

Capítulo 1

Introdução ao Planejamento de Experimentos

Antes de começar uma discussão aprofundada sobre os fundamentos teóricos e numéricos que dão suporte à formulação das diferentes técnicas de planejamento experimental, é conveniente primeiro contextualizar a discussão. Por isso, algumas perguntas filosóficas precisam ser inicialmente respondidas, visando tanto ao desenvolvimento embasado das técnicas matemáticas e numéricas que auxiliam o pesquisador na sua tarefa de investigação experimental, quanto à preparação do próprio pesquisador para lidar com a realidade que o cerca. Por exemplo:

- 1- Em que consiste exatamente planejar experimentos?
- 2- Quais são os objetivos perseguidos durante a tarefa de planejamento experimental?
- 3- Que expectativas o pesquisador deve ter a respeito dos resultados obtidos com o auxílio das rotinas de planejamento experimental?

Embora essas questões possam parecer pouco significativas para aqueles que preferem privilegiar os aspectos mecânicos mais diretamente relacionados à aplicação das técnicas de planejamento experimental, o fato é que a análise dessas questões se reveste de enorme importância prática. A experiência mostra que é enorme a frustração de muitos profissionais, e mesmo de amplos grupos de pesquisadores em certas áreas de estudo, com os resultados obtidos com o auxílio de técnicas de planejamento experimental. Aqueles que militam na área já se acostumaram a ouvir frases depreciativas como "*Nada substitui o bom senso de pesquisador!*" e "*Olha, eu respeito a sua opinião, mas no meu caso isso não me ajudou em nada!*". No entanto, a experiência também mostra que na maior parte dos casos a frustração é consequência da má compreensão a respeito do uso das técnicas de planejamento experimental e dos resultados que de fato podem ser obtidos com o auxílio dessas ferramentas. Por isso, para que se possa desenvolver uma postura correta a respeito de como lidar com problemas de planejamento experimental, é conveniente aprofundar um pouco essas questões de cunho eminentemente filosófico.

1.1. Afinal, o Que é Planejar Experimentos?

Toda vez que se fala em experimentos, supõe-se que algumas observações serão feitas pelo pesquisador. Portanto, supõe-se que algumas variáveis, que constituem o objeto da observação do pesquisador, serão medidas. O conjunto de procedimentos que resulta na coleta de um conjunto de observações experimentais é o que se chama de **experimento**. Em um contexto amplo, portanto, planejar experimentos pode ser definido como:

- a) Definir as variáveis a serem observadas e que, portanto, constituem o universo da investigação;
- b) Definir como e com que equipamentos as variáveis serão medidas e/ou avaliadas;
- c) Definir o conjunto de medidas a serem feitas e as condições da experimentação;
- d) Definir a seqüência experimental a ser seguida, dado que nem todas as observações podem ser feitas ao mesmo tempo.

Colocada nessa forma, vê-se que a rotina de planejamento experimental envolve a execução de um grande número de tarefas, das quais o pesquisador não pode se eximir. É comum que pesquisadores procurem auxílio das técnicas de planejamento experimental com a expectativa de que seja possível desenvolver uma receita miraculosa e universal, que permita resolver qualquer problema de natureza experimental, independentemente do problema investigado. **Essa expectativa é falsa e certamente leva o pesquisador à frustração.** O problema investigado é parte integrante do processo experimental e é impossível desconsiderar as particularidades do problema durante o estabelecimento das rotinas de planejamento. É preciso que o experimentador compreenda que o problema experimental é parte do problema de planejamento, influenciando e sendo influenciado pelas técnicas de projeto experimental utilizadas. Por exemplo, deve ser observado que as seguintes questões secundárias são oriundas da primeira questão proposta no item a:

- a.1) Escolher quantas e quais variáveis serão investigadas;
- a.2) Escolher a região experimental em que as variáveis serão investigadas.

As duas questões apresentadas anteriormente só podem ser definidas pelo próprio investigador. O pesquisador é quem mais conhece o problema analisado e cabe somente a ele definir o universo experimental investigado. Nesse caso, o universo experimental é constituído pelo conjunto de variáveis investigadas e pelo conjunto de condições experimentais que podem e/ou devem ser avaliadas. O pesquisador que entrega a tarefa de escolha das variáveis e das regiões experimentais investigadas a um consultor externo comete um equívoco de enormes proporções (a não ser, claro, que o consultor conheça profundamente a atividade experimental que será executada). Assim, em um problema em que se investiga os efeitos que certas variáveis exercem sobre a velocidade de uma certa reação química, cabe ao pesquisador definir se a temperatura deve ser incluída no conjunto de variáveis experimentais e, caso a resposta seja afirmativa, qual é a faixa admissível para a investigação (por exemplo, 30 °C a 80 °C). É equivocada a noção de que as variáveis e faixas experimentais possam resultar do planejamento experimental. Muito pelo contrário, variáveis e região de experimentação constituem pressupostos básicos do planejamento experimental. O especialista em técnicas de planejamento pode apenas ajudar o pesquisador a avaliar as conseqüências resultantes das escolhas e limitações que o trabalho experimental impõe.

Exemplo 1.1- Para que o leitor se convença das questões apresentadas anteriormente, pode-se raciocinar pelo absurdo.

Suponha que a velocidade com que certos microorganismos decompõem uma certa fonte de matéria orgânica esteja sendo investigada. Suponha ainda que a temperatura seja uma das variáveis investigadas, dado que a temperatura do processo pode ser manipulada de maneira eficiente. Essa escolha parece apropriada, pois sabe-se que a temperatura influencia de fato a atividade de organismos vivos. Escolhe-se então a faixa de 1000 °C a 1200 °C para se

conduzir a investigação. Essa escolha não parece apropriada, pois sabe-se que organismos vivos não resistem a essas condições extremas de temperatura.

A despeito disso, sabe-se que reações de craqueamento (que promovem a quebra de moléculas maiores em moléculas menores) podem ocorrer a altas temperaturas. Por isso, vários processos químicos importantes, como a produção de olefinas a partir de parafinas, são conduzidos a temperaturas altas e próximas da faixa de 1000 °C a 1200 °C. Logo, a escolha da temperatura como variável experimental em condições de experimentação extremamente altas parece apropriada para o estudo do processo de craqueamento.

Finalmente, considere que o desempenho de um certo implante deva ser investigado em condições similares às do organismo humano. Nesse caso, a variação da temperatura ao longo do estudo experimental talvez não faça qualquer sentido, pois se sabe que a temperatura do corpo humano flutua muito pouco, estando quase sempre a cerca de 37 °C. Sendo assim, talvez seja mais apropriado manter a temperatura constante e igual a 37 °C em todos os ensaios.

O leitor atento certamente perceberá que não existe valor intrínseco algum nem escolha da temperatura como variável de investigação nem na faixa de temperaturas de 1000 °C a 1200 °C como faixa apropriada para condução de estudos experimentais. O pesquisador atento deverá avaliar cuidadosamente em cada caso que ações devem ser tomadas. Quanto mais se conhece o problema experimental, melhor se faz o plano experimental.

Deve ser observado no Exemplo 1.1 que a definição do que é ou não é apropriado está fortemente embasada no conhecimento prévio do sistema. Em todos os casos foi usada a expressão "*sabe-se que...*" para julgar o que é apropriado e o que não é. Se o pesquisador não soubesse (ou não desconfiasse) que a temperatura era importante, o efeito dessa importante variável poderia não ser investigado e, pior ainda, essa variável poderia não ser nem controlada e nem mesmo monitorada ao longo do trabalho experimental. Dessa forma, flutuações de temperatura durante os experimentos resultariam em aparente ruído ou flutuação aleatória das medidas experimentais, confundindo o analista. Da mesma forma, a escolha inapropriada da região de experimentação resultaria em medida de atividade microbiana nula, pois todos os microorganismos estariam mortos naquelas condições extremas de temperatura. O experimentador seria levado a concluir que a temperatura não altera a atividade microbiana, que de qualquer maneira é sempre igual a zero. A conclusão, que de fato não estaria errada, estaria limitada àquela região experimental investigada. De forma similar, o estudo das reações de craqueamento a baixas temperaturas resultaria em baixíssimas velocidades de reação e levaria o experimentador a concluir que a temperatura não afeta o processo, o que de fato não é verdade na região experimental apropriada. Finalmente, o estudo sobre o desempenho da prótese em ampla faixa de temperaturas constituiria apenas perda de tempo para a maior parte dos propósitos práticos, dado que a peça será mantida todo o tempo na mesma temperatura.

O Exemplo 1.1 ilustra como o pesquisador e o planejamento experimental interagem fortemente. **Quanto mais se sabe sobre um problema, melhor se planejam os experimentos.** Essa é uma frase reconfortante para aqueles que já acumularam quantidade considerável de informações sobre um problema, mas preocupante para aqueles que pouco conhecem o sistema investigado. O pesquisador é instado a fazer hipóteses e considerações sobre o sistema experimental investigado o tempo todo, o que inclui a definição do conjunto de variáveis investigadas e a região de experimentação. Se as variáveis e/ou região de experimentação são

escolhidas de forma inapropriada, colhem-se resultados pouco conclusivos ou, como dizem os pessimistas, "ruins".

