 6 graus de liberdade.

A melhor organização para um experimento deste tipo se faz pelo chamado **"quadrado latino"** onde cada concentração pode ser realizada em dias diferentes por um operador diferente. Teríamos assim:

	Operadores I	II	III
1º dia	A	B	C
2º dia	C	A	B
3º dia	B	C	A

Neste caso, as comparações feitas entre as médias dos resultados das três concentrações A, B e C são independentes das variações devidas aos dias e das variações devidas aos operadores. A análise da variância será então de acordo com o quadro que segue.

Este plano só dá 2 graus de liberdade para o erro experimental mas este erro será bem pequeno se a influência das linhas e colunas for importante sobre a variação dos rendimentos.

Fonte de Variação	Graus de Liberdade	Variância
Entre Operadores (colunas)	2	s_4^2
Entre Dias (linhas)	2	s_3^2
Entre Concentrações (letras)	2	s_2^2
Resíduo (erro experimental)	2	s_1^2
Total	8	

EXERCÍCIO 40

O experimento consiste em verificar, num nível de significância 0,05, se a concentração de urânio tem influência na porcentagem da extração com um certo agente extrator. Para isto foram usados 4 níveis de concentração: A, B, C e D. No mesmo dia é possível fazer 6 provas em seqüência, mas para usar o plano do "quadrado latino" são feitas somente 4, na seguinte ordem:

A	B	C	D
C	D	B	A
D	C	A	B
B	A	D	C

Os resultados obtidos, como porcentagem da extração, correspondentes a esta ordem foram:

		I	II	III	IV
1º dia	:	89	88	91	94
2º dia	:	90	92	90	90
3º dia	:	95	93	89	89
4º dia	:	91	90	94	92

Neste problema, podemos verificar a influência dos dias (linhas), dos períodos do dia (colunas) em que foi o experimento e das concentrações (letras). Para simplificar, subtraímos 90 de todos os resultados e obtemos os números da Tabela VI.10, onde constam também os respectivos quadrados

Tabela VI.10

Resultados do Exercício 40 depois de Transformados

	I	II	III	IV	Σx	Σx^2
	-1(1)	-2(4)	1(1)	4(16)	2	22
	0(0)	2(4)	0(0)	0(0)	2	4
	5(25)	3(9)	-1(1)	-1(1)	6	36
	1(1)	0(0)	4(16)	2(4)	7	21
Σx	5	3	4	5	17	
Σx^2	(27)	(17)	(18)	(21)		83

$\Sigma x_A = -2$	$\Sigma x_A^2 = 2$
$\Sigma x_B = -2$	$\Sigma x_B^2 = 6$
$\Sigma x_C = 6$	$\Sigma x_C^2 = 14$
$\Sigma x_D = \underline{15}$	$\Sigma x_D^2 = \underline{61}$
17	83

Calcula-se:

a) A soma dos quadrados das 16 observações. Resultado: 83

b) O fator de correção da média: $T^2/N = 17^2/16 = 18{,}0625$

c) A soma dos quadrados das somas das colunas, dividida pelo número de observações em cada coluna

$$\frac{5^2 + 3^2 + 4^2 + 5^2}{4} = 18{,}75$$

d) A soma dos quadrados das somas das linhas, dividida pelo número de observações em cada linha

$$\frac{2^2 + 2^2 + 6^2 + 7^2}{4} = 23{,}25$$

e) A soma dos quadrados da soma dos resultados de cada concentração dividida pelo número de observações para cada concentração.

$$\frac{(-2)^2 + (-2)^2 + 6^2 + 15^2}{4} = 67{,}25$$

Com esses resultados, organiza-se a Tabela (VI.11) para a análise da variância.

Verifica-se que o único valor de F maior que o tabelado é aquele calculado com a variância causada pelas concentrações. Conclui-se então que os períodos e os dias não interferem e a análise da variância pode ser feita levando em conta só o fator concentração.

VI.5.2 — Blocos Incompletos

Suponhamos que, no exercício anterior, temos que comparar a porcentagem de extração correspondente a 7 níveis de concentração de U e dispomos de 3 extratores adaptados a agitadores um pouco diferentes. Se quisermos eliminar o fator agitação do erro experimental, deve-se testar cada concentração nos três extratores, considerando cada extrator como um bloco e distribuindo ao acaso os 7 tratamentos (concentrações) de A a G. Por exemplo, o quadro que segue a Tabela VI.11.

Tabela VI.11

Análise da Variância dos Valores da Tabela VI.10

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Entre Concen-ções	49,1875(e-b)	3	16,3958	$F_{conc} = \frac{16,3958}{1,8125} = 9,04$
Entre Dias (linhas)	4,1875(d-b)	3	1,3958	$F_{dias} = \frac{1,3958}{1,8125} = 0,77$
Entre Períodos (colunas)	0,6875(c-b)	3	0,2292	$F_{período} = \frac{0,2292}{1,8125} = 0,13$
Resíduo (erro expe mental)	10,875	6	1,8125	
Total	64,9375(a-b)	15		$F_{3,6(0,05)} = 4,8$

Extrator	Concentrações						
I	B	E	D	A	F	G	C
II	G	A	D	C	F	B	E
III	A	F	G	C	E	B	D

O erro experimental neste caso vai ter (3 – 1) (7 – 1) = 12 graus de liberdade. Levando-se em conta que cada prova exige 2 dias para ser executada e que só podem ser feitas 3 de cada vez, o experimento completo vai exigir 14 dias de trabalho, isto é, quase 3 semanas, se houver interrupção no sábado e domingo. Como alguns fatores podem variar durante este período, é conveniente, para evitar flutuações, distribuir no tempo o total das provas, em um certo número de blocos menores. Para levar em conta os 2 fatores (tempo e extrator) dividem-se as 21 provas em 7 blocos sucessivos de 3 provas cada um, porque estas são feitas simultaneamente em 3 extratores. Para construir um esquema deste tipo, chamado **"quadrado de Youden"** é necessário que:

a) o número de blocos seja igual ao número de tratamentos;

b) o número de repetições de cada prova seja igual ao número de unidades por bloco;

c) cada tratamento (concentração) apareça uma só vez em cada linha (extrator);

d) cada tratamento só apareça uma vez em cada bloco na mesma posição.

Pode-se ter, por exemplo, o seguinte esquema:

	Blocos						
Extratores	1	2	3	4	5	6	7
I	B	E	D	A	F	G	C
II	G	A	C	D	B	F	E
III	A	F	G	C	E	B	D

Observa-se que cada extrator corresponde a uma linha e aparece uma só vez em cada bloco (coluna). Do mesmo modo, cada concentração só aparece uma vez em cada coluna (bloco) ou linha (extrator).

O problema poderá ser compreendido mais facilmente por meio de um exemplo prático, onde os blocos (dias) constituem as colunas e os tratamentos (concentrações) constituem as linhas.

EXERCÍCIO 41

Foi realizado um experimento de extração de um elemento M, onde se variou a concentração de M (A, B, . . . G) e foram usados 3 tipos de agitação (I, II, III). Para a execução do trabalho foram necessários 14 dias (blocos). A distribuição das provas e os resultados em porcentagem de extração (deduzidos 90 %) estão na Tabela VI.12.

A influência devida à concentração, livre da influência dos blocos, pode ser avaliada da seguinte maneira:

Para cada T_L (soma da linha) calcula-se um valor Q correspondente:

$Q = k(T_L)$ – soma dos totais dos k blocos que contém o tratamento considerado

Por exemplo:

$$Q_A = 3 \times 1{,}5 - (4{,}4 + 5{,}3 + 3{,}3) = -8{,}5$$

onde, entre parentesis, estão os totais das colunas que envolvem o tratamento A.

Calcula-se depois P que é uma estimativa do desvio médio provocado pela variação da concentração, a partir da média geral das 21 provas. Obtém-se dividindo cada valor de Q por $N(k-1)/(t-1)$ *onde* k é o número de extratores e t é o número de concentrações. No nosso caso temos:

$$\frac{N(k-1)}{t-1} = \frac{21 \times 2}{6} = 7$$

Tabela VI.12

Resultados Obtidos para um Experimento com 3 Variáveis

Tratamentos	Blocos (2 dias cada)							T_L	Q	P
(Concentrações)	1	2	3	4	5	6	7			
A	1,2 (1,44) (III)	0,8 (0,64) (II)		-0,5 (0,25) (I)				1,5	-8,5	-1,21
B	0,6 (0,36) (I)				0,0 (0) (II)	0,9 (0,81) (III)		1,5	-9,3	-1,33
C			1,4 (1,96) (II)	1,7 (2,89) (III)			0,8 (0,64) (I)	3,9	-0,9	-0,13
D			0,9 (0,81) (I)	2,1 (4,41) (II)			1,4 (1,96) (III)	4,4	0,6	0,08
E		2.5 (6,25) (I)			1.8 (3,24) (III)		2.0 (4,00) (II)	6,3	5,3	0,76
F		2,0 (4,00) (III)			2,3 (5,29) (I)	1,9 (3,61) (II)		6,2	3,9	0,56
G	2,6 (6,76) (II)		2,8 (7,84 (III)			2,5 (6,25) (I)		7,9	8,9	1,27
T_c	4,4	5,3	5,1	3,3	4,1	5,3	4,2	31,7	0	0

Então:

$$P_A = -\frac{8,5}{7} = -1,21$$

A média geral é: $\bar{\bar{x}} = 31,7/21 = 1,51$

As médias dos resultados para as concentrações A, B, . . . G depois de ajustadas são

$$\begin{aligned}
\bar{A} &= 1,51 - 1,21 = 0,30\\
\bar{B} &= 1,51 - 1,33 = 0,18\\
\bar{C} &= 1,51 - 0,13 = 1,38\\
\bar{D} &= 1,51 + 0,08 = 1,59\\
\bar{E} &= 1,51 + 0,76 = 2,27\\
\bar{F} &= 1,51 + 0,56 = 2,07\\
\bar{G} &= 1,51 + 1,27 = 2,78
\end{aligned}$$

Para analisar os resultados calcula-se como de costume:

a) A soma dos quadrados dos resultados

$$1,44 + 0,36 + \ldots + 3,61 + 6,25 = 63,41$$

b) A soma dos quadrados correspondentes aos blocos

$$\frac{4,4^2 + 5,3^2 + \ldots + 4,2^2}{3} - \frac{31,7^2}{21} = 1,11$$

c) A soma dos quadrados correspondentes aos extratores I, II, III

Totais por extrator:

I) $-0,5 + 0,6 + \ldots + 2,5 = 9,1$

II) $0,8 + 0,0 + \ldots + 2,6 = 10,8$

III) $1,2 + 0,9 + \ldots + 2,8 = 11,8$

$$\frac{9,1^2 + 10,8^2 + 11,8^2}{7} - \frac{31,7^2}{21} = 0,53$$

d) A soma dos quadrados correspondentes às concentrações ajustadas

É dada pela fórmula:

$$\frac{Q^2}{Nk\,(k-1)\,/\,(t-1)}$$

Então temos:

$$\frac{(-8{,}5)^2 + (-9{,}3)^2 + \ldots + 8{,}9^2}{21} = 13{,}45.$$

A soma dos quadrados do erro experimental é achada por diferença. A análise da variância é calculada de acordo com a Tabela VI.13.

Tabela VI.13

Análise da Variância dos Resultados da Tabela VI.12

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Entre Concentrações Ajustadas	13,45	6	2,24	$F_{conc} = \frac{2{,}24}{8{,}05} = 0{,}25$
Entre Extratores	0,53	2	0,26	$F_{extr} = \frac{0{,}26}{8{,}05} = 0{,}03$
Entre Blocos	(1,11)	(6)	(0,18)	$(F_1 = \frac{1{,}11}{8{,}05} = 0{,}13)$
Resíduo (Erro Experimental	48,32	6	8,05	
Total	63,41	20		$F_{2,6(0{,}05)} = 5{,}1$; $F_{6,6(0{,}05)} = 4{,}3$

Vemos que nenhum dos fatores influe sobre o rendimento. A soma dos quadrados entre blocos é uma função complexa; ela é usada para obter a soma dos quadrados do erro mas, quando as outras variáveis têm um efeito significativo, não se pode ter certeza de sua influência por meio de um teste F. Por esse motivo os valores correspondentes estão entre parentesis. O desvio padrão da diferença entre as médias ajustadas de 2 tratamentos é dada pela fórmula:

$$s_{dif} = s_r \sqrt{\frac{2k(t-1)}{N(k-1)}} = \sqrt{8{,}05 \times 0{,}85} = 2{,}62$$

onde s_R^2 é o quadrado médio do resíduo.

Procura-se numa tabela o valor de t para p = 0,05 e 6 graus de liberdade (número de graus de liberdade do resíduo). Acha-se t = 2,45.

Para que a diferença entre as médias ajustadas seja significativa é preciso que duas médias consideradas difiram entre si, de pelo menos:

$$2{,}45 \times 2{,}62 = 6{,}42$$

VI.6 – Valor Aproximado de Resultado Perdidos

Pode acontecer que durante um experimento se perca um ou mais resultados por um acidente qualquer. É possível, principalmente no caso da falta de um só resultado, substituí-lo por um valor aproximado, quando o plano experimental foi organizado de um modo correto. O grau de confiança atribuído a este resultado será evidentemente menor que aquele que se atribuiria se fosse obtido experimentalmente.

Suponhamos que no exercício 40 tenha sido perdido o resultado correspondente à coluna E e linha T_5 (10,2) que na Tabela VI.8, depois da transformação de variável, aparece com o valor 2. Este resultado perdido, por hipótese poderá ser substituído por um valor calculado pela fórmula:

$$y = \frac{rR + cC - S}{(r-1)\,(c-1)}$$

onde
- r é o número de linhas (8)
- c é o número de colunas (5)
- R é a soma da linha onde falta o resultado: -7 - 2 = -9
- C é a soma da coluna onde falta o resultado: -3 - 2 = -5
- S é o total: -66 - 2 = -68

Temos então:

$$y = \frac{8(-9) + 5(-5) + 68}{(8-1)\,(5-1)} = \frac{-97 + 68}{28} = -\frac{29}{28} \cong -1$$

Na análise da variância, o resíduo perde 1 grau de liberdade. Quando o problema se trata de um quadrado latino, a fórmula é:

$$y = \frac{r(R+C+T) - 2S}{(r-1)\,(r-2)}$$

onde:

R, C e S têm o mesmo significado anterior
r é o número de linhas (= colunas = tratamento)
T é a soma de resultados do fator onde falta a observação y

Por exemplo, no exercício 40 vamos admitir que o resultado correspondente à coluna II e à linha "2º dia" tenha sido perdido. Por transformação de variável este resultado, no problema visto era 2. Poderá ser calculado por:

$$y = \frac{4(0+1+9) - 30}{(4-1)\,(4-2)} = \frac{10}{6} = 1{,}67 \cong 2$$

O número de graus de liberdade do resíduo será 5 em vez de 6.

Quando se perdem 2 resultados, o processo é um pouco mais complicado. Se, além do resultado anterior, também tivesse sido perdido o resultado da coluna IV e linha "3º dia", deveria-se tomar para y, a média dos outros 3 resultados da mesma concentração, a saber:

$$y'_1 = \frac{13}{3} \cong 4$$

Introduzindo este valor provisório y'_1 no conjunto dos outros, calcula-se y'_2:

$$y'_2 = \frac{4(7+6+(-1)) - 2 \times 16}{(4-1)\,(4-2)} = \frac{48-32}{6} = 2{,}7 \cong 3$$

Introduzindo o valor aproximado $y'_2 = 3$, recalcula-se o valor de y'_1 para substituir o valor hipotético achado e temos:

$$y''_1 = \frac{4(0+1+13) - 2 \times 15}{6} = \frac{26}{6} = 4{,}333 \cong 4$$

O valor de y''_1 é usado para calcular um novo valor y''_2 e assim por diante até que os valores de y_1 e y_2 achados permaneçam constantes.

CAPÍTULO VII

ESTATÍSTICA APLICADA A DUAS VARIÁVEIS CORRELACIONADAS

VII.1 — Correlação

Quando, num certo tipo de experiências, estão envolvidas duas variáveis, deve-se verificar se elas estão correlacionadas, isto é, se elas não são independentes. Define-se o coeficiente de correlação ρ das variáveis x e y como sendo:

$$\rho = \frac{\operatorname{cov}(x,y)}{\sigma_x\,\sigma_y} \qquad \text{(VII.1)}$$

Se x e y são independentes, a cov (x,y) = 0 e portanto $\rho = 0$.

Se as variáveis x e y estão ligadas por uma relação linear que pode ser expressa por $y = \alpha + \beta x$, esta mesma relação é válida para as médias de x e y, isto é, $\mu_y = \alpha + \beta \mu_x$. Neste caso, a cov (x,y) é:

$$
\begin{aligned}
\text{cov}(x,y) &= E\,[(x-\mu_x)\,(y-\mu_y)] \\
&= E\,[(x-\mu_x)\,(\alpha+\beta x-\alpha-\beta\mu_x)] \\
&= E\,[\beta\,(x-\mu_x)^2] \\
&= \beta\,\sigma_2^2
\end{aligned}
\qquad \text{(VII.2)}
$$

Derivando $y = \alpha + \beta x$ e elevando ao quadrado temos que:

$$
\begin{aligned}
\sigma_y^2 &= \beta^2\,\sigma_x^2 \\
\sigma_y &= |\beta|\,\sigma_x
\end{aligned}
\qquad \text{(VII.3)}
$$

Substituindo os valores de (VII.2) e (VII.3) em (VII.1) temos:

$$\rho = \frac{\beta\,\sigma_x^2}{\sigma_n\,|\beta|\,\sigma_x}$$

$$\rho = \frac{\beta}{|\beta|} = \pm 1 \qquad \text{(VII.4)}$$

Pode-se demonstrar que $-1 \leqslant \rho \leqslant 1$.

O valor verdadeiro do coeficiente de correlação ρ não é conhecido, mas sim a sua estimativa r cujo valor é:

$$r = \frac{\dfrac{\sum_{i=1}^{n} (x_i-\bar{x})\,(y_i-\bar{y})}{n-1}}{\sqrt{\dfrac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}\;\dfrac{\sum_{i=1}^{n} (y_i-\bar{y})^2}{n-1}}}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \qquad \text{(VII.5)}$$

Para simplificar, os índices dos somatórios serão omitidos.

Desenvolvendo o produto do denominador e os quadrados do denominador chega-se à expressão:

$$r = \frac{n \Sigma x_i y_i - \Sigma x_i \ \Sigma y_i}{\sqrt{[n \Sigma x_i^2 - (\Sigma x_i)^2][n \Sigma y_i^2 - (\Sigma y_i)^2]}} \qquad \text{(VII.6)}$$

Valores positivos de r indicam que x e y tendem a variar no mesmo sentido, isto é, para valores crescentes de x temos valores crescentes de y ou vice-versa. Valores negativos indicam o contrário. Valores iguais ou próximos de 1 indicam uma relação linear.

VII.2 – Relações Lineares

Vamos considerar um tipo de experiência em que aparecem duas variáveis correlacionadas, isto é, o valor da variável y depende do valor da variável x. Por exemplo, se irradiarmos a massa x_1 de um elemento M vamos obter uma atividade y_1, para massa x_2 obtemos a atividade y_2, . . . e para a massa x_n, a atividade y_n. Temos então dois conjuntos de valores:

$$x_1, x_2, x_3, \ldots x_n$$

$$y_1, y_2, y_3, \ldots y_n$$

Dentro de um certo intervalo, é válida a expressão

$$y_i = a + bx_i \qquad \text{(VII.7)}$$

Neste exemplo, temos que a atividade é proporcional à massa irradiada, a menos de um certo valor <u>a</u> que seria a atividade de fundo (background). Se fizermos um gráfico com os valores de x nas abcissas e os de y nas ordenadas, vamos obter uma reta. Devem ser considerados dois pontos importantes:

1) como temos n equações com duas incógnitas (a e b), deve-se calcular a equação da reta que melhor se adapta aos n pontos experimentais.

2) deve ser feita uma estimativa dos parâmetros a e b da reta.

VII.2.1 – Método dos Mínimos Quadrados

De um modo geral, uma das variáveis ($\underline{x}$) é afetada de erro desprezível diante do erro da outra (y). A variável $\underline{x}$, em pesquisas relacionadas com a química, pode ser massa, concentração, etc., que podem ser obtidas com boa precisão, quando se dispõe de balanças analíticas e balões volumétricos aferidos. A variável y é o resultado experimental e é afetada de erros decorrentes de uma série de fatores incontroláveis que podem ter um efeito mais ou menos pronunciado na precisão do resultado. O conjunto de valores x_i e y_i pode ser representado pela série de equações

$$
\begin{aligned}
y_1 &= a_1 + b_1 x_1 \\
y_2 &= a_2 + b_2 x_2 \\
&\vdots \\
y_n &= a_n + b_n x_n
\end{aligned}
\qquad \text{(VII.8)}
$$

onde os a_i e b_i diferem por causa do erro introduzido na determinação dos y_i. O problema consiste em achar os melhores valores de a e de b que satisfaçam todas as equações de tal maneira que a soma dos quadrados das diferenças entre os pontos experimentais (x_i, y_i) e os pontos correspondentes da reta, (X_i, Y_i) seja mínima, isto é:

$$(y_i - Y_i)^2 = k \text{ (mínimo)} \qquad \text{(VII.9)}$$

Seja:

$$Y = a + bx$$

a equação da reta que satisfaz a condição de um valor mínimo para k. A condição de mínimo para k deve satisfazer as seguintes condições:

$$\frac{\partial}{\partial a} \{ \Sigma (y_i - a - bx_i)^2 \} = 0 \qquad \text{(VII.10)}$$

$$\frac{\partial}{\partial b} \{ \Sigma (y_i - a - bx_i)^2 \} = 0 \qquad \text{(VII.11)}$$

e a segunda derivada nos dois casos deve ser maior que zero.

Derivando (VII.10) vamos ter:

$$- \Sigma y_i + na + b \Sigma x_i = 0 \qquad \text{(VII.12)}$$

e a segunda derivada em relação a a vai dar n que é maior que zero.

Fazendo o mesmo em relação a (VII.11) vamos ter:

$$- \Sigma x_i (y_i - a - bx_i) = 0$$

ou

$$- \Sigma x_i y_i + a \Sigma x_i + b \Sigma x_i^2 = 0 \qquad \text{(VII.13)}$$

e a segunda derivada em relação a b vai dar Σx_i^2 que é maior que zero. Resolvendo as equações (VII.12) e (VII.13) em relação a a e a b vamos ter, omitindo os índices i das variáveis:

$$a = \frac{\Sigma x^2 \, \Sigma y - \Sigma x \, \Sigma xy}{n \, \Sigma x^2 - (\Sigma x)^2} \qquad \text{(VII.14)}$$

$$b = \frac{n \, \Sigma xy - \Sigma x \, \Sigma y}{n \, \Sigma x^2 - (\Sigma x)^2} \qquad \text{(VII.15)}$$

Os valores a e b achados são os parâmetros da regressão e são estimativas dos parâmetros verdadeiros α e β da equação:

$$Y = \alpha + \beta x \qquad \text{(VII.16)}$$

A equação $Y = a + bx$ (VII.17)

achada pelo método dos mínimos quadrados é a que melhor se adapta às n equações (VII.8) se y_1, $y_2 \ldots y_n$ forem resultados independentes, se obedecerem à distribuição normal e se a hipótese da linearidade for verdadeira. A variância que caracteriza a dispersão dos valores y_i achados em relação aos Y_i da reta se define por:

$$s^2_o = \frac{\Sigma (y_i - Y_i)^2}{n - 2} \qquad \text{(VII.18)}$$

Note-se que o número de graus de liberdade de s^2_o é $n - 2$, porque, para achar os valores Y_i, foi preciso calcular a e b, isto é, perderam-se dois graus de liberdade.

Substituindo em (VII.18) os valores de Y_i tirados de (VII.17) e a pelo seu valor de (VII.14) chega-se ao resultado:

$$(n-2)s_o^2 = \Sigma y^2 - \bar{y}\,\Sigma y - b(\Sigma xy - \bar{x}\,\Sigma y) \qquad \text{(VII.19)}$$

A imprecisão que afeta os parâmetros a e b pode ser calculada aplicando a lei de propagação de erros. Temos:

$$s_a^2 = \left(\frac{\partial a}{\partial y_1}\right)^2 s_1^2 + \left(\frac{\partial a}{\partial y_2}\right)^2 s_2^2 + \ldots + \left(\frac{\partial a}{\partial y_n}\right)^2 s_n^2$$

$$s_b^2 = \left(\frac{\partial b}{\partial y_1}\right)^2 s_1^2 + \left(\frac{\partial b}{\partial y_2}\right)^2 s_2^2 + \ldots + \left(\frac{\partial b}{\partial y_n}\right)^2 s_n^2$$

Por hipótese, os pontos são obtidos com a mesma precisão, isto é:

$$s_1^2 = s_2^2 = \ldots = s_n^2 = s_o^2$$

Assim, as equações se tornam:

$$s_a^2 = \frac{s_o^2\,\Sigma x^2}{n\,\Sigma x^2 - (\Sigma x)^2} \qquad \text{(VII.20)}$$

$$s_b^2 = \frac{n\,s_o^2}{n\,\Sigma x^2 - (\Sigma x)^2} \qquad \text{(VII.21)}$$

Se dividirmos o numerador e denominador de (VII.21) por n, vamos ter:

$$s_b^2 = \frac{s_o^2}{\Sigma (x-\bar{x})^2} \qquad \text{(VII.22)}$$

Este resultado indica que a determinação do coeficiente angular b é tanto mais precisa quanto maior for o intervalo entre os pontos x_1 e x_n, ao longo do eixo dos x. Por exemplo, se quisermos fazer medidas cujos resultados estarão num intervalo definido por y_A e y_B, a curva padrão deve ser construída de tal maneira que a x_1 corresponda $y_1 < y_A$ e a x_n corresponda $y_n > y_B$

VII.3 – Verificação da Hipótese da Linearidade

Sempre que se desenvolve um novo método experimental é preciso verificar se os resultados são lineares ou se existe outra curva que se adapte melhor aos pontos achados.

Existem duas maneiras de fazer esta verificação: a primeira consiste em fazer várias determinações paralelas de cada ponto e depois comparar a variância do método com a variância devida à regressão e a segunda baseia-se na análise da variância, desdobrando a variância total em duas partes: uma devida à regressão e a outra chamada residual.