Mas seriam esses resultados "ruins" mesmo? Uma visão mais serena do problema indica, na realidade, que os resultados usualmente chamados de "ruins" são aqueles que mostram que o pesquisador entende pouco do problema experimental investigado. São, portanto, resultados de fato bons, que mostram ao experimentador que a sua concepção do problema tem que mudar. Na maior parte das vezes, os resultados são chamados de "ruins" apenas porque negam os pressupostos feitos pelo pesquisador. É o conservadorismo do investigador, que muitas vezes já imagina saber (!) o resultado que vai ser obtido, que taxa de ruim os resultados que não confirmam as suas premissas básicas. Quando o experimentador propõe que as reações de craqueamento sejam realizadas a baixas temperaturas e no final conclui que a atividade química é muito baixa nessas condições, ele de fato está aprendendo que compreendia mal o comportamento do sistema. Embora o resultado seja frustrante diante da expectativa inicial, será que nada foi aprendido? Será que o pesquisador não estaria apenas sendo instado a considerar outras variáveis ou outras condições experimentais não consideradas inicialmente? Afinal, por que o resultado pretendido tem que ser sempre obtido após a realização dos primeiros experimentos?

Como se livrar então dessas premissas, que tantas vezes se mostram equivocadas? Uma possível resposta no que diz respeito às variáveis seria considerar todas as possíveis variáveis que podem influenciar o problema experimental. Esse problema, no entanto, não tem solução que não passe pela formulação de hipóteses sobre o comportamento do sistema. Para que uma variável possa ser descartada do estudo experimental (por exemplo, deve-se incluir ou não a luminosidade do dia como fator importante para o estudo sobre atividade microbiana apresentado no Exemplo 1.1?) é necessário que alguma lógica racional seja utilizada (por exemplo, descarte a luminosidade do dia porque os experimentos são realizados sempre em ambientes pouco iluminados) e alguma hipótese seja formulada. Caso contrário, todas as possíveis variáveis do universo teriam que ser investigadas, o que seria de fato impossível. O mesmo diz respeito à região de experimentação. Se qualquer condição experimental interessa e é possível, a experimentação tem que ser conduzida no limite do intervalo $(-\infty, +\infty)$, o que é impossível. Conclui-se, portanto, que **todo planejamento experimental está baseado em pressupostos sobre o comportamento do sistema experimental investigado**. A negação desses pressupostos não indica que os resultados são ruins, mas que o investigador conhece ainda pouco o sistema investigado. A negação de certos pressupostos quase sempre obriga o investigador a refazer os seus planos, seja revendo o conjunto de variáveis investigadas ou a região de experimentação. Portanto, **toda rotina de planejamento experimental deve ser encarada como um processo iterativo e interativo**, ao longo do qual os pressupostos vão sendo revistos e aperfeiçoados em função dos resultados obtidos.

A discussão anterior mostra que o pesquisador que espera sempre obter respostas conclusivas logo após a primeira rodada de observações experimentais pode ficar bastante frustrado com as respostas obtidas. Em particular porque as respostas podem indicar que o pesquisador conhece muito pouco o sistema experimental investigado. Além disso, o pesquisador que entrega o seu problema experimental a um estatístico pode também ficar muito frustrado com os resultados obtidos, já que usualmente ninguém pode agregar mais conhecimento a um planejamento experimental do que o próprio interessado na investigação do problema.

Considere agora as seguintes questões secundárias oriundas da segunda questão proposta no item b:

- b.1) Escolher técnicas e equipamentos adequados para a avaliação das variáveis do problema;
- b.2) Caracterizar os erros experimentais.

De maneira geral, **se a variável não pode ser medida, não há como avaliar seu efeito sobre o resultado dos experimentos**. Portanto, uma das tarefas mais importantes do pesquisador é definir que técnicas e equipamentos devem ser utilizados para acompanhar o trabalho experimental. É necessário inclusive reconhecer as limitações do sistema de medição utilizado e de que maneira a medição influencia o experimento. Por exemplo, em muitos trabalhos de investigação o acompanhamento experimental é realizado por amostragem, o que significa que o pesquisador perturba o sistema de tempos em tempos para "ver" o que está acontecendo. Logo, o pesquisador interage com o sistema investigado. No caso da atividade microbiana do Exemplo 1.1, por exemplo, a amostragem pode sempre resultar em contaminação do meio com outros microorganismos, que assim podem falsear completamente as observações realizadas. O resultado do experimento, portanto, não pode ser analisado de maneira independente das técnicas e equipamentos usados para medição. Isso obriga o experimentador a estar ciente das incertezas inerentes ao seu sistema de observação, do grau de interação existente entre o processo de medição e o resultado experimental e das fontes de perturbação do experimento. Essa tarefa também não pode ser delegada a terceiros e deve ser parte das preocupações fundamentais do pesquisador.

Exemplo 1.2- Para que o leitor se convença da importância das questões apresentadas anteriormente, deve-se raciocinar novamente por absurdo.

Suponha que no problema de craqueamento apresentado no Exemplo 1.1 seja proposta a utilização de um termômetro de mercúrio para monitorar a temperatura do sistema, como normalmente efetuado para medir a temperatura do corpo humano. Como o mercúrio é volátil e entra em ebulição a 357 °C na pressão atmosférica, parece bastante claro que não é possível medir temperaturas extremas de operação com instrumento similar ao usado para monitorar a febre dos bebês, apesar da variável experimental nos dois casos ser a temperatura. Portanto, soluções distintas precisam ser propostas para a medição de uma mesma grandeza física em diferentes problemas experimentais. O uso de um termômetro de mercúrio no problema de craqueamento resultará provavelmente em incapacidade de medir e monitorar a temperatura do experimento de forma apropriada, levando a falhas de procedimento e a insucesso da investigação. Em condições extremas, quase sempre são utilizados instrumentos que monitoram a temperatura do processo com base na quantidade de radiação infravermelha emitida pelos corpos quentes. Esses equipamentos são usualmente pouco úteis a baixas temperaturas, uma vez que os corpos emitem pouca radiação infravermelha nessas condições.

O leitor atento certamente perceberá que não existe valor intrínseco algum na escolha de um sistema de medição de qualquer grandeza física. É necessário que os sistemas de medição estejam devidamente ajustados ao problema experimental considerado, o que inclui as grandezas físicas medidas e as condições de experimentação. Uma vez mais, quanto mais se conhece o problema experimental, melhor se faz o plano experimental.

Por todas as razões discutidas anteriormente e no Capítulo 1 do Volume I desta série de publicações, os experimentos contêm incertezas que precisam ser avaliadas e quantificadas. Para que seja possível avaliar de forma bastante clara a importância das incertezas na formulação do planejamento experimental, são analisados dois casos fictícios abaixo. Suponha primeiramente que a atividade microbiana definida no Exemplo 1.1 mude de 2 quilogramas de material orgânico processado por quilograma de célula viva por hora (2 kg/kg.h) para 2.5 kg/kg.h, quando a temperatura do sistema muda de 35 °C para 40 °C. Suponha agora que duas técnicas distintas de caracterização da atividade microbiana têm precisão de ± 0.1 e ± 1.0 kg/kg.h respectivamente, com 95% de confiança. Se a primeira técnica experimental for usada, o efeito da temperatura será certamente detectado, pois a mudança provocada pela temperatura é muito maior que o possível erro de avaliação cometido. No segundo caso, o efeito da temperatura será completamente obscurecido pelas flutuações devidas aos erros experimentais. No segundo caso, portanto, a influência da temperatura sobre a atividade teria que ser muito maior para que fosse efetivamente detectada. Se o efeito da temperatura sobre a atividade microbiana fosse linear (o que provavelmente não é verdade), seria necessário que a temperatura máxima analisada fosse de 50 °C para que o efeito de temperatura passasse a ser maior que a magnitude do erro experimental. Vê-se, portanto, que a magnitude do erro experimental pode influenciar bastante a definição da região de experimentação, além de ser fundamental para a interpretação apropriada dos resultados obtidos. Essa questão é ilustrada esquematicamente na Figura 1.1 e ajuda a compreender o porquê de normalmente ser sugerido que a região de experimentação seja explorada de forma tão ampla quanto possível.

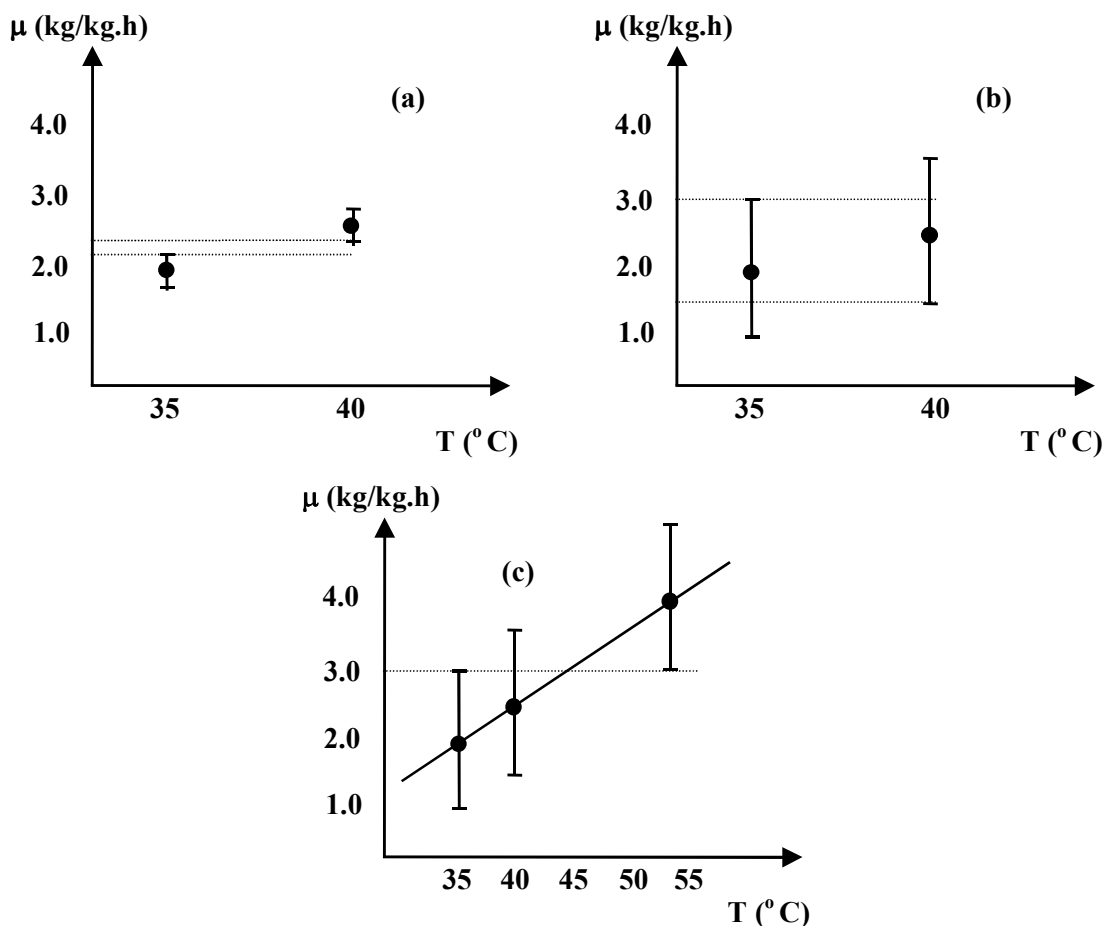


Figura 1.1- Influência dos erros de medição sobre o projeto de experimentos e interpretação dos dados. (a) Erros pequenos mostram a influência de T (temperatura) sobre μ (atividade microbiana). (b) Erros grandes não permitem ver de forma conclusiva a influência de T sobre μ . (c) Se o efeito é linear, ampliação da região de experimentação pode compensar a existência de erros grandes.

Para que seja analisada com maior profundidade a questão do erro experimental, suponha que os erros experimentais mudam de ponto para ponto e não são uniformes. Esse tipo de comportamento é bastante comum, como avaliado no Capítulo 4 do Volume I desta série de publicações. Os erros experimentais podem variar por causa de flutuações não lineares das respostas dos equipamentos usados para medição ou por causa do melhor controle das condições experimentais em certas regiões de experimentação. Por exemplo, suponha que a quantidade de calor dissipado numa manta de aquecimento usada para controlar a temperatura de uma unidade experimental é determinada por uma resistência elétrica. Nesse caso, segundo a Lei de Ohm,

$$\dot{Q} = \frac{U^2}{R} \tag{1.1}$$

onde \dot{Q} é a taxa de dissipação de calor, U é a diferença de potencial e R é a resistência elétrica oferecida. Normalmente a diferença de potencial U é mantida constante e igual ao valor fornecido pela distribuidora de energia elétrica (110 V / 220 V) e disponível na tomada. A resistência R pode ser controlada variando-se a posição de um elemento de contato na forma

$$R = \frac{x}{L} R_0 \tag{1.2}$$

onde x é a posição do elemento de contato, L é o comprimento total do elemento que oferece a resistência elétrica e R_0 é a máxima resistência que pode ser oferecida pelo elemento, quando o contato é estabelecido no final da peça. Portanto, a quantidade de calor dissipada pode ser escrita como

$$\dot{Q} = \frac{L U^2}{R_0} \frac{1}{x} = \frac{C}{x} \tag{1.3}$$

onde C é uma constante característica do sistema de aquecimento. Se as incertezas Δx em relação à posição real x_0 do elemento de contato são pequenas, então

$$\Delta \dot{Q} \approx d\dot{Q} = -\frac{C}{x_0^2} dx \approx -\frac{C}{x_0^2} \Delta x \tag{1.4}$$

ou

$$\sigma_{\dot{Q}}^2 = E\{\Delta \dot{Q}^2\} \approx E\left\{\left[-\frac{C}{x_0^2} \Delta x\right]^2\right\} = \frac{C^2}{x_0^4} E\{\Delta x^2\} = \frac{C^2}{x_0^4} \sigma_x^2 \tag{1.5}$$

de forma que, mesmo que o erro da medida x seja pequeno e constante, como numa régua, os erros associados às flutuações das quantidades dissipadas de calor variam fortemente com a

posição do elemento de contato ao longo da resistência, aumentando à medida que x diminui. Como visto no Capítulo 4 do Volume I desta série de publicações o comportamento dos erros experimentais influencia enormemente a correta formulação do problema de estimação de parâmetros, necessário para uma interpretação quantitativa adequada dos dados experimentais. Portanto, é necessário também avaliar como os erros variam de ponto para ponto na região de experimentação. Em última instância, como discutido na Seção 4.6 do Volume I desta série de publicações, é a distribuição dos erros experimentais que define a forma quantitativa apropriada de analisar os dados obtidos. Assim, **quanto mais informação se tem sobre os erros experimentais, mais adequado é o tratamento quantitativo dos dados experimentais e mais consistentes são as conclusões obtidas**. Caso o experimentador não saiba *a priori* como os erros se comportam na região de experimentação, essa avaliação deve ser parte integrante do planejamento experimental; ou seja, o plano experimental deve ser capaz de informar como as respostas do problema experimental, **incluindo as variâncias de medição**, dependem das condições de experimentação. A Figura 1.2 procura ilustrar a importância do erro experimental para a interpretação quantitativa dos dados em situações distintas de medição experimental. Se o experimentador não sabe nada sobre o comportamento dos erros experimentais, ele é obrigado a formular hipóteses adicionais a respeito do comportamento dos erros para escolher a técnica apropriada para o tratamento quantitativo dos dados. Uma vez mais, essas hipóteses podem ser verdadeiras ou não.

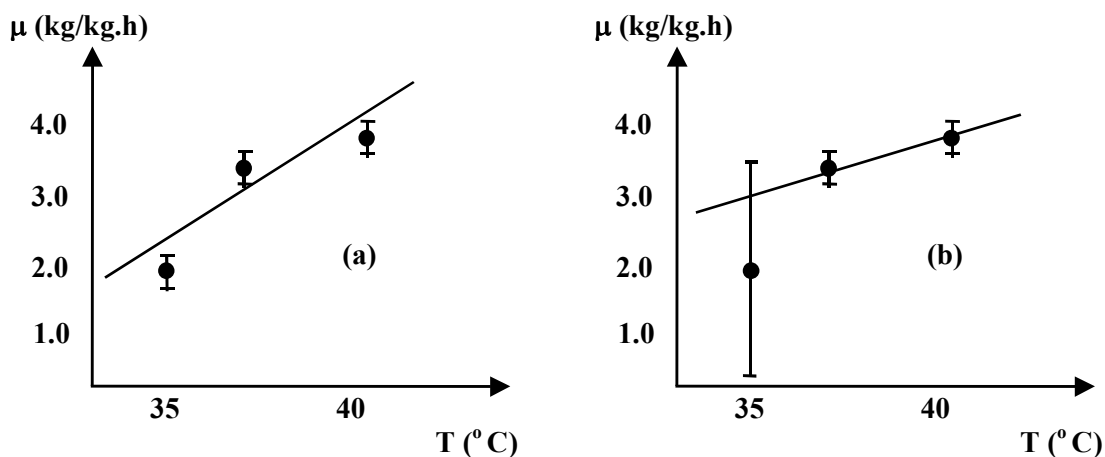


Figura 1.2- Influência dos erros de medição sobre a interpretação dos dados. (a) e (b) ilustram a noção de melhor reta em situações distintas dos erros.

Considere agora as seguintes questões secundárias oriundas da terceira questão proposta no item c:

- c.1) Definir a quantidade de experimentos que devem ser executados;
- c.2) Definir as condições experimentais para a observação.