VII.3.1 — Teste de Linearidade por Determinações Paralelas

Suponhamos que temos m padrões, a partir dos quais vai ser construída uma curva padrão para uma certa medida. Em vez de obter um só resultado para cada padrão, o teste exige que se obtenham n resultados com os quais se deve calcular o valor médio $\bar{y}_i$ e a respectiva variância. Temos assim m valores $\bar{y}_i$ e m variâncias que devem ser testadas quanto à homogeneidade pelos critérios de Bartlett ou de Cochrane. Por hipótese, elas devem ser iguais, porque o método usado é o mesmo para todos. Se houver alguma variância diferente, repetem-se as medidas correspondentes a esse ponto ou simplesmente elimina-se esse ponto. Suponhamos que as m variâncias sejam estatisticamente iguais. Calcula-se a variância total devida ao método pela fórmula:

$$s_y^2 = \frac{\sum_i \sum_j (y_{ij} - \bar{y})^2}{m(n-1)} \qquad (j = 1 \ldots n;\ i = 1 \ldots m) \qquad \text{(VII.23)}$$

A variância que corresponde à dispersão dos valores médios $\bar{y}_i$ em relação à linha de regressão é calculada pela fórmula

$$s_R^2 = \frac{n}{(m-2)} \sum_i (\bar{y}_i - Y_i)^2 \qquad \text{(VII.24)}$$

Existe o fator n, porque foram feitas n determinações cuja média é $\bar{y}$. A hipótese da linearidade se testa pela relação:

$$F = \frac{s_R^2}{s_y^2} \qquad \text{(VII.25)}$$

onde s_R^2 tem $(m - 2)$ graus de liberdade e s_y^2 tem $m(n - 1)$ graus de liberdade. Procura-se na Tabela de F o valor teórico, no nível de significância desejado. O teste é monocaudal, porque as hipóteses são:

$$H_0 \ : \ s_R^2 = s_y^2$$

$$H_1 \ : \ s_R^2 > s_y^2$$

No caso de ser aceita a hipótese da linearidade, s_R^2 e s_y^2 são estatisticamente iguais e ambas são estimativas de σ_o^2. Nessas condições, s_R^2 e s_y^2 podem ser combinadas para se obter uma melhor estimativa de σ_o^2 que é:

$$\bar{s}_o^2 = \frac{m(n-1)\, s_x^2 + (m-2)\, s_R^2}{m\,n-2} \qquad \text{(VII.26)}$$

com $f = mn - 2$ graus de liberdade.

VII.3.2 – Comparação dos Parâmetros a e b da Reta com Valores Esperados

Quando se testa um método de medidas, é costume usar padrões, isto é, conhece-se "a priori" o resultado da medida. Pode-se então construir uma curva, colocando nas abcissas os resultados esperados e nas ordenadas os resultados experimentais. Nestas condições, a equação da reta deve ser $y = x$, isto é, devemos ter $a = 0$ e $b = 1$.

1º Método:

Calculam-se os valores de a, b, s_o, s_b e s_a pelas equações (VII.14), (VII.15), (VII.19), (VII.20) e (VII.21) respectivamente.

As hipóteses formuladas no caso são:

$$H_0 : a = 0$$
$$H_1 : a \neq 0$$
para o parâmetro a

$$H_0 : b = 1$$
$$H_1 : b \neq 1$$
para o parâmetro b

Aplica-se o teste t nos dois casos:

$$t_a = \left| \frac{a - 0}{s_a} \right| \qquad \text{(VII.27)}$$

$$t_b = \left| \frac{b - 1}{s_b} \right| \qquad \text{(VII.28)}$$

O número de graus de liberdade de $\underline{t}$ é o correspondente a s_a e s_b. A aceitação dos valores de t experimentais indica linearidade e ausência de erro sistemático. Se t_a experimental for maior que t tabelado, conclui-se que $a \neq 0$ e isto indica que há um erro sistemático que deve ser subtraído dos resultados (prova em branco). Se t_b for maior que o $\underline{t}$ tabelado, $b \neq 1$ e deve-se admitir a existência de um erro sistemático que depende do padrão usado.

EXERCÍCIO 42

Foram preparadas soluções de urânio com 10, 20, 30, 40, 50 mg/ml. Estas soluções foram analisadas quanto ao teor de urânio para testar um método analítico. Os resultados obtidos estão na Tabela VII.1.

Achar num nível de significância 0,05 se existe algum erro sistemático no método. Para facilitar a resolução do problema, dispõem-se os valores conforme a Tabela (VII.2).

Tabela VII.1

Valores Achados em Análises de Padrões de Urânio

Amostras	U (mg/ml)	
	Colocado	Achado
1	10	10,2
2	20	19,6
3	30	31,1
4	40	40,5
5	50	49,7

Tabela VII.2

Disposição dos Resultados e Cálculos Necessários para a Resolução do Exercício 42

	x	y	x^2	xy	y^2
1	10	10,2	100	102	104,04
2	20	19,6	400	392	384,16
3	30	31,1	900	933	967,21
4	40	40,5	1600	1620	1640,25
5	50	49,7	2500	2485	2470,09
	150	151,1	5500	5532	5565,75

$$\bar{x} = 30{,}0$$

$$\bar{y} = 30{,}22$$

$$a = \frac{151{,}1\ (5500) - 150\ (5532)}{5(5500) - (150)^2} = 0{,}25$$

$$b = \frac{5\ (5532) - 150\ (151{,}1)}{5(5500) - (150)^2} = 0{,}999$$

$$s_o^2 = \frac{5565{,}75 - (30{,}22)\ (151{,}1) - 0{,}999\ (5532 - 30{,}0)\ (151{,}1)}{3} = 0{,}5023$$

$$s_o = 0{,}71$$

$$s_a^2 = \frac{0{,}5023\ (150)}{5(5500) - (150)^2} = 0{,}552566$$

$$s_a = 0{,}74$$

$$s_b^2 = \frac{5\ (0{,}5023)}{5\ (5500) - (150)^2} = 0{,}0005023$$

$$s_b = 0{,}022$$

$$t_a = \frac{0{,}25}{0{,}7333} = 0{,}34$$

$$t_b = \frac{0{,}999 - 1}{0{,}022} = 0{,}04$$

O valor tabelado de $t_{0,05(3)} = 3{,}18$

Aceita-se a hipótese de que não há erro sistemático.

Segundo Nalimov, este método deve ser usado com cuidado, porque há casos em que o teste t pode dar não significativo, quando na realidade as hipóteses devem ser rejeitadas. É mais recomendável trabalhar com as diferenças d_i entre os x_i e y_i, para testar se $a = 0$. Neste caso, como já foi visto, as hipóteses seriam:

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} \neq 0$$

É conveniente, antes de aplicar o teste t, verificar a homogeneidade dos d_i pelo critério do r_{max} e rejeitar os valores estatisticamente diferentes.

2º Método

Consiste em fazer duas medidas de cada padrão, usando massas diferentes. Vamos considerar o padrão p_1 do qual são usadas as massas M_1 e M_2 que devem dar as respostas x_1 e x_2 respectivamente. Pelo método usado as respostas foram y_1 e y_2. Temos então duas equações:

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \end{aligned} \qquad (VII.29)$$

Como o padrão é o mesmo, é *válida* a relação:

$$\frac{x_1}{M_1} = \frac{x_2}{M_2} \tag{VII.30}$$

Substituindo em (VII.30) os valores de x_1 e x_2 de (VII.29), temos:

$$\frac{y_1 - a}{M_1} = \frac{y_2 - a}{M_2} \tag{VII.31}$$

Tirando o valor de $\underline{a}$ da equação (VII.31) vamos obter:

$$a = \frac{y_2 - (M_2/M_1)\, y_1}{1 - (M_2/M_1)} \tag{VII.32}$$

Fazendo o mesmo para n padrões, obtém-se n valores de $\underline{a}$ a partir dos quais se acha $\bar{a}$ e s_a. Calcula-se o valor de t para o valor médio $\bar{a}$ que é:

$$t = \frac{\bar{a} - 0}{s_a}\, n \tag{VII.33}$$

O valor de t achado deve ser comparado com o valor de t tabelado com (n − 1) graus de liberdade.

Para a estimativa de b = 1, também se fazem duas medidas paralelas do mesmo padrão. Usam-se massas iguais para as duas medidas, sendo que numa delas adiciona-se uma quantidade z conhecida daquilo que se vai determinar. Vamos ter então:

$$y_1 = a + bx_1$$

$$y_2 = a + b\,(x_1 + z)$$

de onde se obtém que:

$$b = \frac{y_2 - y_1}{z} \tag{VII.34}$$

Usando este procedimento para cada padrão, vamos ter n valores de b. Calcula-se o valor médio $\bar{b}$ e a variância s_b^2 e aplica-se o teste t.

$$t = \frac{|\,\bar{b} - 1\,|}{s_b} \sqrt{n} \tag{VII.35}$$

Do mesmo modo que foi feito para a média $\bar{x}$ de n resultados, pode-se estabelecer os limites de confiança para $\bar{b}$ e para $\bar{a}$. Deve-se notar que esses métodos para estimar os limites não são completamente corretos, porque os dois parâmetros a e b estão correlacionados. Sabe-se que quaisquer que sejam os valores de a ou de b, a reta passa obrigatoriamente pelo ponto $(\bar{x}, \bar{y})$. Uma variação de a dentro do intervalo $a \pm t\, s_a$, provoca então uma variação de b, conforme se pode ver na Figura 7.1. Em análises estatísticas rigorosas, os limites de confiança para a variação de a não podem ser feitos separadamente dos limites de confiança para a variação de b. Deve-se por isso construir uma região de confiança para os valores possíveis de a e b ao mesmo tempo, isto é, os limites de confiança dos dois parâmetros devem ser compatíveis entre si.

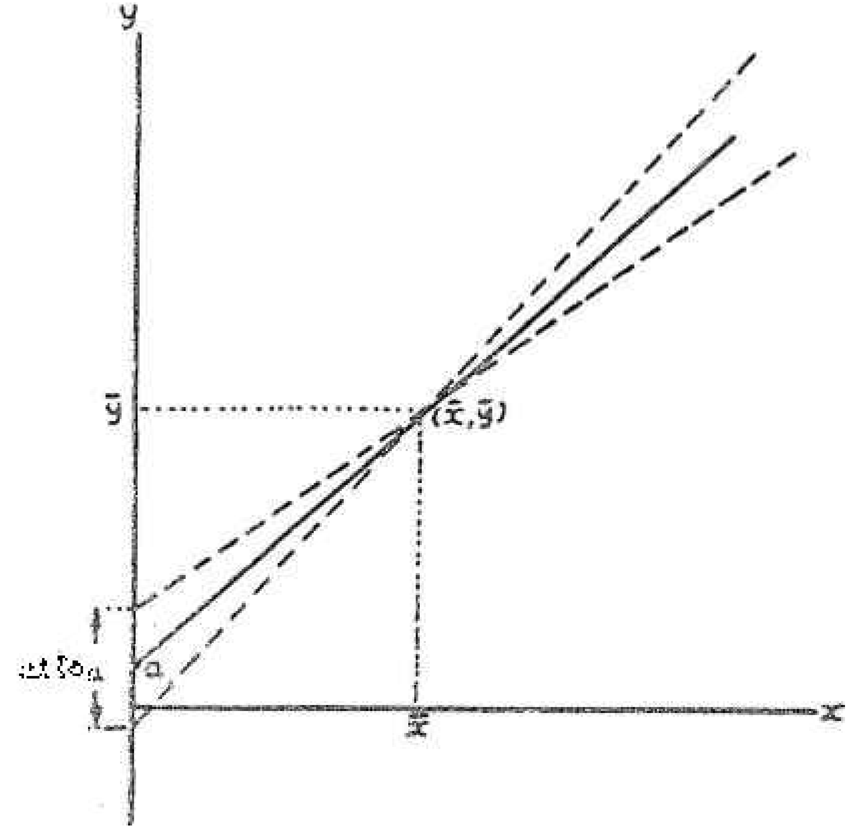

Figura 7.1 – Variação de b com a Variação de a no seu Intervalo de Confiança

EXERCÍCIO 43

Foi feita uma curva padrão para testar um novo método analítico para zircônio. Foram usados padrões de zircônio e pelo método foram achados os resultados que estão na Tabela (VII.3). Na Tabela também se encontram os valores de x^2, y^2 e xy, necessários para o cálculo. Deseja-se saber num nível de significância de 0,05 se o método é afetado de erro sistemático.

Tabela VII.3

Resultados de Análises de Padrões de Zircônio

Colocado $x(\mu g)$	Achado $y(\mu g)$	x^2	y^2	xy
2,0	2,1	4,00	4,41	4,20
3,4	3,6	11,56	12,96	12,24
4,6	4,9	21,16	24,01	22,54
5,5	5,6	30,25	31,36	30,80
6,7	6,9	44,89	47,61	46,23
8,0	8,3	64,00	68,89	66,40
9,8	9,9	96,04	98,01	97,02
10,7	11,0	114,49	121,00	117,70
11,5	11,2	132,25	125,44	128,80
12,8	13,0	163,84	169,00	166,40
14,0	14,4	196,00	207,36	201,60
15,3	15,5	234,09	240,25	237,15
16,5	16,6	272,25	275,56	273,90
18,2	18,5	331,24	342,25	336,70
20,0	20,2	400,00	408,04	404,00
159,00	161,7	2116,06	2176,15	2145,68

$\bar{x} = 10,6 \qquad \bar{y} = 10,78$

$$a = \frac{2116,06\ (161,7) - 159\ (2145,68)}{15\ (2116,06) - (159)^2} = 0,1554$$

$$b = \frac{15\ (2145,68) - 159\ (161,7)}{15\ (2116,06) - (159)^2} = 1,0023$$

$$s_0^2 = \frac{1}{13}\ [2176,15 - 10,78\ (161,7) - 1,0023\ (2145,68 - 10,6\ (161,7))] = 0,028552$$

$$s_0 = 0,169$$

$$s_a^2 = \frac{0,028552 \times 2116,06}{15 \times 2116,06 - (159)^2} = 0,009352$$

$$s_a = 0,097$$

$$s_b^2 = \frac{0{,}028552 \times 15}{15 \times 211{,}06 - (159)^2} = 0{,}0000662982$$

$$s_b = 0{,}00814$$

teste t para verificar se a = 0:

$$t_a = \frac{0{,}1554}{0{,}097} = 1{,}60$$

teste t para verificar se b = 1

$$t_b = \frac{1.0023 - 1{,}0000}{0{,}00814} = 0{,}28$$

$$t_{0,05(13)} = 2{,}16$$

Como os valores de t achados são menores do que 2,16 aceita-se a = 0 e b = 1, isto é, o método não apresenta erro sistemático ao nível de significância adotado.

Método das Diferenças

O mesmo problema pode ser resolvido a partir das diferenças entre os valores reais (x) e os valores achados (y).

Estas diferenças (d_i) estão na Tabela VII.4, onde foram incluídos os valores de $d_i - \bar{d}$ e $(d_i - \bar{d})^2$.

O teste de hipóteses é:

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} \neq 0$$

$$s_d^2 = \frac{\Sigma (d_i - \bar{d})^2}{n - 1} = \frac{0{,}3640}{14} = 0{,}0260$$

$$s_d = 0{,}1612$$

$$t = \frac{0,18}{0,1612}\sqrt{15} = 4,32$$

$$t_{0,05(14)} = 2,15$$

Tabela VII.4

Diferenças entre os Valores Reais (x) e Valores Achados (y) nas Análises de Zircônio

d	$\lvert d - \bar{d} \rvert$	$(d - \bar{d})^2$
0,1	0,08	0,0064
0,2	0,02	0,0004
0,3	0,12	0,0144
0,1	0,08	0,0064
0,2	0,02	0,0004
0,3	0,12	0,0144
0,1	0,08	0,0064
0,3	0,12	0,0144
-0,3	0,48	0,2304
0,2	0,02	0,0004
0,4	0,22	0,0484
0,2	0,02	0,0004
0,1	0,08	0,0064
0,3	0,12	0,0144
0,2	0,02	0,0004
2,7		0,3640

$n = 15$
$\bar{d} = 0,18$

Por este teste não se pode aceitar $\bar{d} = 0$, o que é uma contradição ao teste anterior que permitiu aceitar a = 0.

Se analisarmos os valores de d na Tabela (VII.4), verificamos que só um dos valores é negativo. Aplicando o teste de r_{min} a este valor temos:

$$r_{min} = \left| \frac{-0,3 - 0,18}{0,1612} \right| = \frac{0,48}{0,1612} = 2,98$$

$$r_{min\ (0,05(13))} = 2,49$$

Como o valor de r_{min} achado é maior que o tabelado, o resultado de 11,2 μg de zircônio, quando na realidade o padrão tem 11,5 μg, parece ter um erro acidental e deve ser excluído. Repetindo os cálculos para os 14 resultados restantes vamos obter:

$$\Sigma x = 147{,}5 \qquad \Sigma y = 150{,}5$$

$$\bar{x} = 10{,}536 \qquad \bar{y} = 10{,}75$$

$$\Sigma x^2 = 1983{,}81 \qquad \Sigma y^2 = 2050{,}71$$

$$\Sigma xy = 2016{,}88$$

$$a = \frac{1983{,}81\ (150{,}5) - 147{,}5\ (2016{,}88)}{14\ (1983{,}81) - (147{,}5)^2} = 0{,}1784$$

$$b = \frac{14\ (2016{,}88) - 147{,}5\ (150{,}5)}{14\ (1983{,}81) - (147{,}5)^2} = 1{,}0034$$

$$s_o^2 = \frac{1}{12}\left[2050{,}71 - 10{,}75\ (150{,}5) - 1{,}0034(2016{,}88 - \frac{147{,}5}{14}\ (150{,}5))\right] = 0{,}009477$$

$$s_o = 0{,}0973$$

$$s_a^2 = \frac{0{,}009477\ (1983{,}81)}{14\ (1983{,}81) - (147{,}5)^2} = 0{,}003124$$

$$s_a = 0{,}056$$

$$s_b^2 = \frac{0{,}009477\ (14)}{14\ (1983{,}81) - (147{,}5)^2} = 0{,}00002205$$

$$s_b = 0{,}0047$$

$$t_a = \frac{0{,}1784}{0{,}056} = 3{,}18$$

$$t_b = \frac{1{,}0034 - 1{,}0000}{0{,}0047} = 0{,}72$$

$$t_{0,05(12)} = 2{,}18$$

Como o teste para a = 0 deu um valor t superior ao valor tabelado, não se aceita a hipótese de a = 0, isto é, o método apresenta um erro sistemático, que pode ser corrigido, deduzindo dos valores achados o valor de a ou o de uma prova em branco.

Os intervalos de confiança para os valores de a e de b são dados por:

$$a = 0{,}1784 \pm t\, s_a = 0{,}1784 \pm 0{,}1221$$

$$0{,}0563 \leqslant a \leqslant 0{,}3004$$

$$b = 1{,}0034 \pm t\, s_b = 1{,}0034 \pm 0{,}0102$$

$$0{,}9932 \leqslant b \leqslant 1{,}0136$$

Apesar desses intervalos de confiança serem válidos isoladamente, como os parâmetros a e b não são independentes, o valor de a não pode variar sem afetar o valor de b e vice-versa. A variação de a dentro de seu intervalo de confiança pode trazer como conseqüência valores de b que caem fora do intervalo de confiança de b.

EXERCÍCIO 44

A média de três provas em branco para as análises de zircônio do exercício anterior deu 0,24. Este valor está dentro do intervalo de variação de a, mas deseja-se saber se ele é compatível com b = 1 num nível de significância 0,05.

A região de confiança para a e b simultaneamente é limitada por uma elipse cuja equação é:

$$n(a' - a)^2 + 2(\Sigma x)\,(a' - a)\,(b' - b) + (\Sigma x^2)\,(b' - b)^2 = 2F\, s_o^2 \qquad \text{(VII.36)}$$

onde a e b são os valores achados pelo método dos mínimos quadrados. No nosso caso queremos testar a' = 0,24. Então temos:

$$14(0{,}24 - 0{,}1784)^2 + 2(147{,}5)\,(0{,}24 - 0{,}1784)\,(b' - 1{,}0034) + 1983{,}81\,(b' - 1{,}0034)^2 = 2\,F(0{,}009477)$$

O valor de F é procurado na Tabela e tem 2 e (n − 2) graus de liberdade. No nosso caso $F_{0{,}05(2{,}12)} = 3{,}9$. Resolvendo a expressão acima acha-se

$$1983{,}81\, b'^2 - 1962{,}937908\, b' + 1979{,}068259 = 0$$

Resultados:

$$b' = \begin{cases} 1{,}004 \\ 0{,}993 \end{cases}$$

O intervalo de variação de b' para $a = 0{,}24$ fica sendo então $0{,}993 < b < 1{,}004$, pouco menor que o achado anteriormente. Conclui-se que o valor 0,24 pode ser usado para a prova em branco.

Podemos também querer saber qual o intervalo de variação de a quando assumimos $b = 1$. Neste caso vamos ter:

$$14(a' - 0{,}1784)^2 + (146{,}5)(a' - 0{,}1784)(1{,}000 - 1{,}0034) + 1983{,}81(1{,}000 - 1{,}0034)^2 = 2(3{,}9)(0{,}009477)$$

Resultados: $a' = \begin{cases} 0{,}284 \\ 0{,}144 \end{cases}$

O intervalo de variação de a achado é suficientemente largo, por isso é válido admitir que $b = 1$.

Graficamente (Figura 7.2), o problema pode ser resolvido pela construção da elipse usando os seguintes parâmetros:

$$K^2 = 2 F s_o^2 = 0{,}0739206$$

$$W = \frac{\sqrt{n \Sigma x^2}}{x} = 1{,}1298$$

$$L_b = s_b \sqrt{2F} = 0{,}0131$$

$$L_a = s_a \sqrt{2F} = 0{,}1564$$

$$d_b = K \sqrt{\frac{2W}{(1 + W) \Sigma x^2}} = 0{,}0063$$

$$d_a = \frac{L_a}{L_b} d_b = 0{,}0752$$

Escolhe-se a escala para as coordenadas e marca-se o centro da elipse no ponto (a,b). Por este ponto traçam-se retas paralelas aos eixos. Sobre estas retas, a partir do centro, a uma distância d_b e d_a marcam-se os pontos E, E' e D, D' respectivamente. Sobre o eixo das abcissas, marcam-se os pontos que correspondem aos limites do intervalo de variação de a e por estes pontos traçam-se paralelas ao eixo das ordenadas. Do mesmo modo, no eixo das ordenadas marcam-se os pontos limites da variação de b e traçam-se paralelas ao eixo das abscissas. Traçam-se as retas ED e E'D' até o perímetro do retângulo onde se marcam os pontos R, T, R' e T'. O hexágono RNT R'N'T' limita a elipse. Pelo ponto $a = 0$ e $b = 1$ traça-se uma paralela ao eixo das abcissas. A parte desta reta limitada pela elipse dá a variação de a para $b = 1$ e vemos que é o intervalo achado pela equação (VII.36). Do mesmo modo, pelo ponto $a = 0{,}24$ traça-se uma paralela ao eixo das ordenadas. A variação de b para a = 0,24 é dada pela parte desta reta limitada pela elipse. Vemos, também neste caso, que os valores confirmam os achados pela equação (VII.36). Esta equação pode ser usada também para um teste simultâneo de a = 0 e b = 1. Para isto, devem ser introduzidos na equação (VII.36) os valores a' = 0 e b' = 1. Procura-se o valor de F na Tabela e compara-se com o valor de F experimental. Se o F calculado for maior que o F tabelado, rejeita-se a hipótese de a = 0 e b = 1 simultaneamente.

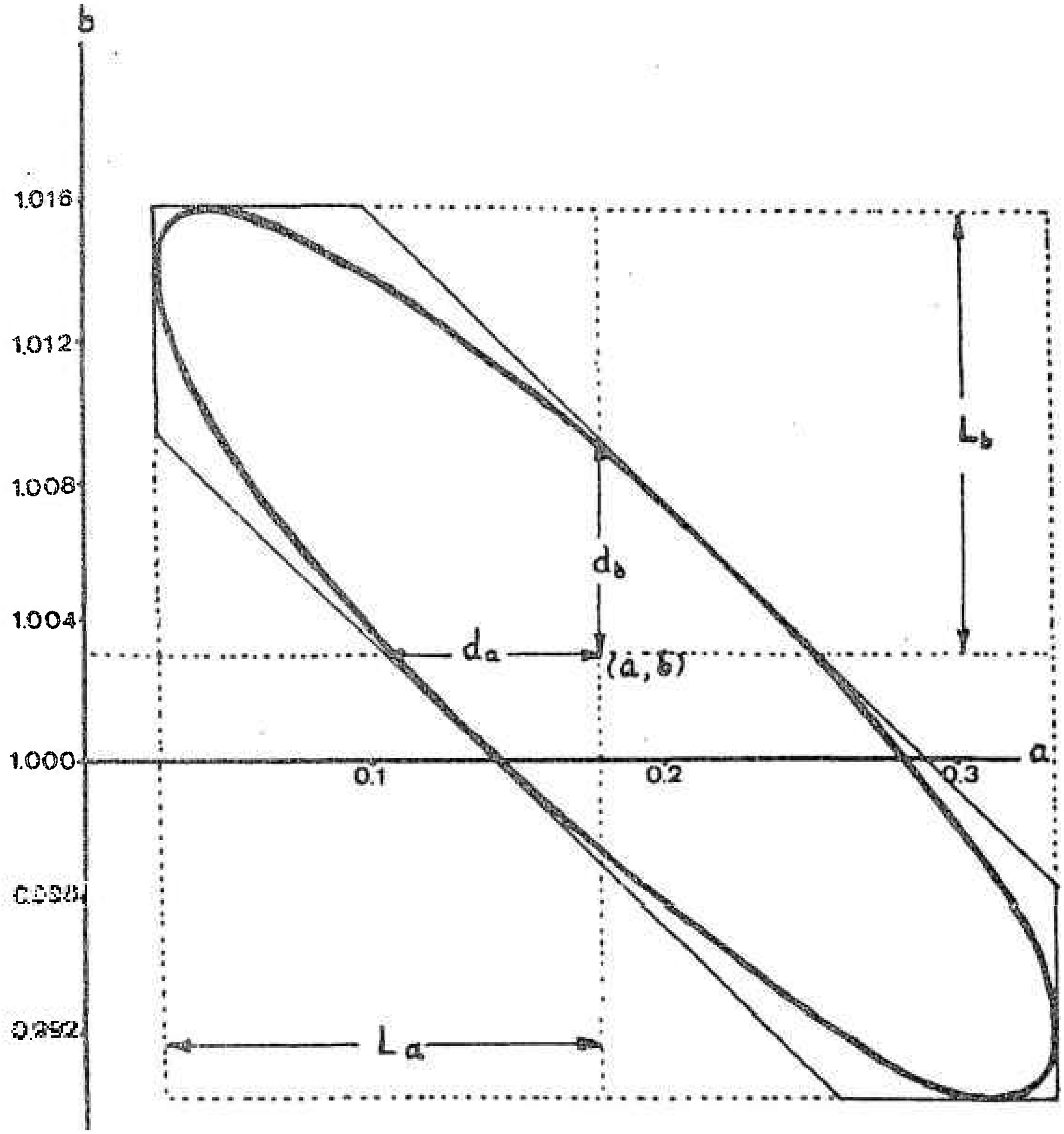

Figura 7.2 — Região de Variação de a e b

EXERCÍCIO 45

Foi traçada uma curva padrão para um método aplicado a análises de tório. Os valores achados (y) e os valores esperados (x) estão na Tabela (VII.5). Verificar se as hipóteses a = 0 e b = 1 podem ser aceitas simultaneamente, num nível de confiança de 0,95.

Tabela VII.5

Resultados Obtidos em Análises de Tório

Colocado $x(\mu g)$	Achado $y(\mu g)$	xy	x^2	y^2
0,20	0,22	0,0440	0,0400	0,0484
1,30	1,26	1,6380	1,6900	1,5876
2,50	2,53	6,3250	6,2500	6,4009
3,60	3,65	13,1400	12,9600	13,3225
4,50	4,44	19,9800	20,2500	19,7136
5,80	5,78	33,5240	33,6400	33,4084
17,90	17,88	74,6510	74,8300	74,4814

$$\bar{x} = 2{,}9833$$

$$\bar{y} = 2{,}98$$

$$a = \frac{74{,}83\ (17{,}88) - 17{,}9\ (74{,}651)}{6\ (74{,}83) - (17{,}9)^2} = 0{,}0133$$

$$b = \frac{6\ (74{,}651) - 17{,}9\ (17{,}88)}{6\ (74{,}83) - (17{,}9)^2} = 0{,}9944$$

$$s_o^2 = \frac{1}{4}\ [74{,}4814 - 17{,}88\ (2{,}98) - 0{,}9944\ (74{,}651 - \frac{(17{,}88)\ (17{,}9)}{6})]$$

$$s_o^2 = 0{,}0023325$$

$$s_o = 0{,}0483$$

Pela fórmula (VII.36) temos:

$$6(0 - 0{,}133)^2 + 2(17{,}9)\ (0 - 0{,}0133)\ (1 - 0{,}9944) + 74{,}83\ (1 - 0{,}9944)^2 = 2F(0{,}0023325)$$

$$F = 0{,}159$$

$$F_{0{,}05(2{,}4)} = 6{,}9$$

Conclui-se que os valores $a = 0$ e $b = 1$ podem ser aceitos simultaneamente.

VII.3.3 — Análise da Variância

A variação total dos resultados (y_i) é dada por $\Sigma(y_i - \bar{y})^2$ que pode ser desdobrada em dois termos: $\Sigma(Y_i - \bar{y})^2$ e $\Sigma(y_i - Y_i)^2$. O primeiro termo é a variância explicada pela regressão e o segundo termo é a variação residual sobre a regressão. A Figura 7.3 permite visualizar melhor o significado desses termos.

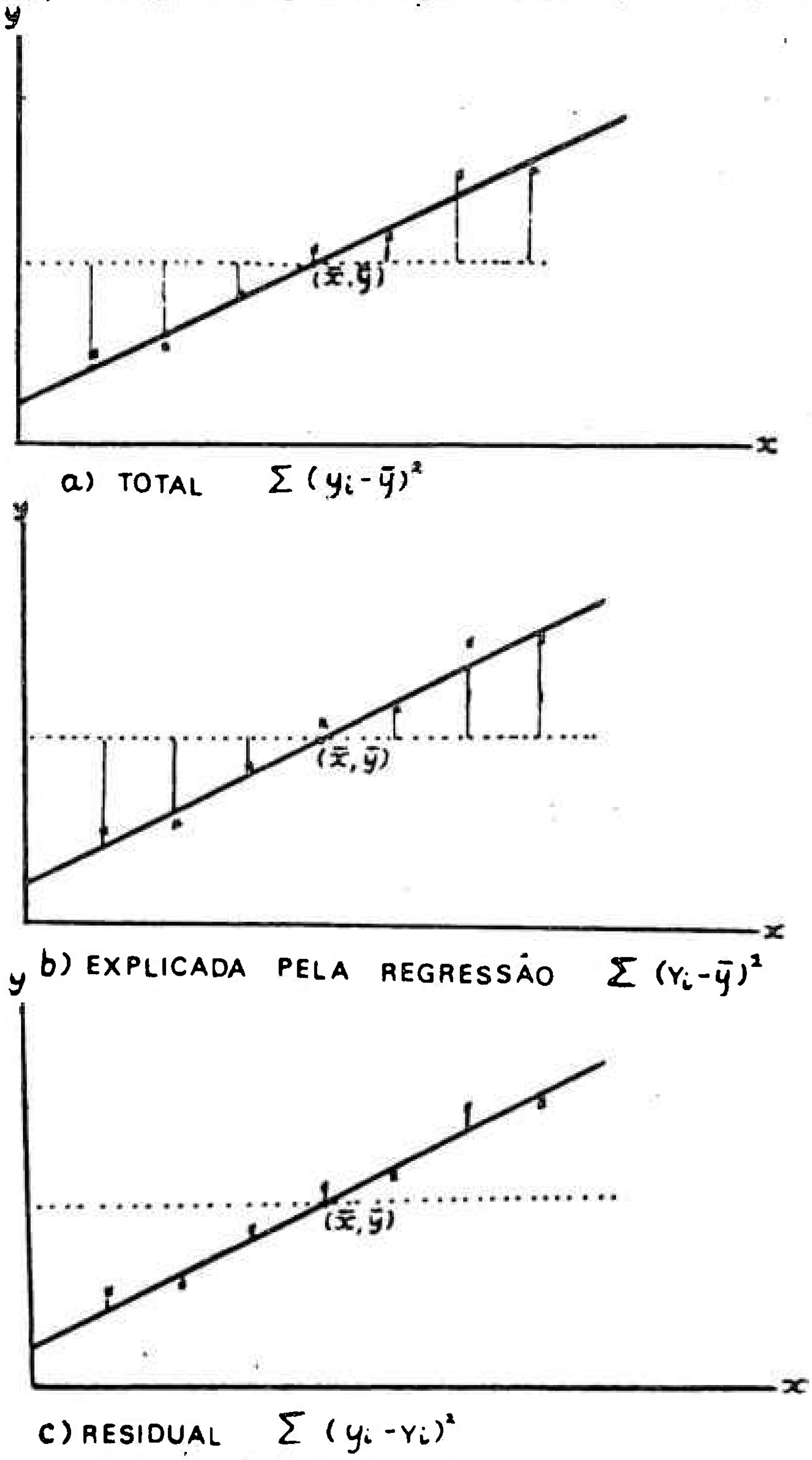

Figura 7.3 — Variação dos Resultados

Como $\Sigma(y_i - \bar{y})^2 / (n - 1)$ é uma estimativa justa de σ_y^2 segue-se que:

$$\frac{\Sigma (y_i - \bar{y})^2}{\sigma_y^2} = \chi_A^2 \qquad \text{(VII.37)}$$

com $(n - 1)$ graus de liberdade.

Podemos ver, pela Figura 7.4 que a relação:

$$\frac{Y_i - \bar{y}}{x_i - \bar{x}} = b$$

é uma definição do coeficiente angular. Sendo assim, a relação

$$(Y_i - \bar{y})^2 = b^2 (x_i - \bar{x})^2$$

é válida e pode ser estendida para todos os pontos. Temos então:

$$\Sigma (Y_i - \bar{y})^2 = b^2 \Sigma (x_i - \bar{x})^2$$

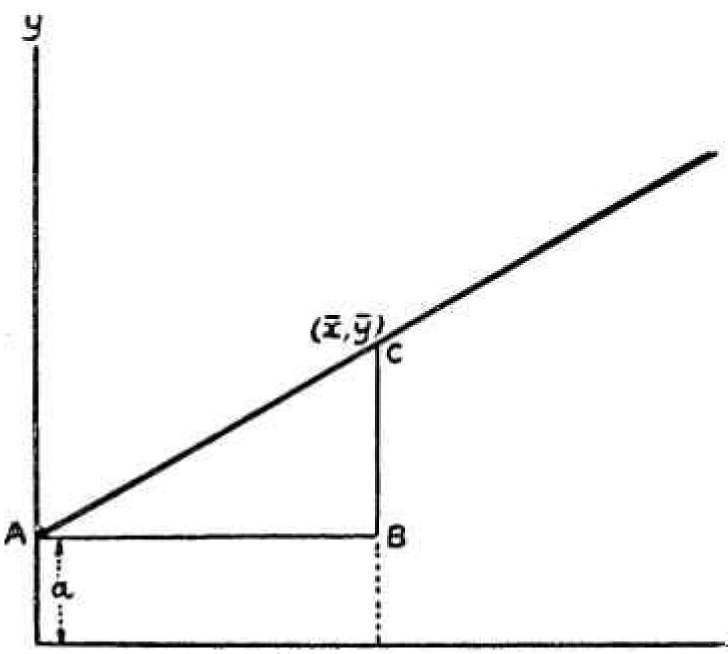

Figura 7.4 – Dependência de a com os Valores de $\bar{x}$ e $\bar{y}$

Quando verificamos a hipótese $\beta = 1$, usamos o teste t pela equação (VII.28) que, num caso mais geral pode ser escrita:

$$t = \frac{b - \beta}{s_b}$$

Vimos também pela equação (VII.22) que:

$$s_b^2 = \frac{s_o^2}{\Sigma (x - \bar{x})^2}$$

onde s_o^2 é estimativa de σ^2. Pode-se então escrever:

$$t = \frac{|b - \beta|}{s_o} \sqrt{\Sigma (x - \bar{x})^2} \qquad \text{(VII.38)}$$

Fazendo $s_o \rightarrow \sigma$, temos $t \rightarrow u$ e a equação se torna:

$$u = \frac{|b - \beta|}{\sigma} \sqrt{\Sigma (x - \bar{x})^2} \qquad \text{(VII.39)}$$

ou

$$\frac{(b - \beta)^2 \; \Sigma (x - \bar{x})^2}{\sigma^2} = u^2 \qquad \text{(VII.40)}$$

Se fizermos a hipótese $\beta = 0$, isto é, não há regressão, temos que:

$$\frac{b^2 \; \Sigma (x - \bar{x})^2}{\sigma^2} = u^2 = \chi_B^2 \qquad \text{(VII.41)}$$

onde χ_B^2 tem um grau de liberdade.

Temos também, por definição, que:

$$s_o^2 = \frac{\Sigma (y_i - Y_i)^2}{n - 2}$$

e sabemos que s_o^2 é uma estimativa de σ^2. Nestas condições temos a relação:

$$\frac{\Sigma (y_i - Y_i)^2}{\sigma^2} = \chi_C^2 \qquad \text{(VII.42)}$$

onde χ^2_C tem (n – 2) graus de liberdade. Os três χ^2 são independentes, porque:

$$\chi^2_A = \chi^2_B + \chi^2_C$$

respectivamente com (n – 1), 1 e (n – 2) graus de liberdade, isto é, obedecem ao teorema de Cochrane. Podem ser então verificadas, pela análise da variância, as hipóteses:

$H_0 : \beta = 0$

$H_1 ; \beta \neq 0$

Neste caso, o quadro da análise da variância é o da Tabela VII.6.

Tabela VII.6

Quadro da Análise da Variância para Ser Aplicado em Testes de Regressão

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Devida à Regressão	$\Sigma(Y_i - y)^2 = b^2 \ \Sigma(x_i - x)^2$	1	$b^2 \ \Sigma(x_i - x)^2$	$F = \dfrac{b^2 \Sigma(x_i - x)^2}{S^2}$
Resíduo	$\Sigma(y_i - Y_i)^2$ (por diferença)	$n - 2$	$\Sigma(y_i - Y_i)^2/(n-2) = S^2$	
Total	$\Sigma(y_i - Y)^2$	$n - 1$		

Se $F_{exp} < F_{p(1,n-2)}$ aceitamos H_0, isto é, não há regressão. A fração da variação de y explicada pela regressão é dada pela expressão:

$$\frac{\text{Soma de Quadrados devida à Regressão}}{\text{Soma de Quadrados Total}} = \frac{b^2 \ \Sigma \ (x_i - \overline{x})^2}{\Sigma \ (y_i - \overline{y})^2} \qquad \text{(VII.43)}$$

Dividindo, na expressão de b dada pela equação (VII.15), o númerador e o denominador por n, e elvando ao quadrado temos:

$$b^2 = \frac{(\Sigma \ xy - \dfrac{\Sigma \ x \ \Sigma \ y}{n})^2}{(\Sigma \ (x - \overline{x})^2)^2} \qquad \text{(VII.44)}$$

Substituindo (VII.44) em (VII.43) vamos ter:

$$\frac{\left(\Sigma\, xy - \dfrac{\Sigma\, x\, \Sigma\, y}{n}\right)^2}{\Sigma\,(y_i - \bar{y})^2\ \ \Sigma\,(x_i - \bar{x})^2} = r^2 .$$

Conclui-se que a fração da variação de y explicada pela regressão é igual ao quadrado do coeficiente de correlação.

VII.4 — Regressão Polinomial

Vamos considerar o caso em que y é um polinômio de grau $\underline{m}$ em x, isto é:

$$y = \beta_o + \beta_1\, x + \beta_2\, x^2 + \ldots + \beta_m x^m$$

cuja estimativa é:

$$Y = b_o + b_1 x + b_2 x^2 + \ldots + b_m x^m$$

Pelo método dos mínimos quadrados, deve-se tornar mínimo o valor de:

$$\sum_{i=1}^{m} (y_i - b_o - b_1 x - b_2 x^2 \ \ldots - b_m x^m)^2$$

Derivando em relação a $b_0, b_1, \ldots b_m$ e igualando a zero vamos ter:

$$\left\{\begin{aligned} \Sigma\, y_i &= m b_0 + b_1\, \Sigma\, x_i + b_2\, \Sigma\, x_i^2 + \ldots + b_m\, \Sigma\, x_i^m \\ \Sigma\, x_i y_i &= b_0\, \Sigma\, x_i + b_1\, \Sigma\, x_i^2 + b_2\, \Sigma\, x_i^3 + \ldots + b_m\, \Sigma\, x_i^{m+1} \\ &\ \ \vdots \\ \Sigma\, x_i^m y_i &= b_0\, \Sigma\, x_i^m + b_1\, \Sigma\, x_i^{m+1} + b_2\, \Sigma\, x_i^{m+2} + \ldots + b_m\, \Sigma\, x_i^{2m} \end{aligned}\right. \qquad \text{(VII.45)}$$

Se a variável x for centrada, isto é, $\Sigma x_i = 0$ (o que se pode obter por mudança de origem, fazendo $\bar{x} = 0$) e se os valores de x forem equi-espaçados, o que ocorre muitas vezes em trabalhos experimentais, as somas das potências impares de x se anulam, isto é, $\Sigma\, x_i^{2k+1} = 0$ $(k = 0, 1, 2, \ldots)$. Deste modo o sistema de equações (VII.45) se simplifica.

VII.4.1 — Determinação do Grau do Polinômio

Se num trabalho experimental temos n pontos, podemos passar um polinômio de grau (n − 1) para ter $y_i - Y_i = 0$ para qualquer valor de i. Isto, porém, não faz sentido, porque y é uma variável aleatória o que quer dizer que deve haver uma certa variação em torno da linha de regressão. Pela análise da variância pode-se estabelecer o grau do polinômio mais conveniente que passa por todos os pontos. No caso da regressão linear, a variação total de y, isto é, $\Sigma(y_i - \bar{y})^2$ foi decomposta em duas parcelas independentes: uma devida à regressão $(\Sigma(Y_i - \bar{y})^2)$ e outra residual $(\Sigma(y_i - Y_i)^2)$. A identidade:

$$\Sigma\,(y_i - \bar{y})^2 = \Sigma\,(y_i - Y_i)^2 + \Sigma\,(Y_i - \bar{y})^2$$

é válida para qualquer Y e portanto para qualquer polinômio.

Se passarmos de um polinômio de grau m para um polinômio de grau m+1, o resíduo diminui e a parte da regressão aumenta, até que, para o grau n – 1, o resíduo é nulo.

O quadro da análise da variância no caso da regressão polinomial é o da Tabela (VII.7).

Tabela VII.7

Análise da Variância para Regressão Polinomial

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F
Polinômio de Grau m	$\Sigma(Y_m - y)^2$	m		
Melhoria	por diferença	1	s_M^2 = Melhoria	$F = \dfrac{S_M^2}{S_R^2}$
Polinômio de Grau m + 1	$\Sigma(Y_{m+1} - y)^2$	(m + 1)		
Resíduo	por diferença	(n – m – 2)	$s_R^2 = \dfrac{\text{Diferença}}{(n - m - 2)}$	
Total		(n – 1)		

As hipóteses são: H_0 ––– Não há melhoria

H_1 ––– Há melhoria

Se $F_{exp} > F_{p,(1,n-m-2)}$ rejeitamos H_0 e concluímos que há melhoria. Se a melhoria for da mesma ordem de grandeza do resíduo, esta melhoria é só aparente, isto é, o polinômio de grau m se adapta tão bem aos pontos quanto o de grau (m+1) e deve-se optar pelo primeiro por ser mais simples.

EXERCÍCIO 46

Num trabalho de recuperação de tório, foi aplicado um método novo e, para verificar se a recuperação é proporcional à massa existente, o método foi aplicado para concentrações conhecidas de tório. Os resultados obtidos estão na Tabela (VII.8) onde x é a massa real de tório e y é a massa do tório recuperado. Verificar num nível de significância de 0,10 se há regressão e qual o polinômio que melhor se adapta aos pontos. Interpretar os resultados.

Tabela VII.8

Resultados de Recuperação de Tório

x (g)	y (g)	xy	x^2	y^2
0,2	0,18	0,036	0,04	0,0324
0,5	0,49	0,245	0,25	0,2401
1,1	1,07	1,177	1,21	1,1449
2,8	2,65	7,420	7,84	7,0225
4,3	4,05	17,415	18,49	16,4025
5,5	5,18	28,490	30,25	26,8324
7,8	7,42	57,876	60,84	55,0564
9,5	8,73	82,935	90,25	76,2129
11,0	10,10	111,100	121,00	102,0100
13,3	12,03	159,999	176,89	144,7209
56,0	51,90	466,693	507,06	429,6750

$\bar{x} = 5{,}60$

$\bar{y} = 5{,}19$

Fonte de variação devida à regressão (SQL):

$$SQL = b^2 \Sigma (x_i - \bar{x})^2 = b^2 \Sigma x_i^2 - \frac{b^2 (\Sigma x_i)^2}{n}$$

$$b = \frac{10\,(466{,}693) - 56{,}0\,(51{,}9)}{10\,(507{,}06) - (56{,}0)^2} = 0{,}9100$$

$$b^2 = 0{,}82814$$

$$SQL = 0{,}82814\,(507{,}06 - \frac{(56)^2}{10}) = 160{,}2120$$

Fonte de variação residual (SQR):

$$SQR = \Sigma (y_i - \bar{y})^2 - SQL$$

$$SQR = \Sigma y_i^2 - \frac{(\Sigma y)^2}{n} - SQL$$

$$SQR = 429,675 - \frac{(51,9)^2}{10} - 160,212 = 0,102$$

O quadro da análise da variância está apresentado na Tabela (VII.9)

Tabela VII.9

Análise da Variância dos Resultados da Tabela VII.8

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Medio	F
Devida 'a Regressão	160,212	1	160,212	$F = \frac{160,212}{0,01275} = 12566$
Resíduo	0,102	8	0,01275	$F_{exp} = F_{0,10(1,8)} = 2,0$
Total	160,314	9		

Como $F_{exp} > F_{tab}$, conclui-se que há regressão

Equação da Reta

$$a = \frac{507,06\,(51,9) - 56\,(466,693)}{10\,(507,06) - (56,0)^2} = 0,0939$$

$$Y_1 = 0,0939 + 0,91x$$

Equação da Curva de 2º Grau

Para achar a equação da curva de 2º grau é preciso calcular os valores apresentados na Tabela (VII.10).

Tabela VII.10

Valores Necessários para Achar a Equação da Curva de 2º Grau

x^3	x^4	x^2y	Y_2	Y_2^2	$(Y_2 - y)^2$
0,008	0,0016	0,0072	0,178	0,031684	25,120144
0,125	0,0625	0,1225	0,471	0,221841	22,268961
1,331	1,4641	1,2947	1,054	1,110916	17,106496
21,952	61,4656	20,7760	2,683	7,198489	6,285049
79,507	341,8801	74,8845	4,095	16,769025	1,199025
166,375	915,0625	156,6950	5,206	27,102436	0,000256
474,552	3701,5056	451,4328	7,290	53,144100	4,410000
857,375	8145,0625	787,8825	8,793	77,316849	12,981609
1331,000	14641,0000	1222,1000	10,093	101,868649	24,039409
2352,637	31290,0721	2127,9867	12,037	144,889369	46,881409
5284,862	59097,5766	4843,1819	51,900	429,653358	160,292358

(1) $51{,}9 = 10\,b_o + 56\,b_1 + 507{,}06\,b_2$

(2) $466{,}693 = 56\,b_o + 507{,}06\,b_1 + 5284{,}862\,b_2$

(3) $4843{,}1819 = 507{,}06\,b_o + 5284{,}862\,b_1 + 59097{,}5766\,b_2$

Resolvendo o sistema de equações, acha-se:

$$b_0 = -0{,}0176$$

$$b_1 = 0{,}9802$$

$$b_2 = -0{,}00555$$

$$Y_2 = -0{,}0176 + 0{,}9802\,x - 0{,}00555\,x^2$$

Com estes dados, pode-se verificar se a curva de segundo grau se adapta aos pontos melhor do que a reta.

Para isto, faz-se o quadro da análise da variância conforme a Tabela (VII.11).

Como o $F_{exp} > F_{0,10(1,7)}$ e mesmo maior que $F_{0,05(1,7)}$ aceita-se a curva de 2º grau como melhor que a reta para descrever os pontos experimentais. Isto indica que há uma variação significativa da recuperação de tório com a concentração deste elemento.

VII.5 – Comparação entre Duas Curvas Padrão

Uma curva de calibração é geralmente construída com padrões acompanhados de certificado e deve ser refeita periodicamente, porque, com o tempo, podem variar algumas condições que afetam o método experimental usado na determinação dos pontos. Pode acontecer também que se façam duas curvas padrão, uma delas com padrões adquiridos (primários) e outra com padrões preparados artificialmente (secundários) e se queira saber se elas são equivalentes. Enfim, podem surgir circunstâncias em que se torne necessário fazer uma comparação entre duas curvas de calibração.

Tabela VII.11

Análise da Variância Aplicada à Curva de 2º Grau

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F
Reta	160,2120	1	0,0804	
Melhoria	0,0804	1		
Curva 2º Grau	160,2924	2		
Resíduo	0,0216	7	$\frac{0,0216}{7} = 0,00308$	$F = \frac{0,0804}{0,00308} = 266,1$
Total	160,314	9		$F_{0,10(1,7)} = 2,0$ $F_{0,05(1,7)} = 5,6$

Sejam:

$$Y_1 = a_1 + b_1 x \qquad (n_1 \text{ pontos})$$

$$Y_2 = a_2 + b_2 x \qquad (n_2 \text{ pontos})$$

as equações correspondentes às curvas que desejamos comparar. O primeiro passo consiste na comparação das variâncias s_{01}^2 e s_{02}^2 pelo teste F. Teoricamente, as variâncias devem ser iguais, porque o método usado para a obtenção dos pontos é o mesmo. Se o teste das variâncias for significativo no nível de confiança desejado, há indicação de que houve algum erro nas medidas e estas devem ser repetidas. Admitindo então que as variâncias sejam iguais, calcula-se a variância média pela equação:

$$\bar{s}_0^2 = \frac{(n_1 - 2)\, s_{01}^2 + (n_2 - 2)\, s_{02}^2}{n_1 + n_2 - 4} \qquad \text{(VII.46)}$$

que tem $(n_1 + n_2 - 4)$ graus de liberdade.

Para saber se duas curvas padrão são equivalentes, verifica-se em primeiro lugar se elas são paralelas, isto é, se $b_1 = b_2$ e, no caso afirmativo, verifica-se a igualdade, $a_1 = a_2$, das ordenadas na origem.

VII.5.1 – Paralelismo entre Duas Curvas Padrão

A variância da diferença entre os dois coeficientes angulares é dada por:

$$s^2_{(b_1-b_2)} = \frac{s^2_{01}}{\Sigma\,(x_1 - \bar{x}_1)^2} + \frac{s^2_{02}}{\Sigma\,(x_2 - \bar{x}_2)^2} \qquad \text{(VII.47)}$$

Como por hipótese s^2_{01} e s^2_{02} são estatisticamente iguais, temos:

$$s^2_{(b_1-b_2)} = \bar{s}^2_{o} \left[\frac{1}{\Sigma\,(x_1 - \bar{x}_1)^2} + \frac{1}{\Sigma\,(x_2 - \bar{x}_2)^2} \right] \qquad \text{(VII.48)}$$

Para simplificar a anotação, chamamos $\Sigma(x_1 - \bar{x}_1)^2 = D_1$ e $\Sigma(x_2 - \bar{x}_2)^2 = D_2$. Aplica-se o teste t à diferença entre os coeficientes angulares e temos:

$$t = \frac{|b_1 - b_2|}{s_{(b_1-b_2)}} \qquad \text{(VII.49)}$$

com $f = n_1 + n_2 - 4$ graus de liberdade. Se o valor de $t_{exp} < t_{tab}$, num nível de significância conveniente p, os coeficientes angulares são iguais, isto é, as retas são paralelas.