De maneira geral, a questão filosófica fundamental que permeia a definição do número de observações que devem ser colhidas é o conflito existente entre a precisão da informação obtida (que aumenta à medida que o número de experimentos aumenta, como discutido amplamente no Volume I desta série de publicações) e o custo do trabalho experimental (medidos em termos de tempo de experimentação e custo financeiro dos experimentos). É impossível resolver esse problema sem a participação efetiva do investigador. Cabe ao

investigador definir os objetivos a serem perseguidos pelo planejamento experimental e especificar os alvos a serem atingidos. Por exemplo, o experimentador pode definir que o objetivo fundamental do plano experimental é determinar com 95% de certeza (alvo relacionado à precisão e qualidade da observação) se o coeficiente linear da reta (alvo relacionado ao tipo de modelo) que relaciona a temperatura à atividade microbiana (alvo relacionado às variáveis do estudo) é significativo ou não (objetivo fundamental do estudo), sendo que o erro da análise não pode ser superior a um certo valor σ^2 (alvo relacionado à incerteza absoluta máxima admitida). Nem sempre é fácil especificar objetivos e alvos, como será discutido na próxima seção, porque nem sempre se conhece com clareza suficiente o sistema experimental para que seja possível formular os alvos de maneira precisa. Por exemplo, e se a temperatura e a atividade microbiana não se relacionarem linearmente? E se o número de experimentos necessários para atingir σ^2 for muito grande? Vemo-nos obrigados de novo a formular hipóteses e estabelecer pressupostos que eventualmente terão que ser revisados, se não forem confirmados pelas observações experimentais. De qualquer forma, mesmo que a definição dos alvos seja trivial e possível, o plano experimental só será efetivo se puder ser enquadrado no orçamento e no cronograma do projeto. Portanto, **não se começa um plano experimental sem um orçamento disponível nem sem um cronograma admissível**. Cabe ao experimentador apresentar o orçamento e o cronograma do projeto, que devem ser confrontados com as metas perseguidas. Caso as metas se mostrem inexecutáveis no orçamento e cronograma propostos, as metas, o orçamento, o cronograma ou todos têm que ser modificados. Para cada orçamento e cronograma, há planos e conjuntos de metas possíveis característicos. Experimentos baratos e rápidos podem ser executados em grandes quantidades, permitindo definição mais justa e precisa de metas a serem atingidas. Experimentos caros e demorados têm que ser executados em menor número possível e, assim, conviver com incertezas maiores e metas de qualidade mais flexíveis. Cabe ao experimentador definir essas questões filosóficas.

Exemplo 1.3- Para que o leitor se convença da importância das questões apresentadas anteriormente, deve-se raciocinar uma vez mais por absurdo.

Suponha que no problema relacionado à medida da atividade microbiana apresentado no Exemplo 1.1 cada experimento demore cerca de uma semana para ser realizado. Ao contrário do que alguns leitores podem imaginar, experimentos que envolvem a transformação de compostos orgânicos por microrganismos podem ser ainda mais longos. Nunca deve ser esquecido que a realização de experimentos requer a preparação de equipamentos antes da experimentação e a análise posterior dos resultados amostrados durante e após as corridas experimentais. Portanto, o tempo necessário para a condução dos experimentos pode ser bastante grande. Para enfatizar ainda mais esse ponto, o leitor deve imaginar os tempos de experimentação necessários para estudar a produtividade de certas plantas na área de agricultura (imagine o tempo necessário para que uma mangueira produza mangas). No contexto apresentado, se os resultados devem ser apresentados em no máximo dois meses (cronograma), o experimentador não poderá conduzir mais do que oito corridas experimentais. Vê-se, portanto, que o plano experimental deve estar subordinado ao cronograma do projeto.

Suponha ainda que cada experimento custe o equivalente a R\$ 20.000,00, por conta dos insumos necessários e da mão-de-obra contratada para executar os serviços. Nesse caso, a realização de oito corridas experimentais resulta num custo total de R\$ 160.000,00 para o projeto. Suponha ainda que o orçamento disponível para o projeto seja equivalente a R\$ 500.000,00. Nesse caso, o experimentador pode ser tentado a manter duas linhas de

experimentação em paralelo, realizando experimentos distintos em cada uma delas e tornando possível a condução de até 16 experimentos, para um custo total de R\$ 320.000,00, compatível com o orçamento e o cronograma propostos. Se os custos associados à aquisição e instalação da nova unidade experimental for equivalente a R\$ 250.000,00, chega-se à conclusão de que é necessário investimento total de R\$ 570.000,00 no projeto, incompatível com o orçamento disponível. Vê-se, portanto, que o plano experimental deve estar subordinado ao orçamento do projeto.

O leitor atento certamente perceberá que não é possível determinar a extensão do plano experimental sem que o pesquisador defina com bastante precisão quais são o cronograma e o orçamento disponíveis para o projeto. Como pode ser compreendido facilmente pelo leitor, o investigador não pode e não deve delegar a terceiros a definição do cronograma e do orçamento do projeto a ser executado. Começar uma atividade de experimentação sem cronograma e sem orçamento é o mesmo que iniciar a construção de uma casa sem a certeza de que há tempo e dinheiro suficientes para que se execute a obra. Portanto, fica claro uma vez mais que, quanto mais se conhece o problema experimental, melhor se faz o plano experimental.

Com relação à definição das condições de experimentação, essa etapa é o coração do processo de planejamento experimental e será amplamente discutida nas seções posteriores. Como veremos, essa escolha está intimamente relacionada com as respostas apresentadas às questões formuladas anteriormente e aos alvos e hipóteses formulados pelo experimentador.

Considere agora as seguintes questões secundárias oriundas da quarta questão proposta no item d:

- d.1) Definir a ordenação dos experimentos;
- d.2) Definir se os planos experimentais devem ser realizados de forma seqüencial ou em batelada.

De maneira geral, como será discutido adiante, os experimentos devem ser realizados em ordem aleatória, como forma de reduzir o impacto de variáveis não monitoradas e não controladas (fatores externos) sobre os resultados da experimentação. Contudo, isso nem sempre é possível ou conveniente. Suponha, por exemplo, que experimentos são realizados em um reator industrial de 200 m³, em que a influência da geometria da pá do agitador sobre o desempenho do reator está sendo analisada. Obviamente não é possível parar o reator e trocar a pá do agitador após finalizado cada experimento, pois isso resultaria num custo absurdo de montagem e desmontagem de equipamentos, que tem que ser feito com o auxílio de mão-de-obra especializada e máquinas de grande porte. Se os experimentos forem realizados dessa forma, é provável que o cronograma e o orçamento do projeto não sejam compatíveis com o projeto. Nesses casos, definir a ordem de execução dos experimentos pode ser tão importante quanto definir as condições gerais de experimentação.

De forma similar, a realização de experimentos em seqüência, como discutido nos Capítulos 4, 5, 6 e 7, pode ser muito vantajosa, já que as novas observações experimentais podem contribuir muito com a melhor compreensão do sistema estudado e, portanto, com a formulação do próximo plano experimental. No entanto, quando a quantidade de informações existentes é pequena ou quando os experimentos são muito longos, é mais adequado formular planos experimentais em batelada, em que vários experimentos são planejados, executados e

analisados em conjunto. Um exemplo típico é a formulação de planos experimentais para a área agrícola e/ou médica. Suponha que o efeito que a formulação de um certo adubo exerce sobre o desenvolvimento de mangueiras deve ser estudado. Dado o tempo que a mangueira leva para crescer, desenvolver e dar frutos, medido na escala de anos, não parece razoável planejar um ou dois experimentos por vez. Se assim for feito, o experimentador vai dedicar toda a sua vida à resolução de uns poucos problemas práticos. Vê-se uma vez mais que o plano experimental depende da natureza do experimento que se realiza e das respostas que são procuradas.

Toda essa discussão mostra que a rotina de planejamento experimental envolve múltiplas tarefas e não dispensa a participação do experimentador. Muito pelo contrário, é o experimentador que deve escolher variáveis, definir condições de experimentação, determinar alvos técnicos, apresentar orçamentos e cronogramas, preocupar-se com as técnicas de medição e os erros experimentais, formular hipóteses, analisar estratégias de experimentação e trazer o máximo possível de conhecimento prévio para a etapa de planejamento experimental. Somente sobre essas bases é que podem ser desenvolvidas as rotinas de planejamento experimental. Se essas tarefas forem delegadas a terceiros, a probabilidade de que os resultados finais obtidos frustrem os pesquisadores envolvidos é usualmente muito grande.

1.2. Afinal, Para Quê Planejar Experimentos?

Essa pergunta trivial pode ser surpreendentemente profunda, pois poucos são os experimentadores que de fato exercitam essa questão no dia-a-dia. Experimentos são frequentemente conduzidos de forma mecânica e automática, sem que sejam feitas muitas indagações sobre os objetivos reais que estão sendo perseguidos. Por isso, não são poucos os experimentadores que não parecem de fato cientes dos objetivos matemáticos e estatísticos perseguidos por seu trabalho experimental. Alguns podem apressadamente ser tentados a dizer até mesmo que seus trabalhos experimentais não guardam qualquer relação com objetivos matemáticos ou estatísticos de qualquer natureza. Essa visão, como mostrado a seguir, está completamente equivocada. **Cada trabalho experimental pode sempre ser vinculado a objetivos matemáticos bastante precisos, cuja compreensão constitui a base das rotinas de planejamento experimental.** O problema é que o número de objetivos distintos pode ser muito grande, tornando bastante complexa a tarefa de classificação exaustiva dos problemas fundamentais relacionados às rotinas de planejamento experimental. Por isso, é proposta a seguir uma classificação não exaustiva dos problemas típicos perseguidos pelo experimentador durante a atividade de investigação.

1.2.1. Verificar se há vínculos entre variáveis de um conjunto

Esse talvez seja o problema mais fundamental de todos, já que usualmente constitui a primeira etapa da investigação. Nesse estágio, o investigador conhece muito pouco o sistema experimental e nem mesmo sabe que variáveis de um conjunto (x) são responsáveis por variações observadas em um segundo conjunto de variáveis (y). Este é o problema básico: identificar as variáveis do conjunto x que provocam flutuações das variáveis do conjunto y . As ferramentas matemáticas básicas utilizadas para esse fim foram apresentadas ao longo do Volume I desta série de publicações: os fatores de correlação e as rotinas de estimação de parâmetros para modelos lineares nos parâmetros. Em ambos os casos, o uso dessas ferramentas torna possível estabelecer de forma precisa se existem vínculos entre as variáveis e

se esses vínculos são significativos ou não. Esses problemas serão tratados ao longo do Capítulo 2 e constituem a classe de **planejamentos experimentais clássicos**.

Uma pergunta típica formulada nessa fase da experimentação é: "*Será que a temperatura influencia a atividade microbiana?*" A resposta obtida pelo experimentador pode ser sim ou não, com grau de confiança estabelecido pela análise. Uma outra pergunta típica é: "*Quem influencia mais a atividade microbiana, a temperatura ou a quantidade de óleo de oliva adicionado ao meio orgânico?*" A resposta depende certamente do estabelecimento de comparações, cujos resultados dependem, dentre outros fatores, dos erros experimentais. Portanto, é feito amplo uso das ferramentas estatísticas de análise de hipóteses (Os erros experimentais são similares? As médias obtidas são similares?) nessa fase do projeto, como analisado nos Capítulos 1, 2 e 3 do Volume I desta série de publicações. O uso de modelos empíricos para análise quantitativa dos dados obtidos também é muito útil nessa fase do projeto.

Quando a região experimental tem forma complexa ou quando os erros experimentais variam ao longo da região de experimentação, nem sempre é possível gerar soluções fechadas para os problemas de planejamento experimental. Nesses casos, é necessário lançar mão de procedimentos numéricos mais avançados para gerar o projeto experimental perseguido. Esses procedimentos são descritos no Capítulo 4. Tradicionalmente, esses planos experimentais são chamados de **planos experimentais ótimos**, por requererem a formulação explícita de um problema de otimização, ao qual a solução encontrada está vinculada.

1.2.2. Desenvolver modelos matemáticos para o sistema experimental

Essa pode ser considerada uma fase mais avançada da atividade experimental. Pressupõe-se que o pesquisador conhece com razoável grau de certeza o conjunto de variáveis que estão correlacionadas em um certo problema. O pesquisador se interessa, portanto, pela forma funcional mais adequada para descrever essa relação. Do ponto de vista matemático, o pesquisador procura a forma funcional $y = f(x)$ (modelo explícito) ou $g(y,x) = 0$ (modelo implícito) que melhor descreve o conjunto de observações experimentais obtido. Nessa fase, o pesquisador costuma ter várias formas funcionais distintas que podem ser utilizadas para descrever os dados experimentais. Esses candidatos a modelos do processo, chamados de modelos rivais, podem ser obtidos através de argumentos teóricos ou através de exercício empírico. O problema fundamental, chamado de **planejamento experimental para discriminação de modelos**, consiste em escolher as condições experimentais que permitam a identificação mais rápida e eficiente do modelo do processo. Esses procedimentos de planejamento experimental são discutidos no Capítulo 5.

Uma pergunta típica formulada nessa fase da experimentação é: "*Será que a temperatura e a atividade microbiana estão relacionadas linearmente ou exponencialmente?*" A resposta obtida pelo experimentador pode eventualmente favorecer um dos modelos, mas pode também resultar na aprovação ou reprovação de ambos os modelos, com grau de confiança estabelecido pela análise.

1.2.3. Estimar parâmetros precisos em modelos matemáticos

Essa pode ser considerada uma fase ainda mais avançada da atividade experimental. Pressupõe-se que o pesquisador conhece com razoável grau de certeza não só o conjunto de

variáveis, mas também o modelo do processo, o qual pode ter sido obtido após o procedimento de discriminação de modelos brevemente descrito no item anterior. Nesse caso, o foco do problema experimental muda para a estrutura interna do modelo matemático, onde estão os parâmetros. Portanto, o objetivo matemático fundamental é a definição precisa do conjunto de parâmetros α no modelo explícito $y = f(x; \alpha)$ ou no modelo implícito $g(x, y; \alpha) = 0$. O problema experimental, chamado de **planejamento experimental para estimação de parâmetros precisos**, consiste em definir o conjunto de condições experimentais que permitam a estimação mais precisa possível de α . Este problema está intimamente ligado a seguinte questão: "*Como minimizar os erros de predição do modelo em uma certa região experimental?*", já que a qualidade dos parâmetros estimados influencia diretamente a qualidade das predições do modelo. Esses procedimentos de planejamento experimental são discutidos no Capítulo 6.

Uma pergunta típica formulada nessa fase da experimentação é: "*Dado que a temperatura e a atividade microbiana estão relacionadas na forma $\mu = K_0 \exp(-\Delta E/T)$, quem são os valores de K_0 e ΔE com a melhor precisão possível?*" A resposta principal obtida pelo experimentador consiste no conjunto de parâmetros e na respectiva matriz de covariâncias dos parâmetros.

1.2.4. Otimizar medidas de desempenho de sistemas experimentais

Esse problema de planejamento experimental pode ser considerado como uma fase avançada da experimentação, quando um modelo é usado como base para definição das condições ótimas de operação, ou como uma linha independente de investigação experimental, quando são valorizados mais fortemente os resultados experimentais brutos e o uso de modelos empíricos simplificados do processo. Esse conjunto de técnicas dá origem aos problemas de **análise de superfície** ou de **superfície de resposta**, apresentados no Capítulo 4. Nesse caso, procedimentos heurísticos e/ou baseados em modelos empíricos simples são formulados visando a atingir um ponto ótimo de experimentação. A definição do ponto ótimo depende obviamente da compreensão do processo e dos objetivos perseguidos pelo investigador. No primeiro caso, o modelo do processo é usado para gerar as condições de experimentação ótima, visando simultaneamente a reduzir os custos da experimentação e a estabelecer uma conexão entre a experimentação e a economia do processo. No Capítulo 3 é feita uma discussão sobre as implicações econômicas que podem estar associadas ao aumento da variabilidade dos dados experimentais e dos parâmetros dos modelos que descrevem o processo.

Uma pergunta típica formulada nessa fase da experimentação é: "*Em que condição experimental devo operar, para que a atividade microbiana seja máxima?*" A resposta principal perseguida pelo investigador é a localização do ponto de ótimo, o desempenho obtido na condição de ótimo e uma avaliação de quão sensível a medida de desempenho é a perturbações experimentais, respeitados os erros experimentais e o grau de confiança da análise.

1.2.5. Reduzir o esforço experimental

Além dos objetivos genéricos descritos acima, objetivos acessórios que permeiam quaisquer das fases experimentais são: reduzir a quantidade de trabalho experimental executado para que seja obtido um certo conteúdo de informação e reduzir o custo da informação e agilizar o processo de obtenção da informação. Por isso, todos os procedimentos

de planejamento experimental estão intimamente relacionados a rotinas de **otimização**, pois sempre se procura o **máximo de informação com um mínimo de esforço experimental**.

Dados a enorme variedade de problemas experimentais avaliados, as características intrínsecas de cada problema experimental e o enorme número de questões com que o pesquisador deve se preocupar ao formular um plano experimental, não parece surpreendente o fato de que **NÃO EXISTE RECEITA UNIVERSAL PARA O PROBLEMA DE PLANEJAMENTO EXPERIMENTAL**. Essa é certamente a maior fonte de frustração para aqueles que usam de forma desavisada as ferramentas de planejamento de experimentos. Não se iluda: é possível formular princípios, avaliar casos, mas não é possível desenvolver um único algoritmo de planejamento que permita enfrentar qualquer problema de natureza experimental. Se você pensa em abrir um livro e consultar uma tabela milagrosa que resolverá todos os seus problemas, prepare-se para a frustração. É preciso antes de tudo compreender a gênese e as limitações das ferramentas propostas. Cada ferramenta de planejamento está atrelada a um objetivo, aos alvos estabelecidos, aos modelos e variáveis consideradas, ao comportamento dos erros, às características e dificuldades experimentais. Por isso, antes de usar um plano, acostume-se a verificar se objetivos e pressupostos usados para formulá-lo são aceitáveis para o seu problema experimental específico. Se não o for, modifique-o de forma apropriada. Nas seções e capítulos seguintes, é dada ênfase sempre aos pressupostos que permitem a formulação dos planos experimentais propostos e a contextualização do conceito de ótimo.