Quando as duas curvas de calibração são obtidas por dois métodos diferentes (comparação de métodos) pode acontecer que $s^2_{01} \neq s^2_{02}$. Neste caso, o número de graus de liberdade f é calculado da seguinte maneira:

$$\frac{1}{f} = \frac{C^2}{f_1} + \frac{1 - C^2}{f_2}$$

onde: $f_1 = n_1 - 2$; $f_2 = n_2 - 2$ e C é dado pela seguinte expressão:

$$C = \frac{s^2_{01} / D_1}{s^2_{01} / D_1 + s^2_{02} / D_2}$$

No caso de $s^2_{01} \neq s^2_{02}$, o valor de t é calculado por:

$$t = \frac{|b_1 - b_2|}{\dfrac{s^2_{01}}{D_1} + \dfrac{s^2_{02}}{D_2}} \qquad \text{(VII.50)}$$

e compara-se com $t_{p,f}$ tabelado

VII.5.2 – Coincidência de Duas Curvas Padrão

Se, pelo teste anterior, se obteve que os coeficientes angulares b_1 e b_2 são estatisticamente iguais, deve-se verificar se a_1 e a_2 são também estatisticamente iguais.

Pela Figura (7.5) vemos que $\bar{y} = AC$ e $b\bar{x} = BC$. Então:

$$a_1 = \bar{y}_1 - b_1 \bar{x}_1$$
$$a_2 = \bar{y}_2 - b_2 \bar{x}_2 \qquad \text{(VII.51)}$$

Assim a condição $a_1 = a_2$ é preenchida pela igualdade.

$$\bar{y}_1 - b_1 \bar{x}_1 = \bar{y}_2 - b_2 \bar{x}_2 \qquad \text{(VII.52)}$$

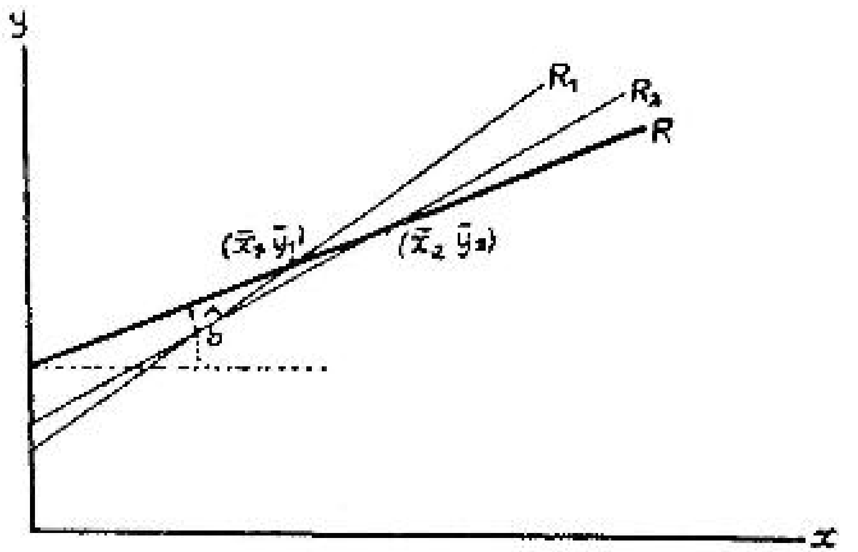

Figura 7.5 – Reta R com Coeficiente Angular $\bar{b}$

Desde que foi admitida a igualdade $b_1 = b_2$, pode-se calcular um valor médio dos coeficientes angulares, substituindo em (VII.51) b_1 e b_2 por $\bar{b}$. Temos então:

$$\hat{b} = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} \qquad \text{(VII.53)}$$

como estimativa do coeficiente angular de todos os pontos. O valor $\bar{b}$ pode ser interpretado como o coeficiente angular de uma reta R que passa pelos pontos médios $(\bar{x}_1\ \bar{y}_1)$ de R_1 e $(\bar{x}_2\ \bar{y}_2)$ de R_2 conforme a Figura (VII.5). Calcula-se também o valor médio $\bar{b}$ fazendo a média ponderada dos coeficientes angulares b_1 e b_2 das duas retas, usando como pesos os valores recíprocos das respectivas variâncias.

$$\bar{b} = \frac{b_1\, D_1/s^2_{01} + b_2\, D_2/s^2_{02}}{D_1/s^2_{01} + D_2/s^2_{02}} \qquad \text{(VII.54)}$$

A estimativa de $\bar{b}$ baseia-se somente no paralelismo das duas retas, enquanto que $\hat{b}$ admite a coincidência das duas retas. A comparação dessas duas estimativas permite testar a hipótese da coincidência.

Pela fórmula (IV.55) sabemos que:

$$s^2_{\bar{x}} = \frac{1}{\Sigma\, 1/s^2_{x_i}}$$

Então a variância da média ponderada $\bar{b}$ dos coeficientes angulares b_1 e b_2, quando s^2_{01} e s^2_{02} são estatisticamente iguais é:

$$s^2_{\bar{b}} = \frac{1}{1/s^2_{b_1} + 1/s^2_{b_2}} = \frac{1}{D_1/s^2_{01} + D_2/s^2_{02}} = \frac{s^2_0}{D_1 + D_2} \qquad \text{(VII.55)}$$

Se $s^2_{01} \neq s^2_{02}$ temos:

$$s^2_{\bar{b}} = \frac{1}{D_1/s^2_{01} + D_2/s^2_{02}} \qquad \text{(VII.56)}$$

Para achar a variância de $\hat{b}$ fazemos:

$$s^2_{\hat{b}} = \left(\frac{\partial \hat{b}}{\partial \bar{y}_1}\right)^2 s^2_{\bar{y}_1} + \left(\frac{\partial \hat{b}}{\partial \bar{y}_2}\right)^2 s^2_{\bar{y}_2}$$

e achamos:

$$s^2_{\hat{b}} = \frac{1}{(\bar{x}_1 - \bar{x}_2)^2} \left[\frac{s^2_{01}}{n_1} + \frac{s^2_{02}}{n_2}\right] \quad \text{(VII.57)}$$

Se $s^2_{01} \neq s^2_{02}$.

No caso de $s^2_{01} = s^2_{02}$, temos:

$$s^2_{\hat{b}} = \frac{\bar{s}^2_0}{(\bar{x}_1 - \bar{x}_2)^2} \left(\frac{n_1 + n_2}{n_1 n_2}\right) \quad \text{(VII.58)}$$

Consequentemente, a variância da diferença $(\hat{b} - \bar{b})$ é a soma das variâncias:

$$s^2_{(\hat{b} - \bar{b})} = \left[\frac{1}{(\bar{x}_1 - \bar{x}_2)^2} \left(\frac{n_1 + n_2}{n_1 n_2}\right) + \frac{1}{D_1 + D_2}\right] \bar{s}^2_0 \quad \text{(VII.59)}$$

No caso de $s^2_{01} \neq s^2_{02}$, temos:

$$s^2_{(\hat{b} - \bar{b})} = \frac{1}{(\bar{x}_1 - \bar{x}_2)^2} \left(\frac{s^2_{01}}{n_1} + \frac{s^2_{02}}{n_2}\right) + \frac{1}{D_1/s^2_{01} + D_2/s^2_{02}} \quad \text{(VII.60)}$$

Quando s^2_{01} e s^2_{02} forem estatisticamente iguais, a coincidência das duas retas se testa pelo valor de t achado por:

$$t = \frac{|\hat{b} - b|}{s_{(\hat{b} - \bar{b})}} \quad \text{(VII.61)}$$

com $f = n_1 + n_2 - 4$ graus de liberdade. Se $t_{exp} < t_{p,f}$ tabelado, aceita-se a coincidência das duas retas.

Quando $s^2_{01} \neq s^2_{02}$ surge uma dificuldade quanto ao número de graus de liberdade que devem ser atribuídos a t. O problema só se resolve se f_1 e f_2 são pelo menos maiores que 10, tomando o valor de $s^2_{\hat{b}}$ como estimativa de $\sigma^2_{(\hat{b} - \bar{b})}$ e comparando o valor:

$$u = \frac{|\hat{b} - b|}{\sigma_{(\hat{b} - \bar{b})}} \quad \text{(VII.62)}$$

com os valores tabelados da distribuição normal.

EXERCÍCIO 47

Foram feitas duas curvas de calibração para um certo método experimental. Na primeira (R_1) foram usados padrões com certificado e na segunda (R_2) foram usados padrões preparados no laboratório. Pergunta-se, num nível de confiança de 0,90, se os padrões artificiais foram bem preparados e podem ser usados em substituição aos primeiros. Os resultados obtidos estão na Tabela (VII.12).

Tabela VII.12

Resultados de Análises de Padrões para a Construção das Curvas R_1 e R_2

x_1	y_1	x_2	y_2
3,0	2,9	2,5	2,6
5,8	6,1	7,6	7,4
7,3	7,2	12,8	13,1
9,7	10,1	19,7	20,0
13,6	13,5	27,4	27,1
18,5	18,6	38,0	38,1
24,6	24,9		
30,3	30,3		
112,8	113,6	108,0	108,3

$\bar{x}_1 = 14,1$ $\qquad \bar{x}_2 = 18,0$

$\bar{y}_1 = 14,2$ $\qquad \bar{y}_2 = 18,05$

$\Sigma x_1^2 = 2240,48$ $\qquad \Sigma x_2^2 = 2810,7$

$\Sigma y_1^2 = 2265,78$ $\qquad \Sigma y_2^2 = 2819,15$

$\Sigma x_1 y_1 = 2252,94$ $\qquad \Sigma x_2 y_2 = 2814,76$

$$a_1 = \frac{113,6\ (2240,48) - 112,8\ (2252,94)}{8\ (2240,48) - (112,8)^2} \simeq 0,0842$$

$$b_1 = \frac{8\ (2252,94) - 112,8\ (113,6)}{8\ (2240,48) - (112,8)^2} = 1,0018$$

$$s_{01}^2 = \frac{1}{6}\left[2265,78 - \frac{(113,6)^2}{8} - \frac{8\left[(2252,94 - \dfrac{(112,8)\ (113,6)}{8}\right]^2}{8\ (2240,48) - (112,8)^2}\right] = 0,049643$$

$$a_2 = \frac{108,3\ (2810,7) - 108\ (2814,76)}{6\ (2810,7) - (108)^2} = 0,0778$$

$$b_2 = \frac{6\ (2814,76) - 108\ (108,3)}{6\ (2810,7) - (108)^2} = 0,9985$$

$$s^2_{02} = \frac{1}{4}\ [2819,15 - \frac{(108,3)^2}{6} - \frac{6\ [(2814,76 - \frac{(108)\ (108,3)}{6}]^2}{6\ (2810,7) - (108)^2}] = 0,078232$$

1) Teste para verificar a igualdade das variâncias.

$$F = \frac{0,078232}{0,049643} = 1,58$$

$$F_{0,10(4,6)} = 4,5 \text{ (bicaudal)}$$

Conclui-se que as variâncias são iguais.

2) Teste para verificar se as duas retas são paralelas.

$$s^2_0 = \frac{6\ (0,049643) + 4\ (0,078232)}{6 + 4} = 0,061079$$

$$s^2_{(b_1 - b_2)} = 0,061079\ [\frac{1}{5200} + \frac{1}{5200,2}] = 0,00002349$$

$$s_{(b_1 - b_2)} = 0,0048$$

$$t = \frac{1,0018 - 0,9985}{0,0048} = 0,69$$

$$t_{0,10(10)} = 1,81$$

Conclusão: As retas são paralelas.

3) Teste para verificar a coincidência das duas retas.

$$\hat{b} = \frac{14{,}2 - 18{,}05}{14{,}1 - 18{,}0} = 0{,}9872$$

$$\bar{b} = \frac{1{,}0018(5200/0{,}049643) + 0{,}9985\ (5200{,}2/0{,}078232)}{(5{,}200/0{,}049643) + (5200{,}2/0{,}078232)} = 1{,}0005$$

$$s^2_{\bar{b}} = \frac{0{,}061079}{5200 + 5200{,}2} = 0{,}0000058728$$

$$s_{\bar{b}} = 0{,}0024$$

$$s^2_{\hat{b}} = \frac{0{,}061079}{(18 - 14{,}1)^2}\ \frac{48}{14} = 0{,}0138$$

$$s_{\hat{b}} = 0{,}117$$

$$s^2_{(\hat{b}-\bar{b})} = 0{,}117 + 0{,}0024 = 0{,}1194$$

$$s_{(\hat{b}-\bar{b})} = 0{,}345$$

$$t = \frac{1{,}0005 - 0{,}9872}{0{,}345} = 0{,}051$$

$$0{,}051 < t_{0{,}10{,}10} = 1{,}81$$

Concluimos que as retas são coincidentes.

EXERCÍCIO 48

Foram feitas duas cruvas de calibração com padrões diferentes em épocas diferentes. Sabe-se que os padrões são equivalentes por testes anteriores. Deseja-se saber se houve alterações nas condições de trabalho durante o tempo transcorrido. Nível de significância 0,10. Os resultados obtidos estão na Tabela (VII.13).

Tabela VII.13

Resultados de Análises de Padrões (x_1 e y_1) Analisados em Épocas Diferentes

x_1	y_1	x_2	y_2
2	1,7	2	2,2
4	3,8	5	5,3
6	6,1	7	7,1
8	7,7	10	10,5
10	9,8	15	15,0
13	13,1	20	20,8
17	16,7	27	27,1
20	19,5	32	33,0
25	24,9	38	38,6
30	30,0	44	44,1
35	34,7	53	53,7
40	40,2		
210	208,2	253	257,4

$x_1 = 17{,}5$ $\qquad x_2 = 23{,}0$

$y_1 = 17{,}35$ $\qquad y_2 = 23{,}4$

$\Sigma x_1^2 = 5428$ $\qquad \Sigma x_2^2 = 8745$

$\Sigma y_1^2 = 5380{,}76$ $\qquad \Sigma y_2^2 = 8993{,}1$

$\Sigma x_1 y_1 = 5404$ $\qquad \Sigma x_2 y_2 = 8867{,}6$

$$a_1 = \frac{208{,}2\ (5428) - 210\ (5404)}{12\ (5428) - (210)^2} = -\ 0{,}2300$$

$$b_1 = \frac{12\ (5404) - 210\ (208{,}2)}{12\ (5428) - (210)^2} = 1{,}0043$$

$$s_{01}^2 = \frac{1}{10}\left[5380{,}76 - \frac{(208{,}2)^2}{12} - \frac{12\left[5404 - \frac{(210)\ (208{,}2)^2}{12}\right]^2}{12\ (5428) - (210)^2}\right] = 0{,}045791$$

$$s_{01} = 0{,}191$$

$$a_2 = \frac{257{,}4\ (8745) - 253\ (8867{,}6)}{11\ (8745) - (253)^2} = 0{,}2318$$

$$b_2 = \frac{11\ (8867{,}6) - 253\ (257{,}4)}{11\ (8745) - (253)^2} = 1{,}0073$$

$$s^2_{02} = \frac{1}{9}\ [8993{,}1 - \frac{(257{,}4)^2}{11} - \frac{11\ [8867{,}6 - \frac{(253)\ (257{,}4)}{11}]^2}{11\ (8745) - (253)^2}] = 0{,}109276$$

$$s_{02} = 0{,}3306$$

1) Teste para verificar se as variâncias são iguais:

$$\frac{0{,}109276}{0{,}045791} = 2{,}38$$

$$F_{0{,}10(9{,}10)} = 1{,}8$$

Como $F_{exp} > F_{0{,}10(9{,}10)}$ conclui-se que as variâncias não são estatisticamente iguais.

2) Teste para verificar se as retas são paralelas:

$$s^2_{(b_1 - b_2)} = \frac{0{,}045791}{21036} + \frac{0{,}109276}{32186} = 0{,}0000055719$$

$$s_{(b_1 - b_2)} = 0{,}00236$$

$$t = \frac{1{,}0073 - 1{,}0043}{0{,}00236} = 1{,}69$$

Número de graus de liberdade de t:

$$c = \frac{0{,}45791\ /\ 21036}{(0{,}045791 / 21036) + (0{,}109276 / 32186)} = 0{,}39$$

$$\frac{1}{f} = \frac{(0{,}39)^2}{10} + \frac{1 - (0{,}39)^2}{9} = 0{,}0565 \qquad f \geq 18$$

$t_{0,10(18)} = 1,73$

Como $t_{exp} < t_{0,10(18)}$, concluímos que as retas são paralelas.

3) Teste para verificar se há coincidência entre as duas curvas padrão.

$$\hat{b} = \frac{23,4 - 17,35}{23,0 - 17,5} = 1,10$$

$$\bar{b}' = \frac{1,0043\,(21036 / 0,045791) + 1,0073\,(32186 / 0,109276)}{(21036 / 0,045791) + (32186 / 0,109276)} = 1,0055$$

$$s^2_{\hat{b}} = \frac{1}{(23 - 17,5)^2}\left(\frac{0,045791}{12} + \frac{0,109276}{10}\right) = 0,0004874$$

$$s^2_{\bar{b}} = \frac{1}{(21036 / 0,045791) + (32186 / 0,109276)} = 0,000001326$$

$$s^2_{(\hat{b}-\bar{b})} = s^2_{\hat{b}} + s^2_{\bar{b}} = 0,000001326 + 0,0004874$$

$$s^2_{(\hat{b}-\bar{b})} = 0,000488726$$

$$s_{(\hat{b}-\bar{b})} = 0,022$$

Neste caso $s^2_{01} \neq s^2_{02}$ e f_1 e f_2 10. Assim

$$u \cong \frac{1,10 - 1,0055}{0,022} = \frac{0,0945}{0,022} = 4,29$$

$$P(u > 4,29) = 0,000008934$$

Conclui-se que as retas não são coincidentes, apesar de paralelas, isto é, durante o tempo considerado, houve uma modificação nas condições de trabalho, motivada por reagentes de pureza diferente, variações de temperatura etc.

VII.6 – Precisão de Resultados Obtidos por Meio de Curvas de Calibração

Suponhamos que a equação da reta obtida por n pontos seja

$$Y = a + bx$$

Essa expressão por (VII.51) também pode ser escrita da seguinte maneira:

$$Y = \bar{y} + b(x - \bar{x}) \qquad \text{(VII.63)}$$

Seja y' um resultado experimental e queremos conhecer x' correspondente por meio da curva padrão. Substituindo em (VII.63), Y e x por y' e x' respectivamente, podemos achar o valor de x' que é:

$$x' = \frac{y' - \bar{y}}{b} + \bar{x} \qquad \text{(VII.64)}$$

Aplicando a lei de acumulação de erros a essa expressão temos:

$$s^2_{x'} = \frac{1}{b^2}\, s^2_y + \frac{(y' - \bar{y})^2}{b^4}\, s^2_b \qquad \text{(VII.65)}$$

onde:

$$s^2_b = \frac{n\, s^2_0}{n \Sigma x^2 - (\Sigma x_i)^2}$$

$$s^2_y = (1 + \frac{1}{m})\, s^2_0$$

Admitimos aqui que, tanto os pontos da curva padrão como o resultado da amostra tenham sido obtidos com o mesmo número de determinações paralelas. Assim o erro do valor x' procurado é:

$$s_{x'} = \frac{s_0}{b} \sqrt{1 + \frac{1}{n} + \frac{(y' - \bar{y})^2}{b^2\, (\Sigma x^2 - n\bar{x}^2)}} \qquad \text{(VII.66)}$$

com f = n – 2 graus de liberdade

Os limites de confiança para o valor de x' será então:

$$x' \pm t_p\, s_{x'} \qquad \text{(VII.67)}$$

onde p é o nível de significância escolhido para t com f = n – 2 graus de liberdade.

EXERCÍCIO 49

A análise de uma amostra que contém zircônio deu $y' = 10{,}2$ e a de outra amostra deu $y' = 4{,}1$. Achar num nível de confiança de 0,05 os valores correspondentes de x' e os respectivos intervalos de confiança.

Vamos usar a curva padrão do Exercício 43 deste Capítulo, para análise de zircônio em amostras desconhecidas. Os valores achados foram:

$a = 0{,}1554$	$b = 1{,}0023$	$s_0^2 = 0{,}028552$	$\bar{x} = 10{,}6$
$s_a^2 = 0{,}009352$	$s_b^2 = 0{,}0000662982$	$s_0 = 0{,}169$	$\bar{y} = 10{,}78$
$s_a = 0{,}097$	$s_b = 0{,}00814$	$n = 15$	$\Sigma x^2 = 2116{,}06$

1) Para $y' = 10{,}2$

$$x' = \frac{10{,}2 - 10{,}78}{1{,}0023} + 10{,}6 = 10{,}02$$

$$s_{x'} = \frac{0{,}169}{1{,}0023}\sqrt{1 + \frac{1}{15} + \frac{(10{,}2 - 10{,}78)^2}{1{,}0023^2\ (2116{,}06 - 15 \times 10{,}6^2)}}$$

$$s_{x'} = 0{,}17$$

$$t_{0{,}05(13)} = 2{,}16$$

$$x' = 10{,}02 \pm 0{,}37$$

2) Para $y' = 4{,}1$

$$x' = \frac{4{,}1 - 10{,}78}{1{,}0023} + 10{,}6 = 3{,}93$$

$$s_{x'} = \frac{0{,}169}{1{,}0023}\sqrt{1 + \frac{1}{15} + \frac{(4{,}1 - 10{,}78)^2}{1{,}0023^2\ (2116{,}06 - 15 \times 10{,}6^2)}}$$

$$s_{x'} = 0{,}20$$

$$x' = 3{,}93 \pm 0{,}43$$

Pela fórmula (VII.66) verificamos que o erro em x' é tanto maior quanto mais afastado estiver o valor y' achado da média $\bar{y}$.

O erro é mínimo quando $y' = \bar{y}$

O valor de y' pode ser representado da seguinte forma:

$$y' = a+b\left[x' \pm \frac{s_o t}{b}\sqrt{1 + \frac{1}{n} + \frac{(y'-\bar{y})^2}{b^2\,(\Sigma x^2 - n\bar{x}^2)}}\right] \qquad \text{(VII.68)}$$

que é a equação de uma hipérbole com as variáveis (x', y'). Os valores possíveis de x', compatíveis com os resultados y' estão compreendidos entre os ramos da hipérbole, conforme a Figura VII.6.

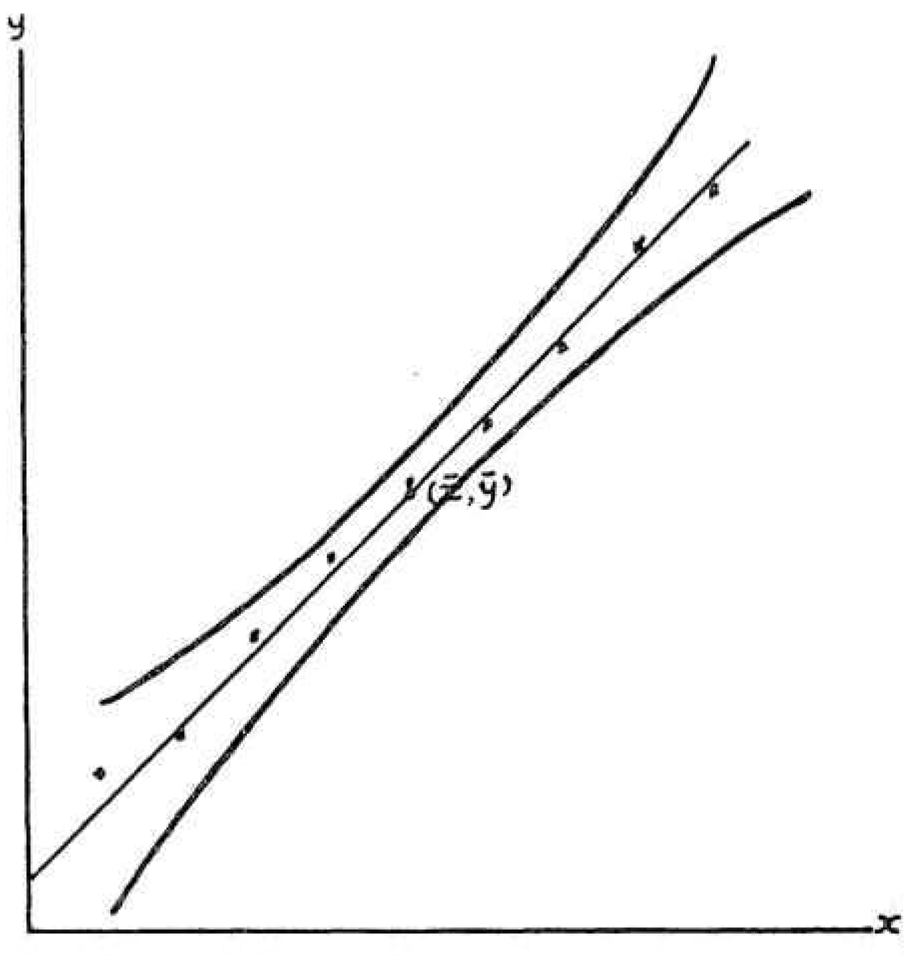

Figura 7.6 — Hiperbole que Determina os Limites de Variação dos Resultados

Quando a hipótese da linearidade é válida, é conveniente construir a curva padrão a partir de n padrões e fazer m determinações paralelas de cada um. O erro na determinação será então:

$$s_{x'} = \frac{s_0}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y' - \bar{y})^2}{b^2 (\Sigma x^2 - n\bar{x}^2)}} \qquad \text{(VII.69)}$$

porque, neste caso:

$$s^2_{(y' - \bar{y})} = (\frac{1}{m} + \frac{1}{n})\, s_0^2$$

Para a construção de uma curva padrão são necessários pelo menos 3 padrões, porque se perdem dois graus de liberdade nas determinações de a e b. Uma curva padrão construída com 3 padrões vai dar resultados baixos para s_0^2 mas os limites de confiança são necessariamente largos porque na fórmula (VII.67) o valor de t é grande ($t_{0,05(1)} = 12,71$)

CAPÍTULO VIII

DELINEAMENTO FATORIAL

Um dos métodos para estudar a influência de duas ou mais variáveis sobre o resultado de um experimento complexo consiste em manter constantes n – 1 variáveis e variar a enésima até esgotar todas as possibilidades. Quando o número de variáveis é grande, o processo torna-se trabalhoso e às vezes inexequível.