Finalmente, antes de aprofundarmos a discussão mais específica do planejamento experimental, é interessante ler a série de perguntas formuladas por Himmelblau em um de seus livros clássicos (Himmelblau, D. M., "*Process Analysis by Statistical Methods*", John Wiley & Sons, New York, 1970) e que se encontra reproduzida abaixo. Segundo proposto por Himmelblau, nenhuma atividade experimental deveria ser iniciada antes que essas questões fossem respondidas a contento pelo investigador. Sugere-se aqui que essas questões sejam formalmente respondidas por todos aqueles interessados na formulação de planos experimentais, como forma de auxiliar a definição de objetivos e alvos e de contextualizar o trabalho experimental desenvolvido. Enquanto houver dúvidas a respeito de quaisquer dessas questões, é conveniente que o investigador continue refletindo sobre a natureza e características do trabalho experimental desenvolvido.

Lista de Perguntas Básicas a Serem Respondidas pelo Pesquisador Sobre a Natureza do Trabalho Experimental antes de Iniciar os Experimentos

Definição dos Objetivos

- 1- Por que o trabalho experimental deve ser feito? Que questões devem ser respondidas com o trabalho?
- 2- Quais são as implicações de não serem detectadas correlações que de fato existem ou de serem detectadas correlações que de fato não existem?
- 3- Qual é a região experimental a ser investigada?
- 4- Qual é o tempo disponível?
- 5- Qual é o orçamento disponível?
- 6- Que informações preliminares podem ser consideradas?
- 7- Procura-se um ponto de ótimo ou apenas as correlações entre as variáveis?

Tipo de Modelagem

- 8- Pretende-se fazer uso apenas de modelos empíricos ou também serão usados modelos teóricos?
- 9- A forma funcional analisada está pré-determinada ou deverá ser desenvolvida?
- 10- Quais são as variáveis dependentes, fixadas arbitrariamente pelo investigador para realizar os experimentos, e independentes, obtidas como respostas experimentais?

Programa Experimental

- 11- Que variáveis serão medidas e controladas? Que variáveis serão apenas medidas? Como as variáveis serão medidas? Qual deverá ser a seqüência de medição?
- 12- Quais são os erros de medição associados às medidas de cada variável?
- 13- Que variáveis podem ser consideradas inicialmente como as mais importantes? E quais são as variáveis menos importantes? Os efeitos podem ser de fato avaliados?
- 14- Que problemas externos podem perturbar os resultados? Como controlar e minimizar estes efeitos?
- 15- Que tipo de controle das variáveis é desejável?
- 16- Até que ponto as variáveis independentes podem ser assim consideradas? As variáveis independentes são funções de outras variáveis?
- 17- Os erros são uniformes ou diferentes em cada região de experimentação?

Réplicas

- 18- Como os experimentos serão replicados? Em seqüência, em grupo?
- 19- Quantas réplicas serão realizadas?
- 20- Como os dados serão analisados e interpretados?

1.3. A Definição do Ideal Estatístico e a Realidade Experimental

Do ponto de vista da análise de dados, é relativamente fácil definir a situação ideal a ser perseguida pelo experimentador: fazer experimentos em infinitas condições experimentais distintas, reproduzindo cada um desses experimentos infinitas vezes. Tanto as condições experimentais quanto o ordenamento dos experimentos devem ser definidos de forma aleatória. Portanto, o ideal estatístico pode ser resumido no seguinte trio de diretivas: fazer infinitos experimentos, reproduzir tudo infinitas vezes e aleatorizar todo o procedimento.

Não é difícil justificar o trio de diretivas definido acima. Primeiramente observe que a execução de experimentos em infinitas condições experimentais distintas permite a definição precisa das relações eventualmente existentes entre as muitas variáveis do problema e a obtenção de precisão absoluta nos parâmetros do modelo (ver Capítulo 4 do Volume I desta série de publicações). Depois, observe que a realização de infinitas réplicas em cada uma das condições experimentais permite a definição precisa das distribuições dos erros de experimentação, permitindo simultaneamente a caracterização dos métodos de análise de dados mais apropriados (ver Capítulo 3 e 4 do Volume I desta série de publicações). Particularmente, a identificação apropriada das distribuições de erros permitem identificar *outliers* e erros grosseiros, perceber como os erros dependem das condições de experimentação e definir as rotinas de máxima verossimilhança adequadas para o problema. Finalmente observe que a aleatorização de todo o procedimento de experimentação impede que o pesquisador introduza artificialmente na análise efeitos externos não controlados e aumenta a chance de identificação de padrões experimentais complexos, como mostrado nos Exemplos 1.4, 1.5 e 1.6.

Exemplo 1.4- Suponha que um experimentador está estudando o comportamento de um catalisador desconhecido. O objetivo fundamental da experimentação é definir se e como a atividade do catalisador depende da temperatura de operação. Para isso, são feitos experimentos a temperaturas distintas, medindo-se a quantidade de produto formada numa certa reação química em cada uma das condições. Cada um dos experimentos é repetido cinco vezes. Os experimentos são inicialmente realizados a temperaturas organizadas seqüencialmente, em ordem crescente. Os resultados são apresentados na Figura 1.3.

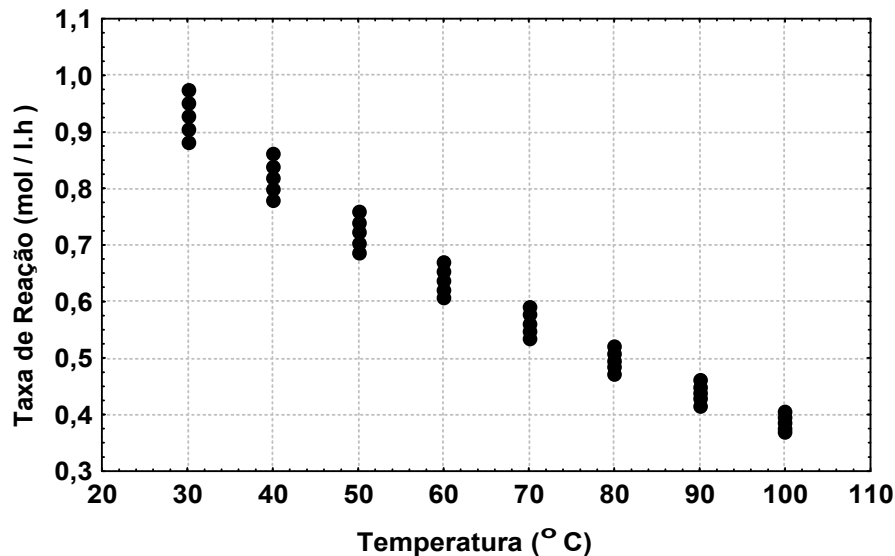


Figura 1.3- Variação da taxa de reação com a temperatura. Ordenamento seqüencial.

O experimentador conclui que a taxa de reação diminui com a temperatura e que o erro de medida também diminui com a temperatura, como pode ser observado na Figura 1.3. Antes de iniciar a etapa de análise quantitativa e modelagem dos dados, um colega sugere que o experimentador refaça os experimentos, aleatorizando o ordenamento das condições experimentais. O experimentador refaz então o estudo, partindo de uma nova batelada de catalisador. Os resultados obtidos são mostrados na Figura 1.4.

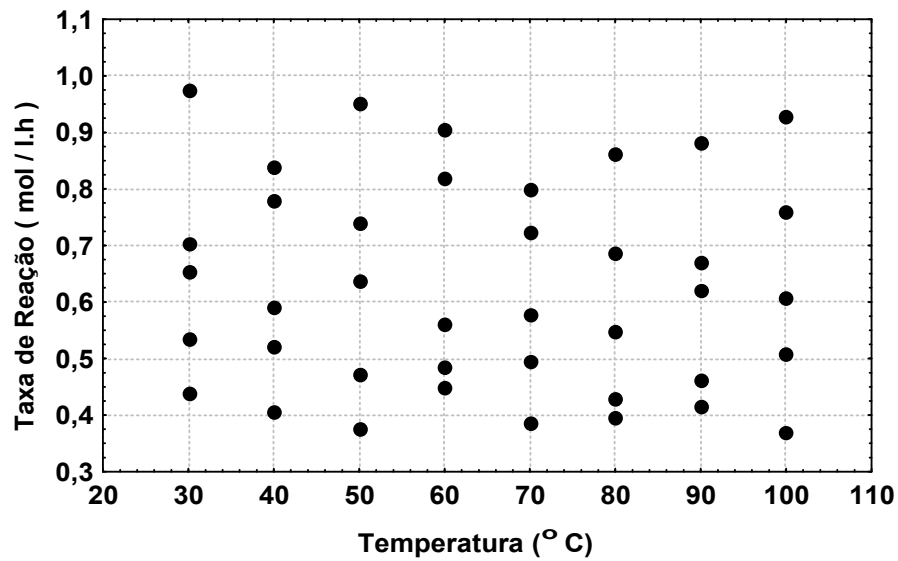


Figura 1.4- Variação da taxa de reação com a temperatura. Ordenamento aleatório.

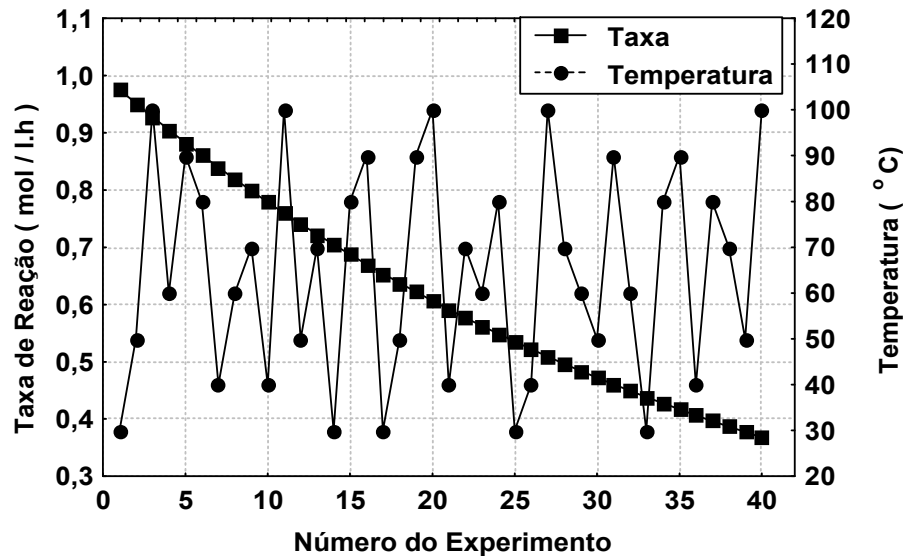


Figura 1.5- Variação da taxa de reação e da temperatura com os experimentos. Ordenamento aleatório.

Os novos resultados obtidos são surpreendentemente diferentes dos anteriores. Não parece haver qualquer relação entre a taxa de reação e a temperatura e os erros da medida são surpreendentemente altos. O experimentador resolve então observar como a atividade do catalisador varia, à medida que os experimentos se sucedem. Os resultados são mostrados na Figura 1.5. Vê-se que, independentemente da temperatura da reação, a atividade do catalisador cai continuamente, à medida que os experimentos se sucedem. O experimentador percebe então que o catalisador perde atividade com o tempo, o que não havia sido imaginado inicialmente. Sabe-se, no entanto, que esse é um problema comum observado em catalisadores químicos. O experimentador verifica assim que sua primeira conclusão estava equivocada e que havia sido induzido ao erro por causa de um fator externo não considerado na análise: o efeito do tempo.

O experimentador aprende que, sempre que possível, para filtrar efeitos externos não controlados, a seqüência experimental deve ser aleatorizada.

Exemplo 1.5- Suponha que um experimentador está estudando como ocorre a dessorção da amônia em um certo catalisador. Esse tipo de experimento é muito comum na catálise. O experimento consiste em impregnar um catalisador sólido com amônia e ver com que velocidade a amônia dessorve (ou seja, desprende do suporte sólido), à medida que a temperatura do sistema é aumentada. O chamado espectro de dessorção (ou TPD - Dessorção a Temperatura Programada), como mostrado na Figura 1.6, pode ser constituído por picos finos, a depender da força de atração existente entre a amônia e o sólido.

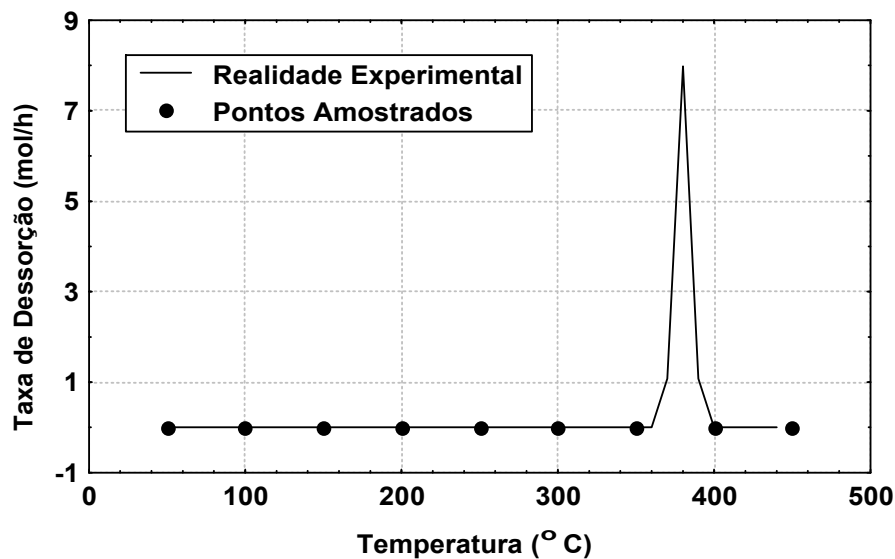


Figura 1.6- Taxa de dessorção como função da temperatura.

Repare que, nessas condições, a utilização de intervalos regulares para definir as condições de experimentação pode ser desastrosa para a análise. Embora na faixa experimental estudada exista uma probabilidade de aproximadamente 10% de que um ponto escolhido aleatoriamente segundo a distribuição uniforme venha a cair exatamente na região onde os efeitos se manifestam, o espaçamento uniforme de até 10 pontos experimentais na faixa de experimentação jamais permitirá que os efeitos de dessorção sejam observados experimentalmente. Por isso, o experimentador deve ter muito cuidado com a formulação de leis determinísticas para seleção de pontos experimentais durante o processo de experimentação.

Exemplo 1.6- Suponha que um experimentador está estudando o comportamento dinâmico de um processo. Para tanto, ele perturba o processo (mudando, por exemplo, uma condição de operação) e observa como a resposta varia no tempo. Esse é um procedimento muito comum em plantas industriais reais e durante a realização de estudos de controle de processos. A Figura 1.7 ilustra uma resposta oscilatória típica que pode ser observada durante esse tipo de experimento. É muito importante verificar que, toda vez que um padrão regular de observação

experimental é imposto ao problema, as oscilações experimentais reais podem ser completamente deturpadas pelo processo de observação. Para isto, basta que a frequência de amostragem seja inferior à frequência natural de oscilação do processo (ou que o intervalo de amostragem seja superior ao período de oscilação), como muito conhecido na área de dinâmica de processos.

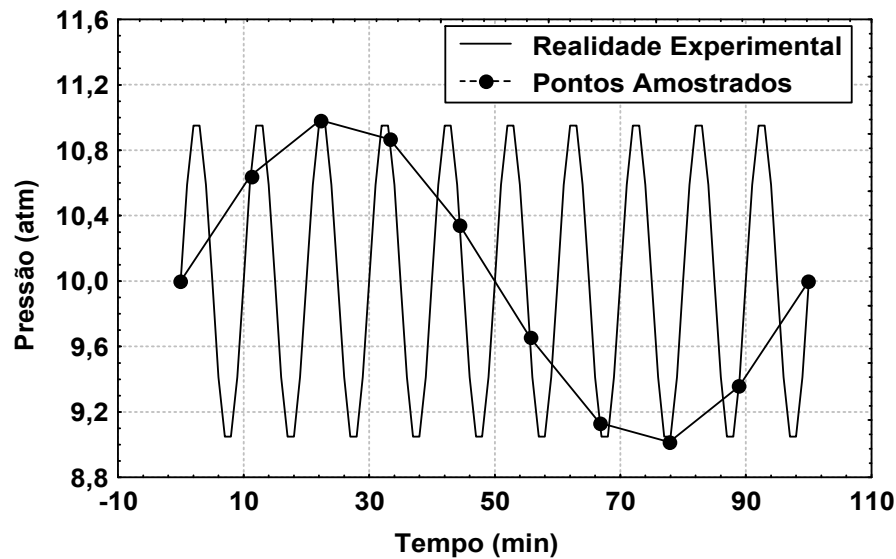


Figura 1.7- Flutuação da pressão no tempo em um experimento dinâmico.

Por isso, mostra-se uma vez mais que o experimentador deve ter muito cuidado com a formulação de leis determinísticas para seleção de pontos experimentais durante o processo de experimentação.

O ideal estatístico sugere, portanto, o algoritmo de planejamento experimental apresentado a seguir.

Algoritmo 1.1- O plano experimental ideal.