O delineamento fatorial permite variar simultaneamente todas as variáveis em estudo e obter o máximo de informações com um número mínimo de provas.

DEFINIÇÕES

Fator

É toda a condição experimental que pode ser variada de um a outro experimento.

Nível

É toda a variação sofrida por um fator.

Tratamento

É o conjunto de condições (diversos níveis de cada fator) sob as quais é executada uma certa prova.

Efeito

É a variação do resultado provocada pela mudança de nível de um determinado fator.

VIII.1 — Estudo da Influência de 2 Fatores

Vamos considerar os fatores A e B nos níveis (1) e (2). Pela experimentação clássica, para avaliar a influência de A, mantemos B fixo e temos:

$$A = A_2B_1 - A_1B_1 \qquad \text{(VIII.1)}$$

e para o fator B temos:

$$B = A_1B_2 - A_1B_1 \qquad \text{(VIII.2)}$$

No total, são feitas 3 provas. Para avaliar o erro experimental, cada prova deve ser feita pelo menos duas vezes o que dá um número mínimo de 6 provas.

Como exemplo, vamos examinar um caso prático.

EXERCÍCIO 50

Foi estudado o efeito do pH e da temperatura no rendimento de uma separação química. Seja A o fator temperatura, nos níveis: $A_1 = 25°C$ e $A_2 = 60°C$ e seja B o fator pH nos níveis B_1 = pH 4 e B_2 = pH 5. Suponhamos que tenham sido obtidos os resultados da Tabela VIII.1.

Tabela VIII.1

Resultados do Rendimento de uma Separação Química Sujeita à Influência de 2 Fatores

Tratamento	1a. Série (%)	2a. Série (%)	Diferença (%)	Média (%)
A_1B_1	89,5	90,7	1,2	90,1
A_2B_2	95,4	96,2	0,8	95,8
A_1B_2	94	94,9	0,8	94,5

A influência de A é: $A_2B_1 - A_1B_1 = 95{,}8 - 90{,}1 = 5{,}7$

A influência de B é: $A_1B_2 - A_1B_1 = 94{,}5 - 90{,}1 = 4{,}4$

A partir dos três pares de resultados, pode-se calcular a variância devida ao erro experimental:

$$s^2 = \frac{d^2}{2n} = \frac{1,2^2 + 0,8^2 + 0,8^2}{2 \times 3} = \frac{2,72}{6} = 0,45$$

A variância da média de duas provas é:

$$\frac{0,45}{2} = 0,23$$

e a variância da diferença entre a média de duas provas é:

$$\frac{0,45}{2} \times 2 = 0,45$$

O desvio padrão é:

$$s = \sqrt{0,45} = 0,67$$

Calcula-se a menor diferença significativa com o valor de t com 3 graus de liberdade.

Para $p = 0,05$, a menor diferença é: $0,67 \times 3,182 = 2,13$

Para $p = 0,01$, a menor diferença é: $0,67 \times 5,841 = 3,91$

Como a influência de A(5,7) e a de B(4,4) são valores maiores que 3,91, conclui-se que os efeitos de A e de B são significativos mesmo para $p = 0,01$.

Este tipo de experimentação (clássica) não fornece, porém, nenhuma informação sobre a interação dos dois fatores.

No delineamento fatorial, são realizadas todas as combinações de tratamentos possíveis, isto é, $2^2 = 4$ tratamentos o que dá um total de 8 provas, no caso de serem feitas só duas provas para cada tratamento. Considerando ainda o exemplo anterior, seria necessário fazer também o tratamento A_2B_2.

EXERCÍCIO 51

Para complementar o exercício anterior, foi feito o tratamento A_2B_2 cujos resultados correspondem ao rendimento da separação química a 60°C e pH 5. Todos os resultados estão na Tabela VIII.2.

A influência do fator A pode ser avaliada nos 2 níveis de B, pelas diferenças $A_2B_1 - A_1B_1$ e $A_2B_2 - A_1B_2$. Se estas diferenças forem iguais, a menos do erro experimental, pode-se concluir que não há interações entre A e B e a média dessas duas observações vai corresponder à influência de A. O mesmo pode ser feito para o fator B.

Tabela VIII.2

Delineamento Fatorial Aplicado ao Rendimento de uma Separação Química Sujeita à Influência de 2 Fatores

	A_1		A_2		Totais
B_1	89,5	90,7	95,4	96,2	371,8
	(Média = 90,1)		(Média = 95,8)		
B_2	94,1	94,9	97,6	98,2	
	(Média = 94,5)		(Média = 97,9)		384,8
Totais	369,2		387,4		756,6

Para verificar se existe interação entre os fatores e se eles têm influência sobre os resultados, usa-se a análise da variância. Calcula-se:

a) A soma total dos quadrados correspondentes às 8 observações que é:

$$SQT = (89,5)^2 + (90,7)^2 + \ldots + (98,2)^2 - \frac{(756,6)^2}{8} = 71622,16 - 71555,445$$

$$SQT = 66,715$$

b) A soma dos quadrados que corresponde à influência de A: é a soma dos resultados no nível A_2 – (soma dos resultados no nível A_1). A diferença obtida é elevada ao quadrado e dividida pelo número de observações.

$$SQA = (387,4 - 369,2)^2 / 8 = 41,405$$

c) A soma dos quadrados que corresponde à influência de B: é o mesmo do ítem anterior só que aplicado ao fator B.

$$SQB = (384,8 - 371,8)^2 / 8 = 21,125$$

d) A soma dos quadrados correspondentes à interação AB é:

$$(A_1B_1 + A_2B_2 - A_1B_2 - A_2B_1)^2 / n$$

$$SQAB = (89,5 + 90,7 + 97,6 + 98,2 - 95,4 - 96,2 - 94,1 - 94,9)^2 / 8 = 2,645$$

A soma dos quadrados correspondente ao erro experimental (resíduo) é calculada por diferença (Tabela VIII.3), mas, a título de controle, pode-se calcular a variância do erro experimental, a partir das diferenças entre as duas observações dos 4 pares de resultados.

$$SQR = \frac{(1,2)^2 + (0,8)^2 + (0,8)^2 + (0,6)^2}{8} = 0,385$$

Tabela VIII.3

Análise da Variância de um Delineamento Fatorial de 2 Fatores com 2 Níveis

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	Componentes da Variância	F
Efeito de A	41,405	$(a-1)=1$	41,405	$\sigma_R^2 + n\,\sigma_{AB}^2 + b\,\sigma_A^2 = s_A^2$	$s_A^2/s^2 = 107$
Efeito de B	21,125	$(b-1)=1$	21,125	$\sigma_R^2 + n\,\sigma_{AB}^2 + a\,\sigma_B^2 = s_B^2$	$s_B^2/s^2 = 55$
Interação AB	6,645	$(a-1)\,(b-1)=1$	2,645	$\sigma_R^2 + n\,\sigma_{AB}^2 = s_{AB}^2$	$s_{AB}^2/s^2 = 6,9$
Resíduo (Erro Experim.)	1,54	$ab(n-1)=4$	0,385	$\sigma_R^2 = s^2$	
Total		7			

a = número de níveis de A = 2
b = número de níveis de B = 2
n = número de interações para cada tratamento = 2

$$F_{0,05(1,4)} = 7,7$$

$$F_{0,01(1,4)} = 21$$

Pela Tabela VIII.3 pode-se ver que a variância devida à interação AB está incluída nas variâncias correspondentes a A e a B. Pelo teste $F = s_{AB}^2/s^2$, acha-se um valor não significativo para a influência da interação AB, mas os valores de F para a influência de A e B são significativos, mesmo num nível de significância $p = 0,01$.

A variância experimental de um resultado é 0,385, portanto a da média de 4 resultados é $0,385/4 = 0,0962$ e a da diferença entre duas médias é: $0,0962 \times 2 = 0,1925$, valor bem mais baixo que no caso da experimentação clássica que era 0,45. O desvio padrão da diferença entre duas médias é $\sqrt{0,1925} = 0,439$. A menor diferença entre duas médias, usando t com 4 graus de liberdade é:

para p = 0,05 : 2,776 x 0,439 = 1,22

para p = 0,01 : 4,604 x 0,439 = 2,02

Pode-se concluir que, não havendo interação AB, cada fator tem um efeito independente, podendo ser estimado com a mesma precisão, como se o experimento estivesse concentrado num único fator.

Na experimentação clássica, a variância da diferença entre os 2 níveis de cada fator é $2s^2/n$, enquanto que no delineamento fatorial é $2s^2/2n = s^2/n$. Assim, com 2 provas a mais (8 em lugar de 6), a variância fica reduzida à metade e, além disso, se houver uma interação entre os fatores, ela é posta em evidência pelo método do delineamento fatorial.

VIII.2 – Estudo de Vários Fatores Envolvidos num Processo

As condições adotadas em processos industriais ou em nível de laboratório dependem de estudos prévios com a finalidade de se obter o melhor resultado (rendimento, pureza, fator de separação de íons, etc.) possível.

Por exemplo, no rendimento da separação de um íon em solução, por uma resina trocadora de íons, podem ser estudados os seguintes fatores:

1) Concentração do íon na solução que vai ser percolada.
2) Temperatura do sistema.
3) Velocidade de percolação.
4) pH da solução a ser percolada.
5) Composição da solução de eluição de um dos íons (quando ambos forem retidos).

Muitos outros fatores poderiam ser mencionados tais como: granulometria da resina, tipo de resina, dimensões da coluna, etc.

Vamos considerar um caso hipotético em que são estudados 4 fatores A, B, C, D, cada um com 2 níveis: mínimo indicado com índice (1) e máximo, com índice (2). As combinações possíveis são $2^4 = 16$, indicadas pelos seguintes símbolos:

$A_1B_1C_1D_1 = (1)$	$A_1B_2C_2D_1 = BC$
$A_2B_1C_1D_1 = A$	$A_1B_2C_1D_2 = BD$
$A_1B_2C_1D_1 = B$	$A_1B_1C_2D_2 = CD$
$A_1B_1C_2D_1 = C$	$A_2B_2C_2D_1 = ABC$
$A_1B_1C_1D_2 = D$	$A_2B_2C_1D_2 = ABD$
$A_2B_2C_1D_1 = AB$	$A_2B_1C_2D_2 = ACD$
$A_2B_1C_2D_1 = AC$	$A_1B_2C_2D_2 = BCD$
$A_2B_1C_1D_2 = AD$	$A_2B_2C_2D_2 = ABCD$

Quando todos os fatores estão no nível mais baixo, $A_1B_1C_1D_1$, o tratamento é representado por (1). O efeito médio de A pode ser calculado por uma série de n diferenças em que A está nos 2 níveis, tais como:

$$
\begin{array}{ll}
A & - (1) \\
AB & - B \\
AC & - C \\
\vdots & \\
ABCD & - BCD
\end{array}
$$

O efeito médio de A é a soma destas diferenças dividida pelo número n que no caso é 8.

$$\bar{A} = \frac{1}{8}(A+AB+AC+AD+ABC+ABD+ACD+ABCD) - ((1)+B+C+D+BC+BD+CD+BCD))$$

Do mesmo modo se calculam os efeitos médios de B, C e D. É fácil observar que, no caso de $\bar{A}$, são positivos todos os termos em que aparece o tratamento A e negativos todos os termos em que não aparece.

As interações de segunda ordem, entre os fatores A e B por exemplo, tem um valor médio $\overline{AB}$ dado pela metade da diferença entre o efeito médio de A quando B está no nível (2) e o mesmo efeito quando B está no nível (1), isto é:

$$A(B \text{ no nível } (2)) = 1/4\,[AB + ABC + ABD + ABCD - (B + BC + BD + BCD)]$$

$$A(B \text{ no nível } (1)) = 1/4\,[A + AC + AD + ACD - ((1) + C + D + CD)]$$

$$\overline{AB} = 1/8\,[AB+ABC+ABD+ABCD+(1)+C+D+CD - (B+BC+BD+BCD+A+AC+AD+ACD)]$$

É bom notar que são positivos todos os termos em que aparece AB e onde não aparece nem A e nem B; todos os outros são negativos. Do mesmo modo se calculam todas as outras interações de 2 fatores.

No caso das interações de 3a. ordem, ABC por exemplo, calcula-se da seguinte maneira:

$$AB(C \text{ no nível } C_2) = 1/4\,[ABC + ABCD + C + CD - (AC + ACD + BC + BCD)]$$

$$AB(C \text{ no nível } C_1) = 1/4\,[AB + ABD + (1) + D - (A + AD + B + BD)]$$

$$\overline{ABC} = 1/8\,[ABC+ABCD+C+CD+A+AD+B+BD - (AC+ACD+BC+BCD+AB+ABD+(1)+D)]$$

Como regra, para memorização fácil, são positivos todos os termos em que aparece ABC e os que aparece um só desses 3 fatores isolados ou com o 4º fator D. As outras interações de 3 fatores são calculadas do mesmo modo.

Pode-se depois calcular a interação ABCD.

Pode-se chegar a esses mesmos resultados desenvolvendo as expressões abaixo

Efeito total de A = (A − 1) (B+1) (C + 1) (D + 1)

Interação total de AB = (A − 1) (B − 1) (C + 1) (D + 1)

Interação total ABCD = (A − 1) (B − 1) (C − 1) (D − 1)

EXERCÍCIO 52

Vamos supor que A, B, C e D sejam fatores que influem no rendimento da produção de UF_4 e que os resultados sejam os da Tabela VIII.4 dos quais foi deduzido 90 % para simplificar os cálculos.

Tabela VIII.4

Resultados do Rendimento da Produção de UF_4 a Menos de 90%

		A_1				A_2			
		B_1		B_2		B_1		B_2	
C_1	D_1	(1)	1,2	B	3,0	A	5,0	AB	5,9
	D_2	D	2,5	BD	3,4	AD	5,5	ABD	6,8
C_2	D_1	C	2,0	BC	3,0	AC	5,1	ABC	6,7
	D_2	CD	2,8	BCD	6,2	ACD	6,9	ABCD	6,5

A ordem de execução das provas deve ser sorteada para conservar a aliatoriedade. Aplicando as expressões dos efeitos e interações nos resultados da Tabela VIII.4 temos:

$$\overline{A} = 1/8\,[5,0 + 5,9 + 5,1 + 5,5 + \ldots - (1,2 + 3,0 + 2,0 + \ldots)] = 3,0375$$

Do mesmo modo se calculam os outros efeitos e intereações médios.

Os resultados são:

$$\overline{B} = 1,3125$$
$$\overline{C} = 0,7375$$
$$\overline{D} = 1,0875$$
$$\overline{AB} = -0,4625$$
$$\overline{AC} = -0,2375$$

$\overline{AD} = -0{,}3375$

$\overline{BC} = -0{,}0875$

$\overline{BD} = -0{,}0125$

$\overline{CD} = 0{,}3125$

$\overline{ABC} = -0{,}3375$

$\overline{ABD} = -0{,}3875$

$\overline{ACD} = -0{,}2625$

$BCD = 0{,}1125$

$ABCD = 0{,}7125$

$(1) = 9{,}0625$

Outro método de cálculo foi proposto por F. Yates, conforme a Tabela VIII.5.

Tabela VIII.5

Cálculo dos Efeitos e Interações Pelo Método de Yates

Combinação Fatorial	Rendimento (−90%)	Col 1	Col 2	Col 3	Col 4	Efeito Médio (Col 4)/8	Soma dos Quadrados $(Col\ 4)^2/16$
(1)	1,2	6,2	15,1	31,9	–	–	–
A	5,0	8,9	16,8	40,6	24,3	3,0375	36,9056
B	3,0	7,1	18,2	13,5	10,5	1,3125	6,8906
AB	5,9	9,7	22,4	10,8	−3,7	−0,4625	0,8556
C	2,0	8,0	6,7	5,3	5,9	0,7375	2,1756
AC	5,1	10,2	6,8	5,2	−1,9	−0,2375	0,2256
BC	3,0	9,7	6,4	−0,3	−0,7	−0,0875	0,0306
ABC	6,7	12,7	4,4	−3,4	−2,7	−0,3375	0,4556
D	2,5	3,8	2,7	1,7	8,7	1,0875	4,7306
AD	5,5	2,9	2,6	4,2	−2,7	−0,3375	0,4556
BD	3,4	3,1	2,2	0,1	−0,1	−0,0125	0,0006
ABD	6,8	3,7	3,0	−2,0	−3,1	−0,3875	0,6006
CD	2,8	3,0	−0,9	−0,1	2,5	0,3125	0,3906
ACD	6,9	3,4	0,6	0,8	−2,1	−0,2625	0,2756
BCD	6,2	4,1	0,4	1,5	0,9	0,1125	0,0506
ABCD	6,5	0,3	−3,8	−4,2	−5,7	0,7125	2,0306
Total	72,5						

Os números da coluna 1 se obtém somando 2 a 2 os rendimentos da coluna anterior, isto é: 1,2 + 5,0 = 6,2; 5,9 + 3,0 = 8,9, etc. Assim se obtém os 8 primeiros números. Os restantes se obtém por subtração: 5,0 – 1,2 = 3,8; 5,9 – 3,0 = 2,9 etc. A coluna 2 se obtém fazendo o mesmo com os números da coluna 1 e assim por diante até formar a coluna 4.

Para interpretar os resultados, usa-se a análise da variância. Neste caso, não temos resultados paralelos para calcular o erro experimental. Substituímos a estimativa deste erro pela variância combinada das interações entre 3 e 4 fatores, porque uma interação entre 3 fatores é pouco provável e entre 4 é menos provável ainda.

Calcula-se então:

1) Soma dos quadrados totais:

$$SQT = (1{,}2^2 + 5{,}0^2 + \ldots + 6{,}2^2 + 6{,}5^2) - (72{,}5)^2/16 = 56{,}0744$$

2) Soma dos quadrados atribuída a cada um dos fatores. Obtém-se elevando ao quadrado o efeito total do fator em questão, dividindo-se depois pelo número total de observações. Por exemplo para A temos:

$$SQA = (24{,}3)^2 / 16 = 36{,}9056$$

Cada soma de quadrados deste tipo tem 1 grau de liberdade, porque cada fator foi estudado em 2 níveis. Nas interações entre fatores, o número de graus de liberdade é o produto do número de graus de liberdade de cada fator. O quadro da análise da variância se apresenta conforme a Tabela VIII.6.

Tabela VIII.6

Análise da Variância de um Delineamento Fatorial de 4 Fatores com 2 Níveis

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade		Quadrado Médio	F	P
A	36,9056	1		36,9056	54,06	< 0,01
B	6,8906	1		6,8906	10,09	< 0,05
C	2,1756	1		2,1756	3,19	
D	4,7306	1		4,7306	6,93	< 0,05
AB	0,8556	1		0,8556	1,25	
AC	0,2256	1		0,2256	0,33	
AD	0,4556	1		0,4556	0,67	
BC	0,0306	1		0,0306	0,04	
BD	0,0006	1		0,0006	0,00	
CD	0,3906	1		0,3906	0,57	
ABC	0,4556	1				
ABD	0,6006	1				
ACD	0,2756	1	5	0,6820		
BCD	0,0506	1				
ABCD	2,0306	1				
Total	56,074			$F_{0,01(1,5)} = 16{,}3$ $F_{0,05(1,5)} = 6{,}6$		

Os efeitos de 3 e 4 fatores foram combinados e a variância correspondente (0,6820) foi usada como estimativa do erro experimental, com 5 graus de liberdade. Vemos também que o efeito dominante é o do fator A, mas os efeitos B e D também são significativos num nível de significância de 0,05. As interações entre os fatores são desprezíveis.

Quando o experimento só tem 3 fatores com 2 níveis, não é possível considerar a soma das interações dos fatores como estimativa do erro experimental, porque para isso deveriam ser usadas as interações entre 2 e 3 fatores. Como as interações entre 2 fatores são frequentes na prática, torna-se necessário fazer uma estimativa do erro experimental por meio de provas paralelas.

VIII.3 – Delineamento Fatorial com Repetição de Provas

É necessário fazer provas paralelas, como já foi dito, quando o número de fatores for menor que 4, para poder estimar o erro experimental. No Exercício 52, vimos que o fator C não é significativo. O experimento pode ser então considerado como tendo 3 fatores com 2 níveis, com repetição de cada tratamento.

EXERCÍCIO 53

Estudo da influência dos fatores A, B e D no rendimento da produção de UF_4, com provas paralelas.

Teremos então os resultados dispostos como na Tabela VIII.7.

Tabela VIII.7

Redução dos Resultados da Tabela VIII.4 a 3 Fatores com Repetição de Provas

A_1				A_2			
B_1		B_2		B_1		B_2	
D_1	D_2	D_1	D_2	D_1	D_2	D_1	D_2
1,2 2,0	2,5 2,8	3,0 3,0	3,4 6,2	5,0 5,1	5,5 6,9	5,9 6,7	6,8 6,5
(1)=3,2	D = 5,3	B = 6,0	BD = 9,6	A = 10,1	AD = 12,4	AB = 12,6	ABD = 13,3

Calcula-se o efeito total de cada fator e interação pelo método de Yates conforme a Tabela VIII.8.

Para ter uma estimativa do erro experimental, calcula-se a diferença entre a soma dos quadrados das 16 observações e a soma dos quadrados dos totais das 2 observações obtidas para cada combinação dividida por 2

$$(1{,}2^2 + 2{,}5^2 + \ldots + 6{,}7^2 + 6{,}5^2) - (3{,}2^2 + 5{,}3^2 + \ldots + 12{,}6^2 + 13{,}3^2)/2 = 5{,}635$$

Tabela VIII.8

Cálculo do Efeito dos Fatores e Interações Pelo Método de Yates

Combinação Fatorial	Rendimento (− 90%)	Col 1	Col 2	Col 3	Efeito Médio (Col 3)/(4x2)	Soma dos Quadrados $(Col\ 3)^2/(8x2)$
(1)	3,2	13,4	32,9	—		
A	10,1	18,6	40,6	24,3	3,0375	36,9056
B	6,0	17,7	13,5	10,5	1,3125	6,8906
AB	12,6	22,9	10,8	− 3,7	− 0,4625	0,8556
D	5,3	6,9	5,3	7,7	0,9625	3,7056
AD	12,4	6,6	5,2	− 2,7	− 0,3375	0,4556
BD	9,6	7,1	− 0,3	− 0,1	0,0125	0,0006
ABD	13,3	3,7	− 3,4	− 3,1	− 0,3875	0,6006

Com esses valores, aplica-se a análise da variância conforme a Tabela VIII.9.

Tabela VIII.9

Análise da Variância dos Resultados da Tabela VIII.7

Fonte de Variação	Nº de Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
A	(a − 1) = 1	36,9056		52,4
B	(b − 1) = 1	6,8906		9,8
D	(d − 1) = 1	3,7056		5,26
AB	(a − 1) (b − 1) = 1	0,8556		
AD	(a − 1) (d − 1) = 1	0,4556		
BD	(b − 1) (d − 1) = 1	0,0006		
ABD	(a − 1) (b − 1) (d − 1) = 1	0,6006		
Erro Experim.	8	5,6350	0,7044	$F_{0,05(1,8)} = 5,3$
Total	15			$F_{0,01(1,8)} = 11,3$

Estes resultados confirmam aqueles do delineamento com 4 fatores, a não ser para o fator D, cujo valor de F é ligeiramente inferior ao do nível de significância 0,05. Para aumentar a sensibilidade do teste, pode-se somar a variância das interações não significativas ao erro experimental e assim temos:

$$\text{Erro Experimental} = \frac{5,6350 + 0,6006 + 0,0006 + 0,4556 + 0,8556}{12} = 0,6289$$

Com este novo valor do erro experimental, os 3 fatores A, B e D são francamente significativos, porque o valor de $F_{0,05(1,12)} = 4,8$ e o F correspondente a D se torna: 5,89.

EXERCÍCIO 54

Foi estudada a influência dos fatores A, B e C no fator de separação entre 2 elementos, por extração com solvente. A relação entre as concentrações dos 2 elementos na solução inicial era 1 e, na fase aquosa, depois da extração, foram obtidos os resultados da Tabela VIII.10, dos quais foi subtraído 130 para simplificar os cálculos.

Tabela VIII.10

Resultados da Separação de 2 Elementos para Estudar a Influência dos Fatores A, B e C

A_1				A_2			
B_1		B_2		B_1		B_2	
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
0,1 0,3	3,8 3,4	0,9 1,1	4,3 3,7	2,6 2,4	3,7 3,5	2,2 1,8	4,9 5,1
(1) = 0,4	C2 = 7,2	B = 2,0	BC = 8,0	A = 5,0	AC = 7,2	AB = 4,0	ABC = 10,0

Estimativa do erro experimental

$$(0,1^2 + 0,3^2 + \ldots + 4,9^2 + 5,1^2) - (0,4^2 + 7,2^2 + \ldots + 10^2)/2 = 0,44$$

Aplicando-se o método de Yates, conforme foi feito no exercício anterior, acha-se a soma dos quadrados correspondente a cada fator e a cada interação e aplica-se a análise da variância que está na Tabela (VIII.11).

Conclui-se que os 3 fatores são altamente significativos. A interação BC não é significativa, se considerarmos um nível de significância de 0,01, enquanto que a interação AC é importante mesmo neste nível. Pode-se aumentar a sensibilidade do teste incluindo no erro experimental as interações de efeito desprezível e vamos obter:

$$s_1^2 = \frac{0,44 + 0,0225 + 0,0225}{10} = 0,0485$$

Nestas condições, a interação BC torna-se significativa mesmo num nível de 0,01 porque $F_{exp} = 0,56/0,0485 = 11,5$ e $F_{0,01(1,10)} = 10,0$.

Tabela VIII.11

Análise da Variância dos Resultados da Tabela VIII.10

Fonte de Variação	Nº de Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
A	1	4,62		84*
B	1	1,10		20*
C	1	27,56		50,1*
AB	1	0,0225		0,40
AC	1	1,32		24*
BC	1	0,56		10,2**
ABC	1	0,0225		0,40
Erro Experim.	8	0,44	0,055	
Total	15		$F_{0,05(1,8)} = 5,3$; $F_{0,01(1,8)} = 11,3$	

Concluímos que os fatores B e C e os fatores A e C não se comportam de maneira independente, isto é, a variação de C afeta o comportamento de A e o comportamento de B. Examinando-se os totais dos resultados onde C está no nível C_1 ou C_2, seguido de A_1 ou A_2 ou também B_1 e B_2 achamos:

(1)

	C_1	C_2	$C_2 - C_1$
A_1	2,4	15,2	12,8
A_2	9,0	17,2	8,2

(2)

	C_1	C_2	$C_2 - C_1$
B_1	5,4	14,4	9,0
B_2	6,0	18,0	12,0

Vemos em (1), que a passagem do nível C_1 para C_2, quando A está no nível A_1, aumenta de 12,8 unidades o fator de separação, enquanto que, no nível A_2, este aumento é de 8,2 unidades. Esta interação está representada graficamente na Figura 8.1. Se a influência do fator C fosse a mesma nos 2 níveis de A, as duas retas seriam paralelas. A existência da interação AC se demonstra pela divergência entre as duas retas.

Fazendo um gráfico com a interação BC (Figura 8.2) também vamos obter retas divergentes o que confirma a existência desta interação.

Pode-se interpretar os resultados das interações AC e BC, dividindo o conjunto em 2 partes: examina-se, na primeira, a influência de A e B com C no nível C_1 e, na segunda, com C no nível C_2, conforme a Tabela VIII.12.

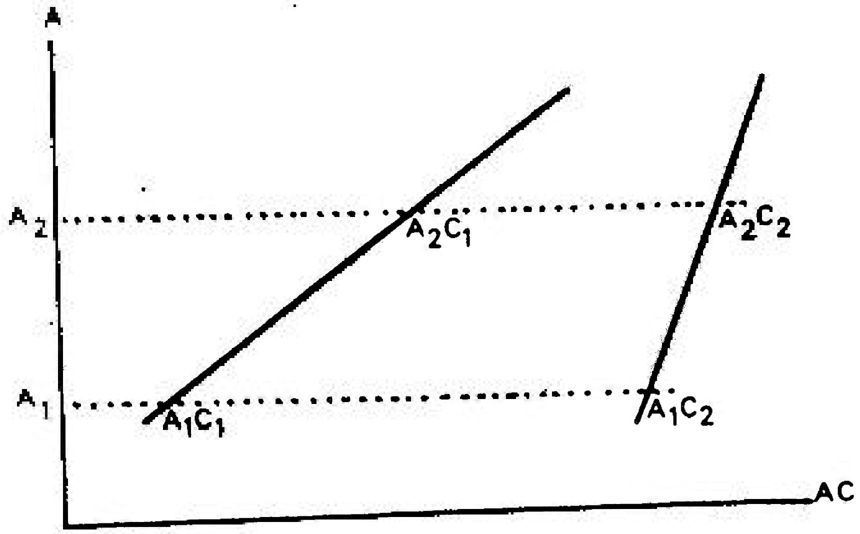

Figura 8.1 – Interação AC

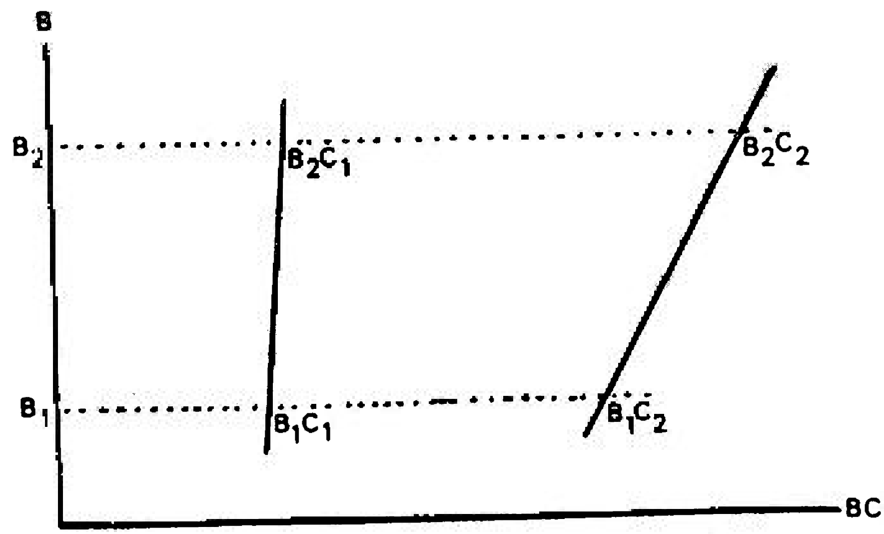

Figura 8.2 – Interação BC

Tabela VIII.12

Resultados Experimentais Divididos em 2 Grupos Para Estudar as Interações AC e BC

	C_1				C_2			
	A_1		A_2		A_1		A_2	
	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
	0,1	0,9	2,6	2,2	3,8	4,3	3,7	4,9
	0,3	1,1	2,4	1,8	3,4	3,7	3,5	5,1
Total	0,4	2,0	5,0	4,0	7,2	8,0	7,2	10,0

Com C no nível C_1 temos os seguintes valores para os efeitos médios de A, B, e AB e também para a soma dos quadrados.

$$\overline{A} = 1/4\,[(5+4) - (2{,}0+0{,}4)] = 1{,}66$$

$$SQA = (1{,}66 \times 4)^2/8 = 5{,}5112$$

$$\overline{B} = 1/4\,[(2+4) - (5{,}0+0{,}4)] = 0{,}1$$

$$SQB = (0{,}1 \times 4)^2/8 = 0{,}02$$

$$\overline{AB} = 1/4\,[(4-5) - (2-0{,}4)] = -0{,}65$$

$$SQAB = (0{,}65 \times 4)^2/8 = 0{,}845$$

Com C no nível C_2:

$$\overline{A} = 1/4\,[(7{,}2+10{,}0) - (7{,}2+8)] = 0{,}5$$

$$SQA = (0{,}5 \times 4)^2/8 = 0{,}5$$

$$\overline{B} = 1/4\,[(8+10) - (7{,}2+7{,}2)] = 0{,}9$$

$$SQB = (0{,}9 \times 4)^2/8 = 1{,}62$$

$$\overline{AB} = 1/4\,[(10-7{,}2) - (8{,}0-7{,}2)] = 0{,}5$$

$$SQAB = (0{,}5 \times 4)^2/8 = 0{,}5$$

Os resultados podem ser interpretados pela análise da variância conforme a Tabela VIII.13.

Tabela VIII.13

Análise da Variância nos Níveis C_1 e C_2

Fonte de Variação	Graus de Liberdade	C no Nível C_1 Quadrado Médio	F	C no Nível C_2 Quadrado Médio	F
A	1	5,5112	157,5	0,5	6,7
B	1	0,02	0,6	1,62	21,6
AB	1	0,845	24,1	0,5	6,7
Erro Experim.	4	0,035		0,075	
					$F_{0,05(1,4)} = 7,7$
					$F_{0,01(1,4)} = 21,2$

Estimativa do erro experimental do conjunto no nível C_1:

$$(0,1^2 + 0,3^2 + \ldots + 1,8^2) - (0,4^2 + \ldots + 4,0^2)/2 = 0,14$$

Estimativa do erro experimental do conjunto no nível C_2:

$$(3,8^2 + 3,4^2 + \ldots + 5,1^2) - (7,2^2 + \ldots + 10^2)/2 = 0,3$$

A análise da variância mostra que o fator A é altamente significativo no nível C_1, enquanto que no nível C_2 não afeta os resultados. Quanto ao fator B, observa-se o contrário; ele é significativo no nível C_2. Por este quadro de análise de variancia, observa-se que existe interação AB no nível C_1.

VIII.4 – Fatores com mais de 2 Níveis

Os fatores que podem influir no comportamento de um processo qualquer podem ser classificados como quantitativos (temperatura, pH, pressão, etc..) e qualitativos (natureza do solvente, tipo de resina, etc.).

VIII.4.1 – 1º Caso: Fatores Qualitativos

No caso de fatores qualitativos, toma-se como nível mais baixo o de uso rotineiro. Por exemplo, se, num determinado trabalho, usa-se sempre um solvente S e queremos testar comparativamente a eficiência de outros dois solventes, consideramos o primeiro como nível mais baixo S_1 e os outros dois como S_2 e S_3.

Suponhamos que um fator A tenha 3 níveis. Como no caso de 2 níveis, o efeito principal do fator A é dado pela diferença entre a média dos resultados onde aparece A_3 e a média onde aparece A_1, isto é, $(\bar{A}_3 - \bar{A}_1)$ e também pelas diferenças $(\bar{A}_3 - \bar{A}_2)$ e $(\bar{A}_2 - \bar{A}_1)$.

EXERCÍCIO 55

Deseja-se estudar a influência de 3 fatores qualitativos na recuperação de urânio de uma solução complexa. Estes fatores são:

A – 3 tipos de resinas para retenção seletiva de urânio de uma solução
B – 2 soluções eluentes (mesmo volume) de composição diferente
C – 2 processos de precipitação do diuranato de amônio.

O diuranato de amonio depois de precipitado foi filtrado e dissolvido em 100 l de HNO_3 de concentração conveniente. Em todos os tratamentos foi dosada a concentração final de urânio na solução e verificou-se que a pureza do urânio, em todos os casos, foi satisfatória. Os resultados obtidos em g/l, dos quais foi deduzido 10 estão na Tabela VIII.14. Deseja-se saber quais os fatores e interações entre fatores que favorecem o rendimento.

Tabela VIII.14

Concentração de U (g/1 – 10) em Soluções

	A_1		A_2		A_3	
	B_1	B_2	B_1	B_2	B_1	B_2
C_1	1,8	2,1	2,4	2,7	2,9	3,3
C_2	2,0	2,0	2,5	2,6	2,8	3,1

Para calcular os efeitos principais dos fatores A, B e C e as interações entre 2 fatores, desdobra-se a Tabela (VIII.14) em 3 Tabelas menores: A x B, A x C e B x C. Na Tabela A x B, os resultados dos 2 níveis de C são somados para cada combinação de A e B. Na Tabela A x C, somam-se os 2 níveis de B e, na Tabela B x C, somam-se os 2 níveis de A.

Tabela A x B

	A_1	A_2	A_3	Soma	Média
B_1	3,8	4,9	5,7	14,4	2,4
B_2	4,1	5,3	6,4	15,8	2,633
Soma	7,9	10,2	12,1		
Média	1,975	2,55	3,025		

Tabela A x C

	A_1	A_2	A_3	Soma	Média
C_1	3,9	5,1	6,2	15,2	2,533
C_2	4,0	5,1	5,9	15,0	2,5
Soma	7,9	10,2	12,1		
Média	1,975	2,55	3,025		

Tabela B x C

	B_1	B_2	Soma	Média
C_1	7,1	8,1	15,2	2,533
C_2	7,3	7,7	15,0	2,5
Soma	14,4	15,8		
Média	2,4	2,633		

Para a Tabela A x B, calculam-se:

a) Soma dos quadrados dos 6 resultados, dividida por 2 (2 níveis de C): $(3{,}8^2 + 4{,}1^2 + \ldots + 6{,}4^2)/2 = 78{,}4$

b) Fator de correção:

$(1{,}8 + 2{,}0 + 2{,}1 + \ldots + 3{,}3 + 3{,}1)^2/12 = 76{,}00$

c) Soma dos quadrados totais:

$SQT = 78{,}4 - 76{,}0 = 2{,}4$

d) Soma dos quadrados de A: $(7{,}9^2 + 10{,}2^2)/4 - 76{,}0 = 2{,}215$

e) Soma dos quadrados de B: $(14{,}4^2 + 15{,}8^2)/6 - 76{,}0 = 0{,}167$

A análise da variância dos resultados da Tabela A x B deve ser disposta conforme a Tabela VIII.15.

Tabela VIII.15

Análise da Variância dos Resultados da Tabela A x B

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Fator A	2,215	2	1,107	123
Fator B	0,167	1	0,167	18,5
Interação AB	0,018	2	0,009	
Total	2,4	5		

Procede-se do mesmo modo com as outras duas Tabelas para calcular a soma dos quadrados correspondentes a C, a AC e a BC. Acha-se:

Soma dos quadrados de C = 0,007

Soma dos quadrados de AC = 0,018

Soma dos quadrados de BC = 0,026

Para ser possível uma análise da variância total é preciso calcular a soma dos quadrados totais a partir dos 12 resultados da Tabela VIII.14.

$$(1{,}8^2 + 2{,}0^2 + \ldots + 3{,}3^2 + 3{,}1^2) - 76{,}00 = 2{,}46$$

A interação ABC é obtida por diferença. A análise da variância total está na Tabela VIII.16.

Tabela VIII.16

Análise da Variância dos Resultados da Tabela VIII.14

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F
Fator A	2,215	2	1,1075	113
Fator B	0,167	1	0,167	17,1
Fator C	0,007	1	0,007	
Interação AB	0,018	2	0,009	
Interação AC	0,018	2	0,009	
Interação BC	0,026	1	0,026	
Interação ABC	0,009	2	0,0045	
Total	2,460	11	$S^2_{comb} = 0{,}00975$	

Os valores de F achados para o fator C e para as interações AB, AC e BC não são significativos. Assim as somas de quadrados de C, AB, AC e BC podem ser adicionados à soma de quadrados de ABC e o total dividido pelo número de graus de liberdade obtido por esta soma

$$S^2_{comb} = \frac{0{,}007 + 0{,}018 + 0{,}018 + 0{,}026 + 0{,}009}{8} = 0{,}00975$$

O valor de $F_{0,001(1,8)} = 11{,}3$, portanto a influência de A e B é significativa mesmo neste nível, no rendimento do urânio.

Conclusão:

Para cada média de A, o erro experimental é 0,00975/4 = 0,00244 e o da diferença entre duas médias é: (2 x 0,00975)/4 = 0,00488. O valor de t para p = 0,05 e 8 graus de liberdade é 2,31. Então a menor diferença significativa entre as médias de A é:

$$0{,}00488 \times 2{,}31 = 0{,}011$$

As médias obtidas para A foram:

$$\bar{A}_1 = 1{,}97$$

$$\bar{A}_2 = 2{,}55$$

$$\bar{A}_3 = 3{,}02$$

e concluímos que o melhor rendimento é obtido com a resina (3), depois com a (2) e finalmente com a (1) que era usada nos trabalhos de rotina. No caso do fator B, temos só duas médias:

$$\bar{B}_1 = 2{,}40$$

$$\bar{B}_2 = 2{,}63$$

Como o teste deu significativo, concluímos que a composição do 2º eluente é mais eficiente que a composição do primeiro.

VIII.4.2 – Fatores Qualitativos e Quantitativos com mais de 2 Níveis

Quando se testam fatores quantitativos é conveniente, para simplificar a análise da variância, espaçar os níveis em quantidades iguais.

EXERCÍCIO 56

Como conseqüência do Exercício 55 foi escolhida a resina nº 3 por ser mais eficiente e se quer saber agora se a temperatura do sistema influe nos resultados. Ao mesmo tempo, deseja-se testar outros

2 tipos de eluentes simultaneamente com aquele que deu maior rendimento no caso anterior. Sejam $A_1, \ldots, A_4$ os níveis de temperatura, B_1 o eluente usado no problema anterior e B_2 e B_3 os novos eluentes usados. A parte final do processo é a mesma do problema anterior. Os resultados obtidos dos quais foram deduzidos 10 g/l estão na Tabela VIII.17.

Tabela VIII.17

Concentração de U em Soluções (g/1 – 10)

Eluentes	Temperatura				Total	Média
	A_1 = 25°C	A_2 = 45°C	A_3 = 65°C	A_4 = 85°C		
B_1	3,0	3,1	3,4	3,6	13,2	4,4
B_2	2,9	3,0	3,4	3,5	12,8	4,267
B_3	3,1	2,8	3,5	3,6	13,0	4,333
Total	9,0	8,9	10,4	10,7		
Média	2,25	2,225	2,6	2,675		
Coeficientes:						
Linear	− 3	− 1	+ 1	+ 3		
Quadrativo	+ 1	− 1	− 1	+ 1		
Cúbico	− 1	+ 3	− 3	+ 1		

Procede-se como no caso anterior como se os fatores fossem qualitativos. Soma dos quadrados de todos os resultados:

$$3{,}0^2 + 2{,}9^2 + \ldots + 3{,}5^2 + 3{,}6^2 = 127{,}7$$

Fator de correção:

$$(3{,}0 + 2{,}9 + \ldots + 3{,}5 + 3{,}6)^2/12 = 136{,}75$$

Soma dos quadrados totais:

$$SQT = 127{,}7 - 126{,}75 = 0{,}95$$

Soma dos quadrados correspondentes a A:

$$SQA = (9{,}0^2 + 8{,}9^2 + 10{,}4^2 + 10{,}7^2)/3 - 126{,}75 = 0{,}87$$

Soma dos quadrados correspondentes a B:

$$SQB = (13,2^2 + 12,8^2 + 13,0^2)/4 - 126,75 = 0,02$$

Soma dos quadrados correspondentes a AB:

$$SQAB = 0,95 - (0,87 + 0,02) = 0,06$$

A estimativa do erro é feita pela interação AB, porque só se dispõe de um resultado para cada tratamento. Pela análise da variância (Tabela VIII.18) observa-se que a variação da temperatura tem influência significativa, enquanto que os 2 novos eluentes são igualmente eficientes ao melhor do experimento anterior.

Tabela VIII.18

Análise da Variância dos Resultados da Tabela VIII.17

Fonte de Variação	Soma dos Quadrados	Nº de Graus de Liberdade	Quadrado Médio	F
Temperaturas	0,87	3	0,29	29
Eluentes	0,02	2	0,01	
Interação AB	0,06	6	0,01	
Total	0,95	11		

A análise da variancia no caso de fatores quantitativos pode ser extendida com o exame do efeito da variação de temperatura em cada eluente e verificando se a relação temperatura-resultado difere de um eluente para outro. Como as temperaturas são igualmente espaçadas, em lugar das temperaturas podem ser usados coeficientes polinomiais que permitem subdividir, para cada nível do fator qualitativo B, a soma dos quadrados de A em um componente linear (que se incumbe do ajuste linear das respostas às temperaturas), um componente quadrático e um cúbico (que traduzem uma curvatura possível da reta). Estes coeficientes são encontrados em Tabelas em função do número de pontos n. (Fischer et Yates – "Statistical Tables for Biological, Agricultural and Medical Research" – Oliver and Boyd, London, 1949).

Coeficientes Polinomiais

No caso da relação entre as variáveis x e y ser descrita pelo polinômio:

$$Y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots \qquad (VIII.3)$$

o polinômio pode ser substituído pela equação:

$$Y' = b'_0\, x_0 + b'_1\, x_1 + b'_2\, x_2 + b'_3\, x_3 \qquad \text{(VIII.4)}$$

onde os diferentes valores x, seus quadrados, cubos, etc. são substituídos pelos coeficientes:

x_1 = coeficiente linear usado no cálculo de b_1, no ajuste linear

x_2 = coeficiente quadrático, usado no cálculo de b_2

x_3 = coeficiente cúbico, usado no cálculo de b_3, etc.

O coeficiente x_0 é igual a 1. Estes coeficientes permitem calcular as constantes b_i de cada termo do polinômio e a soma dos quadrados correspondente a este termo.

As constantes b' da equação (VIII.4) são dadas por:

$$b'_i = \frac{\Sigma\, x_i\, T_d}{n\, \Sigma\, x_i^2} \qquad \text{(VIII.5)}$$

A soma dos quadrados correspondente a este termo é:

$$SQ = \frac{(\Sigma\, x_i\, T_d)^2}{n\, \Sigma\, x_i^2} \qquad \text{(VIII.6)}$$

e a variância é:

$$S^2_{b'} = \frac{S^2}{n\, \Sigma\, x_i^2} \qquad \text{(VIII.7)}$$

onde S^2 é a variância do erro experimental e T_d é a soma parcial.

Voltando ao problema, calculam-se sucessivamente, para cada eluente, a soma dos produtos dos coeficientes pela resposta correspondente. Por exemplo, para B_1, a soma dos produtos correspondentes ao termo linear é:

$$(-3 \times 3) + (-1 \times 3{,}1) + (1 \times 3{,}5) + (3 \times 3{,}6) = 2{,}2$$

Os termos lineares, quadráticos e cúbicos estão na Tabela (VIII.19)

Os divisores são as somas dos quadrados dos coeficientes polinomiais.

A soma dos quadrados correspondentes às diferenças de temperaturas poderá ser sub-dividida em 3 componentes que correspondem aos termos linear, quadrático e cúbico. Isto vai permitir decidir se a relação temperatura-resultado pode ser descrita por uma reta ou por uma curva de 2º ou 3º grau.

Tabela VIII.19

Termos Lineares, Quadráticos e Cúbicos

Fator Qualitativo	Soma dos Produtos		
	Termo Linear	Termo Quadr.	Termo Cúbico
B_1	2,2	0	− 0,6
B_2	2,2	0	− 0,5
B_3	2,2	0,4	− 1,6
Total	6,6	0,4	− 2,7
Divisor	20	4	20

A soma dos quadrados do termo linear é:

$$SQL = 6{,}6^2/(20 \times 3) = 0{,}726$$

onde 20 é o divisor e 3 é o número de níveis do fator qualitativo (eluente). Para o termo quadrático temos:

$$SQQ = 0{,}4^2/(4 \times 3) = 0{,}0133$$

e para o cúbico:

$$SQC = (-2{,}7)^2/(20 \times 3) = 0{,}1215$$

A soma de SQC, SQQ e SQL deve ser igual à soma dos quadrados correspondente às temperaturas, a menos de diferenças de aproximação. A soma dos quadrados da interação AB também pode, do mesmo modo, ser subdividida em 3 componentes:

Interação do termo linear A_ℓ com B = $A_\ell B$

Interação do termo quadrático A_q com B = $A_q B$

Interação do termo cúbico A_c com B = $A_c B$

As variações de A_ℓ para cada nível de B estão na 2a. coluna da Tabela VIII.19.

A soma dos quadrados que corresponde à interação $A_\ell B$ é:

$$SQA_\ell B = [(2{,}2)^2 + (2{,}2)^2 + (2{,}2)^2 - (6{,}6)^2/3] / 20 = 0$$

Para A_qB temos:

$$SQA_qB = [(0{,}4)^2 - (0{,}4)^2/3] / 4 = 0{,}0268$$

e para A_cB:

$$SQA_cB = [(-0{,}6)^2 + (-0{,}5)^2 + (-1{,}6)^2 - (-2{,}7)^2/3] / 20 = 0{,}037$$

A soma destes 3 termos deve ser igual à soma dos quadrados da interação AB. Com esses resultados, faz-se a análise da variância, conforme a Tabela VIII.20.

Tabela VIII.20

Fonte de Variação	Soma de Quadrados	Número de Graus de Liberdade	Quadrado Médio	F
Eluentes (B)	0,0200	2	0,0100	0,94
Temperatura (A):				
Termo linear	0,7260	1	0,7260	68,5
Termo quadrático	0,0133	1	0,0133	1,3
Termo cúbico	0,1215	1	0,1215	11,5
Interações:				
$A_\ell B$	0,0000	2	0,0638/6 =	
A_qB	0,0268	2	0,0106	
A_cB	0,0370	2		
	0,9446			

Valores de $F_{Tabelado}$ $\begin{cases} F_{0,05(1,6)} = 6{,}0 \\ F_{0,01(1,6)} = 13{,}7 \end{cases}$

Como só temos um resultado para cada tratamento, tomamos como erro experimental a interação que tem menos chance de ser verdadeira que é A_cB. Visto que $A_\ell B$ e A_qB são menores que A_cB, pode-se reunir as 3 interações e temos o valor 0,0638 com 6 graus de liberdade que dá um quadrado médio igual a 0,0106.

Pelos valores de F obtidos (Tabela VIII.20), vê-se que o termo linear é preponderante, se bem que o termo cúbico é significativo num nível de 0,05.

Pode-se concluir que não há diferença entre os 3 eluentes sobre o rendimento de urânio e que a influência da temperatura no intervalo examinado é aproximadamente linear.

VIII.4.3 – 3º Caso – Fatores Quantitativos com mais de 2 Níveis

Quando todos os fatores são quantitativos, pode-se fazer uma análise mais detalhada dos resultados.

Procede-se como no caso anterior, só que, como os 2 fatores são quantitativos, aplicam-se os coeficientes polinomiais para os dois fatores, decompondo-se a soma dos quadrados nas contribuições dos termos linear, quadrático e cúbico.

Se os fatores são A e B, vamos ter interações do tipo $A_\ell B_\ell$, $A_\ell B_q$, $A_q B_\ell$ etc.

As interações da terceira ordem $A_\ell B_\ell$, $A_q B_c$, $A_c B_\ell$, $A_c B_\ell$ e $A_c B_q$ são somadas e a soma dos quadrados correspondentes é usada como estimativa do erro experimental. Isto é possível, porque este tipo de interações é muito pouco provável.

Com este tipo de interpretação dos resultados pode-se não só estabelecer se existe interação entre dois fatores, mas também como se comporta esta interação.

CAPÍTULO IX

DELINEAMENTO FATORIAL FRACIONADO

Quando um experimento envolve vários fatores com mais de 2 níveis a sua execução pode exigir um período longo de tempo, pode ser necessário o uso de mais de um aparelho para provas paralelas, pode também exigir lotes diferentes de matéria prima (casos industriais) etc. As diferenças introduzidas pelas variações de condições com o tempo, com os aparelhos ou com as matérias primas vão aumentar o erro experimental e portanto diminuir a sensibilidade do teste.

Existe um método, "confounding" em inglês, que permite separar os erros introduzidos por fatores deste tipo. Baseia-se na distribuição sistemática do conjunto de provas em um certo número de blocos para confundir os erros citados com interações de ordem elevada, isto é, com pouca chance de existir.

Como exemplo, vamos considerar um caso simples de 3 fatores com 2 níveis. Para resolver o problema são necessários, no mínimo, 8 tratamentos. Na execução dos tratamentos são usados 2 aparelhos semelhantes mas que podem dar resultados ligeiramente diferentes. É possível distribuir os 8 tratamentos em 2 blocos de tal maneira que a diferença causada pelos 2 aparelhos (blocos) se confunda com a interação entre os 3 fatores. Esta distribuição racional é a da Tabela IX.1. Para calcular a interação ABC desenvolve-se o polinômio:

$$(A-1)\,(B-1)\,(C-1) = (ABC + C + B + A) - (BC + AC + AB + (1))$$

que é precisamente a diferença entre os 2 aparelhos, isto é, esta interação se confunde com a interação ABC.

Vamos supor que exista uma diferença sistemática $\underline{x}$ entre o 1º e o 2º aparelho. Esta diferença poderia ser causada por uma tendência do 2º aparelho dar resultados mais altos. Com este tratamento, esta diferença sistemática desaparece. O efeito total do fator A, por exemplo, que se obtém desenvolvendo $(A-1)\ (B+1)\ (C+1)$, levando em conta a diferença $\underline{x}$ seria:

$$A+(AB+x)+(AC+x)+ABC-B-C-(BC+x)-((1)+x) = A+AB+AC+ABC - (B+C+BC+(1))$$

Por esta expressão vemos que o erro sistemático desaparece. O mesmo acontece para todos os fatores e interações. A ordem da execução das provas deve ser sorteada para manter o efeito de aleatoriedade.

Tabela IX.1

Distribuição de um Delineamento Fatorial de 3 Fatores com 2 Níveis em 2 Blocos

Aparelho (1)	Aparelho (2)
A	AB
B	AC
C	BC
ABC	(1)

IX.1 – Aplicação a um Conjunto de Tratamentos 2^4

No caso de 4 fatores com 2 níveis, podemos ter 2 blocos de 8 tratamentos ou 4 de 4 tratamentos. Se a distribuição for feita em 2 blocos, usa-se a interação ABCD para ser confundida com a diferença entre os blocos. Desenvolve-se o polinômio $(A - 1)\ (B - 1)\ (C - 1)\ (D - 1)$ e colocam-se num bloco os termos positivos e no outro os termos de sinal negativo. No caso de se fazer a distribuição em 4 blocos, 3 interações devem ser usadas para serem confundidas com o efeito dos blocos. Neste caso deve-se observar que:

a) Devem ser confundidas interações de ordem elevada, porque há pouca chance de que existam.

b) Quando se confundem, 3 interações, só duas podem ser escolhidas livremente, porque a terceira resulta da multiplicação algébrica das outras duas, igualando a 1 todas as letras com potência 2. Se escolhermos a interação ABCD (de ordem mais elevada) e uma de 3a. ordem, ABC por exemplo, a terceira seria obrigatoriamente dada por:

$$ABCD \times ABC = A^2B^2C^2D = D$$

Esta escolha não é conveniente porque levaria a confundir o fator principal D. Para que isto não aconteça, devem-se escolher 2 interações de 3 fatores e a terceira será de 2 fatores. Aqui deve entrar o bom senso do pesquisador para escolher qual das interações de 2 fatores tem menos chance de existir, do ponto de vista técnico ou científico. Suponhamos que um exame do problema tenha permitido concluir que a interação AB é pouco provável. Devem-se escolher então as interações ACD e BCD cujo produto vai dar AB.

Para determinar quais os tratamentos que devem aparecer nos 4 blocos, desenvolvem-se os polinômios correspondentes a duas das interações escolhidas. Por exemplo:

$$ACD = (A-1)(C-1)(D-1)(B+1)$$

$$ABCD + ACD + AB + BC + BD + A + C + D - (ABC+ABD+BCD+AC+AD+CD+B+(1))$$

$$BCD = (A+1)(B-1)(C-1)(D-1)$$

$$ABCD+BCD+AB+AC+AD+B+C+D - (ABC+ABD+ACD+BC+BD+CD+A+(1))$$

Colocam-se num mesmo bloco os tratamentos que tem nos 2 grupos os sinais (+,+), (–,–) (+,–) e (–,+). No nosso caso, a distribuição seria a da Tabela IX.2.

Tabela IX.2

Distribuição de um Conjunto de Provas 2^4 em 4 Blocos

Bloco I (+, +)	Bloco II (–, –)	Bloco III (+, –)	Bloco IV (–, +)
C	(1)	A	B
D	CD	BC	AC
AB	ABC	BD	AD
ABCD	ABD	ACD	BCD

O bloco principal é o que contém o elemento (1)

EXERCÍCIO 57

No estudo da separação de um elemento M por meio de resinas foram levados em conta 4 fatores:

A – 2 concentrações de M na solução de percolação: A_1, A_2.

B – 2 tipos de resina (mesma capacidade).

C – 2 níveis de temperatura: C_1, C_2.

D – 2 velocidades de percolação: D_1, D_2.

Fazer as 16 provas em seqüência seria muita perda de tempo. Foram então preparadas 4 colunas semelhantes com a mesma massa de resina o que permitiu conduzir 4 tratamentos simultaneamente. A interação de 2a. ordem que pareceu menos provável foi BD, isto é, tipo de resina e velocidade de percolação. Uma das interações de terceira ordem pode ser ACD. A outra será obrigatoriamente:

$$ACD \times BD = ABC$$

Aplicando a técnica de distribuição vista, obtém-se o quadro de tratamentos da Tabela IX.3.

Tabela IX.3

Distribuição de Tratamentos em 4 Blocos
Confusão das Diferenças entre Blocos com as Interações BD, ACD, e ABC

Coluna I	Coluna II	Coluna III	Coluna IV
A	(1)	B	D
C	AC	AD	AB
BD	ABD	CD	BC
ABCD	BCD	ABC	ACD

O eluente de M retido na resina foi o mesmo nos 4 blocos, tanto em composição como em volume e depois foi determinada a concentração de M no eluído. A ordem cronológica dos tratamentos foi sorteada. Os resultados em g/l, dos quais foram deduzidos 10 g/l estão na Tabela IX.4.

Tabela IX.4

Resultados dos Tratamentos de um Conjunto 2^4 Distribuídos em 4 Blocos

Tratamentos		A_1		A_2	
		B_1	B_2	B_1	B_2
C_1	D_1	2,56	2,48	1,44	1,60
	D_2	0,92	1,10	0,60	0,68
C_2	D_1	2,82	2,70	1,72	1,64
	D_2	1,20	1,24	0,74	0,80

Aplicando o método de Yates, calculam-se o efeito médio e a soma dos quadrados. O conjunto de resultados obtidos por esse cálculo está na Tabela IX.5.

A análise da variância se encontra na Tabela IX.6, onde, no erro, estão englobadas as interações ABCD, BCD e ABD.

Comparando os valores de F experimentais com os valores de F tabelados, concluímos que o fator A (concentração de M na solução de percolação), o fator C(temperatura), o fator D (velocidade de percolação) e a interação AD (concentração x velocidade de percolação) são altamente significativos. Conclui-se também que os dois tipos de resina dão o mesmo rendimento de separação, porque o fator B no teste deu um resultado não significativo.

Tabela IX.5

Efeito Médio e Soma dos Quadrados dos Tratamentos

Tratamento	Coluna nº	Concentração g/l – 10	Efeito Médio	Soma de Quadrados
(1)	II	2,56	–	–
A	I	1,44	– 0,7375	2,175625
B	III	2,48	0,0175	0,001225
AB	IV	1,50	0,0125	0,000625
C	I	2,82	0,1975	0,156025
AC	II	1,72	– 0,0275	0,003025
BC	IV	2,70	–0,0425	0,007225
ABC	III	1,64	(0,0025)	(0,000025)
D	IV	0,92	– 1,0725	4,601025
AD	III	0,60	0,3275	0,429025
BD	I	1,10	0,0725	(0,021025)
ABD	II	0,68	– 0,0325	0,004225
CD	III	1,20	– 0,0275	0,003025
ACD	IV	0,74	(– 0,0125)	(0,000625)
BCD	II	1,24	0,0025	0,000025
ABCD	I	0,80	0,0275	0,003025
Total		24,14		

Totais por Blocos: I) 6,16; II) 6,20; III) 5,92; IV) 5,86

Tabela IX.6

Análise da Variância dos Resultados Experimentais da Tabela IX.4

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio (Variância)	F
Dif. entre colunas	(0,021675)	3	0,007225	2,98
Fator A	2,175625	1	2,175625	897 *
Fator B	0,001225	1	0,001225	0,50
Fator C	0,156025	1	0,156025	64,3 *
Fator D	4,601025	1	4,601025	1897 *
Interação AB	0,000625	1	0,000625	0,25
Interação AC	0,003025	1	0,003025	1,24
Interação AD	0,429025	1	0,429025	176,9 *
Interação BC	0,007225	1	0,007225	2,98
Interação CD	0,003025	1	0,003025	1,25
Erro	0,007275	3	0,002425	

Valores tabelados de F : $F_{0,01(3,3)} = 29,5$ $F_{0,05(3,3)} = 9,3$

$F_{0,01(1,3)} = 34,1$ $F_{0,05(1,3)} = 10,1$

IX.2 – "Confusão" Parcial

Quando um delineamento fatorial só tem 3 ou 4 fatores, não é conveniente usar as interações de ordem elevada como estimativa do erro, porque o número de graus de liberdade é pequeno e o teste se torna pouco sensível. É mais aconselhável repetir os tratamentos algumas vezes para se ter uma estimativa direta do erro. Se, por outro lado, as condições exigirem que os tratamentos sejam divididos em dois blocos, pode-se escolher interações diferentes para serem confundidas com a diferença entre os blocos, em cada série de tratamentos. Assim é possível obter uma estimativa das interações confundidas, com menos sensibilidade obviamente que as dos efeitos principais, mas, nem por isso, menos válidas.

Vamos considerar um experimento com 3 fatores com 2 níveis cada um, realizado em 2 blocos e repetido 3 vezes. É possível confundir a interação ABC na primeira série, AC na segunda e BC na terceira série de réplicas, conforme o plano.

1a. série	Bloco I	: (1) – AB – AC – BC	(confusão de ABC)
	Bloco II	: A – B – C – ABC	
2a. série	Bloco I	: (1) – B – AC – ABC	(confusão de AC)
	Bloco II	: A – AB – C – BC	
3a. série	Bloco I	: (1) – A – BC – ABC	(confusão de BC)
	Bloco II	: B – AB – C – AC	

EXERCÍCIO 58

No estudo da separação dos elementos X e Y (em concentração igual numa solução), por meio de resina de troca iônica, foi usado um eluente em 2 níveis de concentração (A_1 e A_2), 2 níveis de pH (B_1 e B_2) e dois níveis de temperatura (C_1 e C_2). Os efluentes foram recolhidos em frascos de 2 ml e foi examinada a concentração de X e Y em cada fração pela radioatividade dos traçadores usados. Verificou-se que a maior parte de X sai nas primeiras 10 frações eluídas. Estas 10 frações foram agrupadas e determinou-se a relação das concentrações [X] / [Y]. Os resultados estão na Tabela IX.7.

O efeito principal de cada fator independe dos blocos e é calculado como de costume, a partir dos totais por tratamento. Para ter um efeito médio, divide-se a diferença por 4 x 3 = 12 (4 tratamentos e 3 réplicas por tratamento).

Acha-se:

$$A = 1/12\,[(A-1)\,(B+1)\,(C+1)] = \frac{121}{12} = 10{,}08$$

$$B = \frac{44{,}8}{12} = 3{,}73$$

$$C = -\frac{10{,}2}{12} = -0{,}85$$

A interação AB que não foi confundida também deve ser calculada

$$AB = \frac{18{,}8}{12} = -1{,}57$$

Tabela IX.7

Resultados da Separação dos Elementos X e Y por Troca Iônica

Colunas Diferentes	Série I		Série II		Série III	
(Blocos)	1	2	1	2	1	2
Resultados das Relações ([x] / [y]) − 20 em g/1	ABC = 22,1 A = 19,0 B = 15,2 C = 8,7	AB = 20,3 AC = 16,3 BC = 9,4 (1) = 7,0	ABC = 18,4 AC = 20,2 B = 12,8 (1) = 5,6	AB = 18,4 BC = 10,2 A = 17,0 C = 5,4	ABC = 23,6 BC = 10,8 A = 20,2 (1) = 8,4	AB = 21,2 AC = 18,3 B = 14,5 C = 6,0
Total dos Blocos	65,0	53,0	57,0	51,0	63,0	60,0
Total da Série	118		108		123	
Interação Confundida	ABC		AC		BC	
Total por Tratamento	(1) = 21,0 A = 56,2 B = 42,5 AB = 59,9		C = 20,1 AC = 54,8 BC = 30,4 ABC = 64,1			
Total Geral	349					

A interação AC, como foi confundida com os resultados da série II, pode ser avaliada com os resultados das outras duas séries. Deve-se observar, porém, que o divisor não é mais 12 mais sim: $4 \times 2 = 8$. Do mesmo modo se calculam BC e ABC

$$AC = -\frac{16,8}{8} = -2,10$$

$$BC = -\frac{8,6}{8} = -1,07$$

$$ABC = \frac{4,8}{8} = 0,60$$

As somas dos quadrados correspondentes se calculam como de costume, elevando ao quadrado a diferença total e dividindo pelo número de observações. Por exemplo:

$$A = [A]^2/24 = 121^2/24 = 610,04$$

Para poder fazer uma análise da variância, é suficiente calcular a soma dos quadrados totais:

$$22,1^2 + 19,0^2 + \ldots + 14,5^2 + 6,0^2 - 349^2/24 = 5856,62 - 5075,04 = 781,58$$

e a soma dos quadrados devida à diferença entre blocos:

$$(65{,}0^2 + 53{,}0^2 + \ldots + 60{,}0^2)/4 - 5075{,}04 = 38{,}21$$

O quadro da análise da variância está na Tabela IX.8.

Tabela IX.8

Análise da Variância dos Resultados da Tabela IX.7

Fonte de Variação	Nº de Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
Diferença entre:				
Blocos	5	38,21	7,64	60,6
Fator A	1	610,04	610,04	968
Fator B	1	83,63	83,63	132,7
Fator C	1	4,33	4,33	6,9
Interação AB	1	14,73	14,73	23,4
Interação AC	1	17,64	17,64	28
Interação BC	1	4,62	4,62	7,3
Interação ABC	1	1,44	1,44	2,3*
Erro (Resíduo)	11	6,94	0,63	
Total	23	781,58		
	Valores de F Tabelado		$F_{0,01(1,11)} = 9{,}7$ $F_{0,05(1,11)} = 4{,}8$ $F_{0,05(5,11)} = 3{,}2$	

Só não é significativa a interação ABC cujo valor de F está assinalado com um asterisco. O estudo das interações pode ser esclarecido dispondo os resultados conforme a Tabela IX.9.

Tabela IX.9

Estudo das Interações e Fatores

Fatores	Resposta Média	Resposta com:					
		A^-	A^+	B^-	B^+	C^-	C^+
A	10,00	–	–	11,65	8,51	12,18	7,98
B	3,73	5,30	2,16	–	–		
C	– 0,85	1,26	– 2,95	0,22	– 1,92		

Observação: Os sinais (+) e (–) como expoentes de A, B e C indicam respectivamente presença ou ausência do fator.

A resposta média do fator A em ausência do fator B é:

$$A - AB = 10{,}08 - (-1{,}57) = 11{,}65$$

e em presença de B é:

$$A + AB = 10{,}08 + (-1{,}57) = 8{,}51$$

Do mesmo modo se calculam as outras respostas médias.

A variância de uma observação é 0,63. Cada resposta média é calculada pela diferença entre as médias de 2 grupos de 12 observações. A variância de cada média é então 0,63/12 = 0,0525. A variância média da diferença entre 2 médias é 2 x 0,0525 = 0,105 e o desvio padrão é 0,32. O cálculo do desvio padrão de cada resposta diferencial se faz do mesmo modo, por exemplo: para o grupo (Fator A — Fator B), a variância de A e a variância de AB são iguais, isto é, 0,105 e a variância da soma ou diferença é 0,21 e desvio padrão é 0,46. Nos outros grupos aparecem as interações AC e BC que, estando confundidas com as séries II e III respectivamente, são obtidas pelas diferenças entre as médias de 8 observações e a variância é (0,63 x 2)/8 = 0,1575. A variância da resposta diferencial obtida por adição ou diferença com A, B ou C cuja variância é 0,63/6 será:

$$0{,}63\left(\frac{1}{6} + \frac{1}{4}\right) = 0{,}2625$$

e o desvio padrão é 0,51.

Pode-se calcular, com o auxílio dos valores dos desvios padrão, os limites de confiança de cada resultado.

IX.3 — Fracionamento

Quando o número de fatores for maior ou igual a 5, o número de tratamentos torna-se tão grande que, mesmo confundindo algumas fontes de variação, o experimento se torna de realização difícil. Neste caso é conveniente, como primeira orientação, estudar o efeito de alguns fatores, isto é, fazer uma fração da análise completa.

A título de exemplo, vamos ver como se faria no caso de 3 fatores com 2 níveis (8 tratamentos) divididos em 2 blocos em que se confunde a interação ABC.

Bloco I : (1), AB, AC, BC

Bloco II : A, B, C, ABC

Vamos levar em consideração só o Bloco I e calcular os efeitos médios de cada fator com os 4 resultados deste bloco.

Num experimento completo, o efeito médio de A seria:

$$A = 1/4\,[A + AB + AC + ABC - ((1) + B + C + BC)]$$

No delineamento fracionado, o efeito médio de A é:

$$A = 1/2\,[A + ABC - (B + C)]$$

Para a interação BC temos:

Experimento completo : $BC = 1/4\,[A + BC + ABC + (1) - (B+C+AB+AC)]$

Experimento fracionado : $BC = 1/2\,[A + ABC - (B+C)]$
que é o mesmo calculado para A.

Segue-se que, quando se faz só a metade das operações, A e BC são dados pela mesma equação e deve-se considerar então que, na realidade, temos:

$$1/2\,[A + ABC - (B+C)] = A + BC$$

Se a interação BC for pequena ou não existir, esta equação dá uma boa estimativa de A. Caso contrário, é impossível saber qual a contribuição de A e qual a de BC.

Em geral, quando se confunde a interação ABC entre 2 blocos e se executa somente um deles, cada fator principal está associado a uma interação que se obtém multiplicando o fator principal pela interação confundida, lembrando a regra que estabelece que $A^2 = B^2 = C^2 = 1$

A está associado a $A \times ABC = BC$
B está associado a $B \times ABC = AC$
C está associado a $C \times ABC = AB$

Quando se calculam os efeitos médios, usa-se a notação A (BC) para indicar que o valor de A é válido se a interação BC for desprezível. No caso de 3 fatores, o fracionamento não é aconselhável, porque as interações entre 2 fatores não são raras. Quando o número de fatores aumenta, o delineamento fatorial fracionado é de grande importância, porque permite avaliar os efeitos médios principais e os de 2a. ordem.

IX.3.1 – Aplicação ao Caso de 6 Fatores

O experimento consiste de $2^6 = 64$ provas. Pode-se fazer a metade deste número, confundindo o efeito ABCDEF e vamos ter os 32 tratamentos seguintes:

(1) – AB – AC – AD – AE – AF – BC – BD – BE – BF – CD – CE – CF – DE – DF – EF – ABCD – ABCE – ABCF – ABDE – ABDF – ABEF – BCDE – BCDF – BCEF – CDEF – ACDE – ACDF – ADEF – BDEF – ACEF – ABCDEF.

Cada efeito principal está associado a uma interação de 5ª ordem e cada interação de 2 fatores está associada a uma interação de 4ª ordem, como se pode ver na Tabela IX.10. As interações de mais de 3 fatores tem pouca chance de existir. Assim temos uma boa estimativa dos efeitos principais e das interações entre 2 fatores. Por outro lado, o erro experimental constituído pela soma dos quadrados das interações de 3 fatores e de seus associados tem 10 graus de liberdade o que permite uma boa sensibilidade.

Tabela IX.10

Associação de Fatores e Interações num Experimento 2^6 Realizado pela Metade.

Fonte de Variação	Nº de Graus de Liberdade
A (BCDEF)	1
B (ACDEF)	1
C (ABDEF)	1
D (ABCEF)	1
E (ABCDF)	1
F (ABCDE)	1
AB (CDEF)	1
AC (BDEF)	1
AD (BCEF)	1
AE (BCDF)	1
AF (BCDE)	1
BC (ADEF)	1
BD (ACEF)	1
BE (ACDF)	1
BF (ACDE)	1
CD (ABEF)	1
CE (ABDF)	1
CF (ABDE)	1
DE (ABCF)	1
DF (ABCE)	1
EF (ABCD)	1
Erro:	10
ABC (DEF)	
ABD (CEF), etc.	
Total	31

IX.3.2 – Tratamento Confundido num Delineamento Fatorial Fracionado

Mesmo que se efetue só uma parte de um experimento, é possível, por razões técnicas (aparelhos, amostras, etc. diferentes), dividir o conjunto em blocos.

Voltando ao exemplo de 6 fatores, as diferenças entre blocos não podem ser confundidas com a interação ABCDEF, porque já foi usada para estabelecer a metade das provas. As interações de 4 e 5 fatores não podem ser usadas, porque estão associadas respectivamente às interações de 2 fatores e aos fatores principais. Escolhem-se então as interações de 3 fatores.

Se quisermos formar 2 blocos (por exemplo, 2 aparelhos) e escolhermos a interação ABC (DEF) para ser confundida com a diferença entre os blocos, vamos ter no Bloco I os tratamentos com um número par de letras em comum com ABC ou com DEF e no Bloco II os tratamentos restantes, conforme a Tabela IX.11.

Tabela IX.11

Distribuição dos Tratamentos de um Experimento 2^6
Realizado pela Metade, em 2 Blocos

Bloco I	Bloco II
(1) – AB – AC – BC – DE – DF – EF – ABDE – ABDF – ABEF – ACDE – ACDF – ACEF – BCDE – BCDF – BCEF.	AD – AE – AF – BD – BF – CD – BE CE – CF – ABCD – ABCE – ABCF – ADEF – BDEF – CDEF – ABCDEF

Para formar 4 blocos, é preciso confundir 3 interações: 2 de 3 fatores, por exemplo, ABC e ADE e a 3a. será ABC x ADE = BCDE que tem como associado AF, interação de 2 fatores. Na prática, o que se faz é escolher a interação de 2 fatores que tem pouca chance de existir e depois se determinam as outras duas de 3 fatores.

Na Tabela IX.12 está representada, como exemplo, uma divisão em 4 blocos em que se confundem as interações ABC(DEF); ADE(BCF) e AF(BCDE).

Tabela IX.12

Distribuição dos Tratamentos de um Experimento 2^6,
Realizado pela Metade, em 4 Blocos.

Bloco I	Bloco II	Bloco III	Bloco IV
(1) BC – DE – BCDE – ABEF – ACEF – ACDF – ABDF	AB – AC – EF – DF – ABDE – ACDE – BCEF – BCDF	AD – AE – CF – BF – ABCD – ABCE – BDEF – CDEF	AF – BD – CD – BE – CE – ADEF – ABCF – ABCDEF –

EXERCÍCIO 59

Deseja-se saber quais os fatores que influem na extração de um elemento X, considerando-se os seguintes:

A – 2 valores de pH ($A_1 < A_2$)

B – 3 concentrações de agente extrator ($B_1 < B_2$)

C – 2 concentrações de X ($C_1 < C_2$)

D – 2 concentrações salinas ($D_1 < D_2$)

E – 2 tipos de diluentes (E_1, diluente rotineiro; E_2, a ser testado)

F – 2 valores de temperatura ($T_1 < T_2$)

Em todos os casos, o metal é retro-extraído da fase orgânica e a solução final de X levada a um volume de 50 ml de onde são tiradas alíquotas de 1 ml que são levadas para um contador monocanal (contagem total). Comparando-se a atividade colocada com a atividade extraída, obtém-se o rendimento procurado.

As operações são feitas em 4 extratores e 4 agitadores mecânicos que se supõem iguais. Admitindo-se uma possível diferença entre eles, principalmente quanto à agitação, adota-se o método de confusão, usando ABC, ADE e AF (interação de 2 fatores menos provável) para confundir os efeitos. Como seriam 64 provas no total, faz-se a metade delas distribuída em 4 blocos. Os resultados expressos em rendimento de extração (%) dos quais foi deduzido 90% estão na Tabela IX.13. Calcula-se o efeito total de cada fator pelo método já visto.

Tabela IX.13

Resultados dos Tratamentos de Extrações do Metal X.
Os Resultados são Dados em Porcentagem Menos 90%.

Bloco I	Bloco II	Bloco III	Bloco IV
(1) = - 1,2	AB = 2,8	AD = 0,3	BD = 2,4
BC = 2,0	AC = - 0,7	ABCD = 0,9	CD = 0,9
DE = 0,4	ABDE = 0,3	AE = 1,8	BE = 5,1
BCDE = 5,2	ACDE = 3,6	ABCE = 3,4	CE = 2,9
ABEF = 0,5	EF = 2,4	BDEF = 3,6	ADEF = 2,1
ACEF = 3,1	BCEF = 4,6	CDEF = 4,0	ABCDEF = 4,5
ACDF = 1,3	BCDF = 3,8	CF = 0,0	ABCF = 5,0
ABDF = 2,1	DF = - 1,4	BF = 1,8	AF = - 1,8
13,4	15,4	15,8	21,1

Por exemplo, para A temos que somar todos os resultados dos tratamentos onde aparece o fator A e desta soma deduzir a soma dos resultados dos tratamentos onde não aparece A. Temos assim o efeito total de A que vamos representar por A_t

$$A_t = 29{,}2 - 36{,}5 = -7{,}3$$

O efeito médio é obtido dividindo este resultado por 16

$$A = (-7{,}3) / 16 = -0{,}45625$$

No caso das interações de 2 fatores, por exemplo AB, somam-se todos os resultados dos tratamentos em que aparece AB com aqueles em que não aparece nem A e nem B; desta soma se deduz a soma dos outros resultados (tratamentos com A ou B).

$$AB_t = 27,5 - 38,2 = -10,7$$

e a interação média é:

$$AB = (-10,7) / 16 = -0,66875$$

O resultado dos cálculos está na Tabela IX.14.

Tabela IX.14

Resultados Obtidos para o Efeito dos Fatores e Interações

Fonte de Variação	Efeito Total (E_T)	Efeito Médio $(E_T)/16$	Graus de Liberdade	Soma dos Quadrados $(E_T)^2/32$	Quadrado Médio
A (BCDEF)	− 7,3	−0,45625	1	1,6653125	
B (ACDEF)	30,3	1,89375	1	28,6903125	
C (ABDEF)	23,3	1,45625	1	16,9653125	
D (ABCEF)	2,3	0,14375	1	0,1653125	
E (ABCDF)	29,3	1,83125	1	26,8278125	
F (ABCDE)	5,5	0,34375	1	0,9453125	
AB (CDEF)	−10,7	−0,66875	1	3,5778125	
AC (BDEF)	3,4	0,21250	1	0,3612500	
AD (BCEF)	− 0,3	−0,01875	1	0,0028125	
AE (BCDF)	−10,5	−0,65625	1	3,4453125	
BC (ADEF)	− 1,7	−0,10625	1	0,0903125	
BD (ACEF)	− 7,1	−0,44375	1	1,5753125	
BE (ACDF)	−16,5	−1,03125	1	8,5078125	
BF (ACDE)	2,1	0,13125	1	0,1378125	
CD (ABEF)	9,7	0,60625	1	2,9403125	
CE (ABDF)	11,7	0,73125	1	4,2778125	
CF (ABDE)	10,7	0,66875	1	3,5778125	
DE (ABCF)	2,5	−0,15625	1	0,1953125	
DF (ABCE)	6,5	0,40625	1	1,3203125	
EF (ABCD)	− 1,3	−0,08125	1	0,0528125	
Dif. entre Blocos			3	4,0559375	1,3520
Erro			8	10,8215625	1,3527
Total			31	120,1996875	

Valores tabelados de F : $F_{0,05(1,8)} = 5,3$

$F_{0,01(1,8)} = 11,3$

Para poder julgar o significado destes resultados, calcula-se a soma dos quadrados correspondente a cada um (efeito total ao quadrado dividido por 32).

Calcula-se também:

a) A soma dos quadrados devida às diferenças entre os blocos

$$(13,4^2 + 15,4^2 + 15,8^2 + 21,1^2)/8 - (13,4 + 15,4 + 15,8 + 21,1)^2/32 =$$

$$= 138,94625 - 134,8903125$$

$$= 4,0559375$$

b) A soma dos quadrados totais

$$(-1,2)^2 + (2,0)^2 + \ldots + (5,0)^2 + (-1,8)^2 - 134,8903125 = 120,1996875$$

A soma dos quadrados atribuída ao erro experimental, calculada por diferença, é aquela das 8 interações de 3 fatores. Na realidade, um experimento 2^6 tem 20 interações de 3 fatores, mas como foi feita a metade dos experimentos, elas se confundem duas a duas. Além disso foram usadas duas interações e suas associadas ABC(DEF) e ADE(BCF) para dividir os resultados em 4 blocos. Estas aparecem nas somas dos quadrados atribuídas às diferenças entre os blocos. O terceiro componente desta soma de quadrados é BCDE(AF) e por esse motivo a interação AF não aparece na Tabela.

A variância do erro experimental é 1,3527 e um teste F vai mostrar que os fatores B, C e E têm uma influência significativa, mesmo num nível de significância de 0,01, enquanto que a interação BE é significativa num nível de 0,05.

Conclui-se que o fator C(concentração de X) tem influência independente de qualquer fator. O aumento da concentração favorece a extração. A análise dos efeitos B e E, em virtude da existência da interação BE, exige um desmembramento dos resultados conforme a Tabela IX.15. O valor de B_1E_1 se obtém da Tabela IX.14, somando os 8 resultados onde não aparece nem B e nem E e dividindo o total por 8. O valor de B_2E_1 se obtém dividindo por 8 a soma dos oito resultados em que aparece B e não aparece E, etc.

Tabela IX.15

Análise dos Efeitos: Concentração do Agente Extrator (B) e pH(E)

Concentração do Agente Extrator (B)	pH (E) E_1	E_2
B_1	- 2,60	2,54
B_2	2,60	3,40

Vê-se pelos resultados da Tabela IX.15 que, aumentando a concentração do agente extrator, aumenta o rendimento nos dois valores de pH, mas, em pH mais alto (E_2), este aumento é muito mais acentuado.

Conclusão:

O tratamento que dá o melhor resultado é:

A — pH mais elevado: A_2
B — concentração do agente extrator maior: B_2
C — concentração de X mais elevada: C_2
D — concentração salina indiferente nos níveis testados
E — indiferente: os 2 tipos de diluentes têm o mesmo efeito
F — indiferente: efeito de T_1 = efeito de T_2

Conclusão: Deve ser escolhido o tratamento ABC

IX.4 — Delineamento Fatorial com um Número Reduzido de Fatores

Nos experimentos que envolvem um grande número de fatores de difícil controle, é necessário obter uma estimativa do erro experimental baseada nos próprios resultados. Neste caso, o número de interações de ordem elevada deve ser suficiente para que o número de graus de liberdade do erro possa ser grande para não prejudicar a sensibilidade do teste. Alguns experimentos (agrícolas e em biologia em particular) são de duração prolongada o que exige que sejam feitas simultaneamente todas as provas previstas.

Neste tipo de experimentos, os tratamentos podem ser realizados em pequenas séries sucessivas, utilizando uma estimativa do erro experimental obtido em provas anteriores. Muitas vezes também é possível prever se dois fatores são independentes ou se são sujeitos a uma interação. Pode-se então aplicar o delineamento fracionado limitado a 3 ou 4 fatores.

Vamos considerar o estudo da influência dos fatores A e B sobre o rendimento de uma reação e supor que não exista nenhuma razão técnica que justifique uma interação entre eles. Adota-se a notação de acordo com o seguinte esquema:

Fator ou Interação A	B	AB	Tratamento	Rendimento (x)
−	−	+	(1)	x_1
+	−	−	A	x_2
−	+	−	B	x_3
+	+	+	AB	x_4

Cada efeito total se obtém atribuindo ao rendimento o sinal correspondente.

Por exemplo:

$$A_t = -x_1 + x_2 - x_3 + x_4$$

Se os fatores A e B forem independentes, o valor de AB deve ser da ordem do erro experimental. Se substituirmos AB por um fator C que não tenha interação com A e B, os sinais correspondentes a AB poderão servir para avaliar o efeito total C.

O esquema, neste caso, toma o seguinte aspecto.

Fatores A	B	C	Tratamento	Rendimento (x)
−	−	+	C	x_1
+	−	−	A	x_2
−	+	−	B	x_3
+	+	+	ABC	x_4

Os 4 tratamentos feitos correspondem aos de um bloco de um experimento 2^3 dividido em 2 blocos, em que se confunde o tratamento ABC. Quando se faz um experimento fracionado, igualando um fator suplementar a uma interação de um experimento fatorial completo (C=AB), o grupo de interações confundidas se obtém multiplicando esta igualdade pelo novo fator

$$(C = AB) \times C = (I = ABC)$$

O outro bloco seria obtido igualando C a − AB

Neste experimento dividido por 2, cada fator está associado a uma interação de 2 fatores: A(BC), B(AC), C(AB).

Por isso, foi preciso admitir a priori a independência dos 3 fatores para poder obter uma estimativa satisfatória dos respectivos efeitos.

Muitas vezes esta hipótese feita a priori é difícil de justificar; por este motivo, costuma-se fracionar em 2 blocos um experimento com 4 fatores de 2 níveis (16/2 = 8 tratamentos) em que cada fator está associado a uma interação de 3 fatores.

EXERCÍCIO 60

Estudo da separação radioquímica dos elementos X e Y por precipitação, onde podem influir os seguintes fatores:

A — Massa de carregador: A_1 = 5 mg; A_2 = 20 mg

B — Massa de precipitante: B_1 = 50 mg; B_2 = 100 mg

C — Massa de agente quelante: C_1 = 40 mg; C_2 = 80 mg

D — Temperatura: D_1 = 25°C; D_2 = 80°C

O experimento fracionado se faz substituindo a interação ABC pelo fator D. Os 8 tratamentos que devem ser feitos correspondem a um bloco dos 2 blocos obtidos confundindo D x (D = ABC) = (I = ABCD).

No trabalho experimental, foram usados traçadores radioativos de X e Y.

A relação [x] / [y] na solução original era 0,75 e, após a separação, foram obtidos os resultados da Tabela IX.15. Para facilitar os cálculos, de cada resultado foi subtraído o valor 110.

Tabela IX.15

Resultados de um Estudo de Separação Radioquímica dos Elementos X e Y

Fatores				Tratamento	Resultado [X]/[Y]−110	Interação		
A	B	C	D = ABC			AB(CD)	AC(BD)	AD(BC)
−	−	−	−	(1)	3,28	+	+	+
+	−	−	+	AD	3,28	−	−	+
−	+	−	+	BD	1,76	−	+	−
+	+	−	−	AB	1,04	+	−	−
−	−	+	+	CD	3,20	+	−	−
+	−	+	−	AC	3,28	−	+	−
−	+	+	−	BC	− 0,08	−	−	+
+	+	+	+	ABCD	0,48	+	+	+
− 0,08	− 9,84	− 2,48	1,2			0,24	1,36	− 2,32

Cada efeito principal está associado a uma interação de 3 fatores e as interações de 2 fatores são agrupadas duas a duas. Pode-se ver, por exemplo, que os sinas da coluna AB(CD) são obtidos tanto pelos produtos das colunas A e B como C e D. Os efeitos totais (última linha) são obtidos aplicando os sinais das colunas aos resultados e fazendo a soma algébrica. O mesmo resultado pode ser obtido pelo método de Yates. Os tratamentos são dispostos numa ordem bem determinada em relação aos fatores A, B, C. Nesta ordem, os sinais são os seguintes para cada fator:

A : − + − + − + − + ...

B : − − + + − − + + ...

C : − − − − + + + + ...

D : − (x8) + (x8) ....

Os tratamentos são colocados nesta ordem, em função dos fatores A, B e C, e aqueles onde intervem D = ABC são indicados por D entre parentesis. O cálculo é feito do mesmo modo como já indicado anteriormente e se obtém os resultados da Tabela IX.16.

Tabela IX.16

Cálculo de Efeito dos Fatores e Interações

Tratamento	Resultado	Coluna (1)	Coluna (2)	Coluna (3)	Efeito e Associado	Efeito Médio (col 3)/4	Soma de Quadrado (col 3)²/4
(1)	3,28	6,56	9,36	16,24	–	–	–
A(D)	3,28	2,80	6,88	– 0,08	A,BCD	– 0,02	0,0008
B(D)	1,76	6,48	– 0,72	– 9,84	B,ACD	– 2,46	12,1032
AB	1,04	0,40	0,64	– 0,24	AB,CD	– 0,06	0,0072
C(D)	3,20	0,00	– 3,76	– 2,48	C,ABD	– 0,62	0,7688
AC	3,28	– 0,72	– 6,08	1,36	AC,BD	0,34	0,2312
BC	– 0,08	0,08	– 0,72	– 2,32	BC,AD	– 0,58	0,6728
ABC(D)	0,48	0,56	0,48	1,20	ABC,D	0,30	0,1800

Não é possível fazer uma estimativa do erro experimental com este grupo de tratamentos. Este tipo de delineamento fatorial exige que o erro experimental seja conhecido por provas anteriores. Suponhamos que, no caso do exercício, a variância seja 0,01 baseada num número suficiente de resultados para tomar $n \cong \infty$. Nestas condições vamos ter $F_{0,05(1,\infty)} = 3,8$. Para que um quadrado médio seja significativo neste nível é preciso que tenha um valor mínimo de 3,8 x 0,01 = 0,038.

Vemos que o fator B é altamente significativo e que as interações AC(BD) e AD(BC) provavelmente existem. Como, á priori, não se pode dizer quais as interações que realmente existem, porque estão associadas duas a duas, é preciso examinar cada uma destas interações quanto ao seu comportamento na separação dos elementos X e Y.

Interação AC

A mg de Carregador	C – mg Agente Quelante 40	80
5	112,52	111,56
20	112,16	111,88

Interação BD

B mg de Precipitante	D – Temperatura 25°C	80°C
50	113,28	113,24
100	110,48	111,12

Para E e C no nível mais baixo temos:

$$\frac{BD + (1)}{2} = \frac{1,76 + 3,28}{2} = \frac{5,04}{2} = 2,52$$

Para C no nível mais baixo e A no nível mais alto temos:

$$\frac{A(D) + AB}{2} = \frac{3{,}28 + 1{,}04}{2} = \frac{4{,}32}{2} = 2{,}16$$

Para C no nível mais alto e A no mais baixo vamos ter:

$$\frac{C(D) + BC}{2} = \frac{3{,}20 - 0{,}08}{2} = \frac{3{,}12}{2} = 1{,}56$$

Para A e C no nível mais alto temos:

$$\frac{AC + BC\,(D)}{2} = \frac{3{,}28 + 0{,}48}{2} = 1{,}88$$

Do mesmo modo se calculam os valores para a interação BD. As melhores condições, por estas duas análises, seriam:

A — quantidade de carregador = 25 mg
B — massa de precipitante = 50 mg
C — massa de agente quelante = 40 mg
D — temperatura: indiferente

A análise do grupo AD(BC) é feita do mesmo modo

Interação AD

A mg de Carregador	D — Temperatura 25°C	80°C
5	111,6	112,48
10	112,16	111,88

Interação BC

B mg de Precipitante	C — mg Agente Quelante 40	80
50	113,28	113,24
100	111,40	110,20

O exame destas duas Tabelas permite concluir que as condições favoráveis são:

A — quantidade de carregador = 5 mg
B — massa de precipitante = 50 mg
C — massa de agente quelante = indiferente
D — temperatura: 80°C

Usando os 2 grupos de conclusões, opta-se pelas seguintes condições

A — 5 mg de carregador
B — 50 mg de precipitante
C — 40 mg de agente quelante
D — temperatura 80°C

IX.4.1 — Possibilidades de Maior Fracionamento

Se for possível saber que certos fatores são independentes, seja por resultados já obtidos ou por considerações teóricas, pode-se estudar um fator a mais com o mesmo número de tratamentos.

Sejam A, B e C três fatores entre os quais sabemos que B e C são independentes, isto é, a interação BC não existe, mas AB e AC são prováveis. Como a interação ABC é muito pouco provável (3a. ordem) pode-se introduzir um fator D igualando-o a ABC e um fator E igualando-o a BC conforme a Tabela IX.17.

Tabela IX.17

Substituição de Interações Pouco Prováveis ou Inexistentes por Novos Fatores

Fatores					
A	B	C	D(ABC)	E = (BC)	Tratamentos
−	−	−	−	+	E
+	−	−	+	+	ADE
−	+	−	+	−	BD
+	+	−	−	−	AB
−	−	+	+	−	CD
+	−	+	−	−	AC
−	+	+	−	+	BCE
+	+	+	+	+	ABCDE

Os sinais de D e de E são respectivamente os produtos de A, B e C e B e C na mesma linha.

Um experimento de 5 fatores com 2 níveis, se completo, exigiria $2^5 = 32$ tratamentos. Neste plano, temos a quarta parte dos tratamentos que corresponde a um dos 4 blocos que se obteriam confundindo 3 interações. Estas interações são dadas por:

D x (D = ABC) = (I = ABCD)
E x (E = BC) = (I = BCE)

A terceira interação é obtida pelo produto das 2 primeiras e será ADE. Como só se faz a quarta parte do experimento completo, o efeito de cada fator estará associado a 3 interações que se obtém multiplicando o fator pelas interações confundidas. Por exemplo, o fator A está associado a:

$$A \times ABCD = BCD$$
$$A \times BCE = ABCE$$
$$A \times ADE = DE$$

No caso das interações temos, por exemplo, para AB:

$$AB \times ABCD = CD$$
$$AB \times BCE = ACE$$
$$AB \times ADE = BDE$$

Desprezando as interações de 3a ordem como sendo pouco prováveis, obtemos as seguintes associações:

$$A = DE$$
$$B = CE$$
$$C = BE$$
$$D = AE$$
$$E = BC = AD$$
$$AB = CD$$
$$AC = BD$$

Vemos assim que os 5 fatores e as duas interações principais AB e AC não estão confundidas entre si e se pode obter uma estimativa correta de todos com 8 tratamentos em vez de 32.

Se os resultados obtidos com estes 8 tratamentos não forem satisfatórios, é sempre possível fazer mais um quarto do tratamento completo, usando o bom senso na escolha.

IX.5 — Delineamento Fatorial com Fracionamento Progressivo

Quando se realiza uma pesquisa que requer uma série de experimentos do mesmo tipo, sabe-se, geralmente, depois de realizado um número razoável de provas, qual é o erro experimental. Se nesta pesquisa for interassante estudar a influência de vários fatores (mínimo 5), é conveniente racionalizar as provas de maneira a ser executado o menor número necessário.

No decorrer da realização das provas, depois da análise de cada conjunto, será possível eliminar fatores que não afetam o problema, mudar os níveis dos fatores de interesse ou também introduzir outros fatores.

EXERCÍCIO 61

Estudo da retenção de um elemento M em resina catiônica ("batch"), utilizando um radioisótopo de M. Em todas as provas foi usada a mesma massa de resina e o mesmo volume de solução de M. Pretende-se estudar a influência dos seguintes fatores:

A — tipo de resina......... I — II

B — tipo de agitação....... lenta — rápida

C — tempo de contato..... 5 min 20 min

D — temperatura........... 25°C 60°C

E — pH da solução........ 4 5

Depois de cada tratamento, a resina é filtrada e lavada com 5 ml de água. A resina é colocada num tubo de contagem e o filtrado é levado a um volume conveniente e tirada uma alíquota para contagem. Os resultados correspondem à relação

$$\frac{\text{Ativ de M/g de resina}}{\text{Ativ de M/ml de filtrado}}$$

Um delineamento fatorial completo exigiria 32 tratamentos, mas decide-se fazer somente 8, escolhendo os fatores A, B e C e introduzindo D = ABC e E = AB. Os 8 tratamentos escolhidos correspondem a um dos 4 blocos em que se confundiram as interações (1), ABCD, AB e CDE. Os tratamentos e respectivos resultados estão na Tabela IX.18.

Tabela IX.18

Tratamentos e Resultados do Estudo da Retenção de M por Resinas

Fatores A	B	C	D = ABC	E = AB	Tratamento	Resultado − 100	Efeito Principal e Interação Dupla	Efeito Médio	Soma de Quadrados
−	−	−	−	+	E	1,50	A+BE	−4,935	48,70845
+	−	−	+	−	AD	1,98	B+AE	1,725	5,95125
−	+	−	+	−	BD	9,84	C+DE	6,795	92,34405
+	+	−	−	+	ABE	− 0,54	D+CE	−2,625	13,78125
−	−	+	+	+	CDE	6,84	E+AB+CD	−8,055	129,76605
+	−	+	−	−	AC	12,60			
−	+	+	−	−	BC	18,06			
+	+	+	+	+	ABCDE	2,46			
−19,74	6,90	27,18	−10,50	−32,22					

Na Tabela IX.18 aparecem o efeito principal e seus associados de 2 fatores. Admitimos que as interações de mais de 2 fatores são desprezíveis. Sabe-se que, nesse tipo de experiências, por 60 resultados anteriores, a variância do erro experimental é 21.0 valor de $F_{0,05(1,60)} = 4$. Então o efeito do fator é real se a sua variância for superior a $4 \times 21 = 84$.

Vimos que C+DE e E+AB+CD têm um efeito significativo sobre o rendimento. A variância mais elevada é a de E+AB+CD e seria interessante saber qual a parte correspondente a E e qual a correspondente às duas interações.

Para isso, faz-se mais um quarto do delineamento total. Pode-se escolher um dos 3 blocos restantes que podem ser determinados da seguinte maneira:

b) D = − ABC E = − AB Interações confundidas : (I) − ABCD, − ABE, CDE

c) D = − ABC E = AB Interações confundidas : (I) − ABCD, ABE, − CDE

d) D = ABC E = − AB Interações confundidas (I) ABCD, − ABE, − CDE

É conveniente escolher o último, porque, nos tratamentos já feitos (Tabela IX.18), E está associado a ABCDE, AB e CD. No bloco d), confundindo ABCD, -ABE e -CDE teremos E associado a ABCDE, -AB e -CD. Fazendo:

Bloco a) $E + AB + CD + ABCDE = X_1$

Bloco b) $E - AB - CD + ABCDE = X_4$

pela soma das 2 equações temos:

$$E + ABCDE = 1/2\,(x_1 + x_4)$$

e pela diferença:

$$AB + CD = 1/2\,(x_1 - x_2)$$

Temos assim o efeito principal E associado a uma interação de 5 fatores (ABCDE) que é desprezível e um grupo de interações de 2 fatores AB e CD. Por ser um delineamento fatorial 2^5, isto é, 32 tratamentos dos quais se realizam 16, cada fator ou interação só tem um associado.

Neste fracionamento, os tratamentos têm um número par de letras em comum com a interação ABCD. Os 8 resultados da Tabela IX.18 e os outros 8 resultados obtidos nas provas posteriores estão na Tabela IX.19. Os resultados estão dispostos na ordem clássica em relação a A, B, C e E.

O efeito total de A é calculado somando os resultados dos 8 tratamentos onde aparece A e desta soma se deduzem os resultados dos 8 tratamentos onde não aparece A. O efeito médio é o efeito total dividido por 8. A soma dos quadrados é o efeito médio ao quadrado multiplicado por 4. Os resultados do efeito médio na Tabela IX.19 foram aproximados, por isso a soma dos quadrados, que foi calculada com os valores reais, não é exatamente o resultado do cálculo indicado.

Se tomarmos $S^2 = 21$ como variância do erro experimental, nenhuma das interações de 2 fatores é significativa. Só 3 fatores são significativos.

A − resina I dá melhor rendimento

C − um tempo de contato de 20 minutos aumenta o rendimento

E − pH 4 dá rendimento maior que pH 5

Com 16 provas, em vez de 32, é possível resolver o problema. Depois de ter realizado as primeiras 8 provas, foi obtido o resultado:

$$E + AB + CD = -8{,}055$$

para o efeito médio. Poderia-se deduzir que a resina II, isto é, E no nível mais alto é desfavorável, levando em conta que as interações AB e CD não devem ser tão pronunciadas. Poderia-se, porém, correr o risco dessas duas interações existirem e serem responsáveis pelo valor deste efeito médio.

Tabela IX.19

1ª Complementação dos Tratamentos e Resultados da Retenção de M por Resinas

Fatores A	B	C	E	D = ABC	Tratamento	Resultado (− 100)	Efeito Principal e Associado	Efeito Médio	Soma de Quadrados
−	−	−	−	−	(1)	9,12	—	—	—
+	−	−	−	+	A(D)	1,98	A+BCD	− 7,74	242,42
−	+	−	−	+	B(D)	9,48	B+ACD	1,02	4,16
+	+	−	−	−	AB	2,22	AB+CD	− 2,58	26,32
−	−	+	−	+	C(D)	18,06	C+ABD	4,50	80,46
+	−	+	−	−	AC	12,60	AC+BD	− 2,22	20,25
−	+	+	−	−	BC	18,06	BC+AD	—	—
+	+	+	−	+	ABC(D)	4,32	D+ABC	0,36	0,52
−	−	−	+	−	E	1,50	E+ABCDE	− 5,46	120,56
+	−	−	+	+	AE(D)	3,42	AE+BCDE	0,72	1,98
−	+	−	+	+	BE(D)	8,76	BE+ACDE	2,82	32,49
+	+	−	+	−	ABD	− 0,54	ABE+CDE	− 0,36	0,56
−	−	+	+	+	CE(D)	6,84	CE+ABDE	− 3,00	35,64
+	−	+	+	−	ACE	− 3,36	ACE+BDE	− 1,14	5,20
−	+	+	+	−	BCE	13,20	BCE+ADE	2,28	20,52
+	+	+	+	+	ABCE(D)	2,46	DE+ABCE	2,28	21,34

Se tivéssemos concluído pelas primeiras 8 provas que a resina I é mais favorável, o fator E estaria excluído, trabalharíamos só com a resina I e estudaríamos só os fatores A, B, C e D cujo estudo completo exigiria $2^4 = 16$ tratamentos. Destes 16 poderia-se fazer a metade, igualando D a ABC. A interação confundida seria (1) = ABCD e o bloco principal seria formado pelos tratamentos:

(1) ; AB ; AC ; AD ; BC ; BD ; CD e ABCD

Comparando este bloco com a Tabela (IX.18), vemos que 4 tratamentos já foram feitos e resta fazer (1); AB; CD e ABCD. Os resultados destes 4 tratamentos e mais aqueles que já tinham sido obtidos estão na Tabela IX.20.

Ve-se que as interações de 2 fatores são desprezíveis. Os fatores A e C são significativos e, como o efeito médio para A é negativo, o nível mais baixo (Resina I) desse fator é o mais favorável, enquanto que para C é mais favorável um tempo de contato de 20 minutos (nível mais alto). Os fatores B e D são

Tabela IX.20

2ª Complementação dos Tratamentos e Resultados
da Retenção de M por Resinas

Tratamentos	Resultados (– 100)	Método de Yates			Efeitos Associados	Efeito Médio	Soma de Quadrados
		Col 1	Col 2	Col 3			
(1)	9,12	11,10	23,16	–	–	–	–
A(D)	1,98	12,06	53,04	– 33,96	A+BCD	– 8,49	144,16
B(D)	9,84	30,66	– 14,76	– 7,32	B+ACD	– 1,83	6,70
AB	2,22	22,38	– 19,20	– 8,76	AB+CD	– 2,14	9,59
C(D)	18,06	– 7,14	0,96	29,88	C+ABD	7,47	111,60
AC	12,60	– 7,62	– 8,28	– 4,44	AC+BD	– 1,11	2,40
BC	18,06	– 5,46	– 0,48	– 9,24	BC+AD	– 2,31	10,67
ABC(D)	4,32	– 13,74	– 8,28	– 7,80	D+ABC	– 1,95	7,60
Interação Confundida : ABCD							

indiferentes. Chega-se assim à mesma conclusão com 12 tratamentos em vez de 16. Se AB + CD fossem significativos, a eliminação de E não teria sido feita com base suficiente, seriam necessários os outros 4 tratamentos: ADE, BDE, ACE, BCE e se chegaria então à Tabela IX.19.

REFERÊNCIAS BIBLIOGRÁFICAS

1. CRUMPLER, T. B. & YOE, J. H. *Chemical computations and errors.* New York, Wiley, 1947.

2. NALIMOV, V. V. *The application of mathematical statistics to chemical analysis.* Reading, Mass., Addison Wesley, 1963.

3. PHILIPPE, J. *Les méthodes statistiques en pharmacie et en chimie: applications à la recherche, à la production et au controle.* Paris, Masson, 1967.

4. TORRES, O. F. F. *Curso de estatística.* São Paulo, Escola Politécnica, 1967.