1. Defina o conjunto de variáveis independentes $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_{NX})$;
2. Defina as faixas experimentais de investigação para cada variável $x_j^{min} \leq x_j \leq x_j^{max}$, $j = 1 \dots NX$;
3. Defina o número de condições experimentais NL que serão realizadas, com NL suficientemente grande;
4. Defina o número de réplicas NR que serão realizadas em cada condição experimental, com NR suficientemente grande;
5. Para cada condição experimental $i = 1 \dots NL$ faça:
 - 5.1. Para cada variável x_j , $j = 1 \dots NX$ faça:
 - 5.1.1. Selecione a condição experimental x_{ji} aleatoriamente na faixa experimental especificada;

- 5.2. Faça $\mathbf{x}_i^T = (x_{1i} \ x_{2i} \ \dots \ x_{NX \ i})$;
- 5.3. Para a quantidade de réplicas $k = 1 \dots NR$ faça:
- 5.3.1. $\mathbf{X}_n^T = \mathbf{x}_i^T$, $n = (i-1)NR + k$;
6. Selecione a ordem dos experimentos $n = 1 \dots NL \times NR$ aleatoriamente na matriz de condições experimentais \mathbf{X}^T .

Apesar do Algoritmo 1.1 ter sido gerado a partir de raízes matemáticas e práticas muito claras, é de fato impossível atingir o ideal estatístico. Primeiramente, a noção de suficientemente grande imposta a NL e NR é imprecisa. O que é um número suficientemente grande? Como definir a quantidade de condições e réplicas de maneira clara? O plano ideal do Algoritmo 1.1 também pressupõe que a região experimental tem a forma de um hiper-cubo, dado que qualquer combinação de variáveis que atenda as limitações impostas pelas restrições experimentais definidas no passo 2 para cada variável independentemente é aceita. No entanto, em muitos casos certas combinações não são de fato possíveis experimentalmente. Por exemplo, devido a limitações impostas por procedimentos de segurança, freqüentemente é possível trabalhar com pressões altas se as temperaturas são baixas e vice-versa, mas não é possível trabalhar em condições de altas pressões e temperaturas simultaneamente. Além disso, a aleatorização da escolha das condições experimentais e do ordenamento dos experimentos pode resultar em problemas como aqueles discutidos na Seção 1.1, quando ao menos uma das variáveis experimentais analisadas é muito difícil de manipular. Problemas adicionais são discutidos no Exemplo 1.7 apresentado a seguir.

Exemplo 1.7- Um pesquisador deseja conhecer o efeito causado pela adição de um certo composto A (um agente anti-oxidante) na estabilidade química de um plástico que recobre móveis de jardim submetidos constantemente à radiação solar. Esse tipo de problema é conhecido classicamente como um problema de formulação. Para tanto, o pesquisador monitora o tempo necessário para amarelecimento do plástico (medido como o tempo necessário para que a transmissão da luz em um certo comprimento de onda atinja um valor 90% inferior ao do composto original), quando os móveis são expostos ao ambiente natural, sem qualquer proteção. Para tanto, o pesquisador usa o Algoritmo 1.1, do plano experimental ideal.

Segundo o Algoritmo 1.1, o vetor de variáveis independentes é constituído por uma única variável ($NX = 1$), $\mathbf{x}^T = (x_1)$, que é a concentração de agente anti-oxidante utilizada. O vetor de variáveis dependentes é constituído por uma única variável ($NY = 1$), $\mathbf{y}^T = (y_1)$, que é o tempo necessário para que a transmissão de luz caia a 90% do valor original, como medido em um espectrofotômetro. A faixa de interesse é definida pelo usuário como $0 \leq x_1 \leq 100$ ppm (1 ppm - partes por milhão - equivale a 1 mg de anti-oxidante em 1 kg de material plástico), em função de restrições econômicas, sanitárias e ambientais, que limitam a quantidade máxima admissível de anti-oxidante no produto final. O número máximo admissível de condições experimentais é definido como $NL = 8$. De forma similar, o número máximo admissível de réplicas experimentais é definido como $NR = 8$. O plano experimental será composto, portanto, por 64 experimentos ($NE = NL \times NR = 8 \times 8 = 64$) e a matriz experimental \mathbf{X} terá $NX=1$ linhas e $NE=64$ colunas. Admite-se aqui que essa quantidade de experimentos é compatível com o orçamento e o cronograma do projeto.

Um gerador de números aleatórios semelhante aos apresentados na Seção 2.4 do Volume I desta série de publicações é usado para definir as condições experimentais apresentadas na Tabela 1.1 e na Figura 1.8.

Tabela 1.1- Condições experimentais selecionadas pelo Algoritmo 1.1.

i	1	2	3	4	5	6	7	8
x_{1i}	0.114722	0.375215	0.592371	3.119766	25.858977	26.331239	58.218344	73.472172

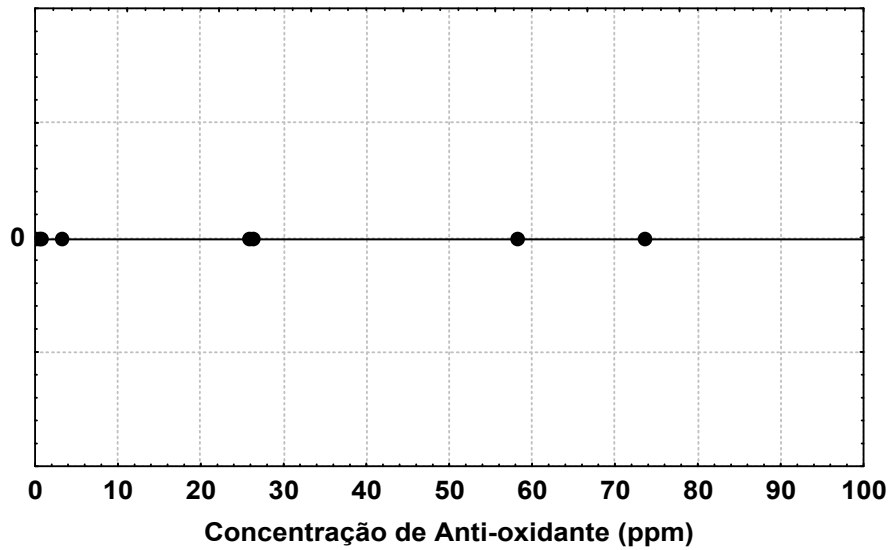


Figura 1.8- Representação gráfica das condições experimentais selecionadas para o Exemplo 1.7.

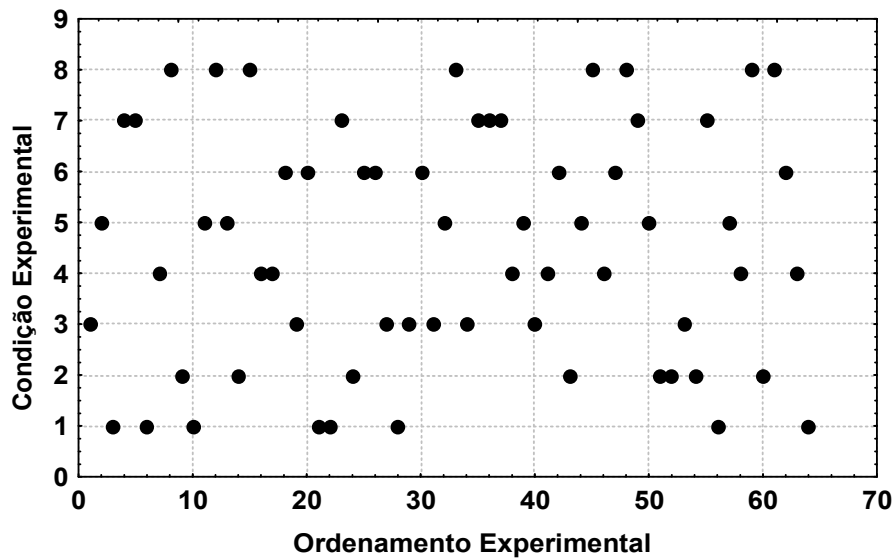


Figura 1.9- Ordenamento experimental aleatório para o Exemplo 1.7.

Cada condição experimental é repetida 8 vezes. A seqüência de execução dos experimentos é gerada aleatoriamente, como mostrado na Figura 1.9. Os dados experimentais

obtidos são organizados na forma de uma tabela em que, para cada condição experimental, são armazenados o tempo médio de amarelecimento do plástico e a variância experimental obtida para essa medida. Dessa forma, é possível avaliar com o teste t para as médias e com o teste F para as variâncias (ver Seções 3.3.1 e 3.3.3 do Volume I desta série de publicações) se a estabilidade média do material e se os erros de medida mudam, à medida que muda a concentração de anti-oxidante utilizada.

Independentemente dos resultados obtidos pelo pesquisador, o plano experimental foi apresentado a dois profissionais, que apresentaram os seguintes conjuntos de críticas:

1- Sobre o número de condições experimentais:

Professional 1: $NL = 8$ condições experimentais é muito pouco! Não é possível construir uma dependência funcional precisa com tão poucas condições experimentais.

Professional 2: $NL = 8$ condições experimentais é muito, pois os experimentos são longos e caros!

2- Sobre o número de réplicas:

Professional 1: $NR = 8$ réplicas é muito pouco! Não é possível identificar as distribuições dos erros experimentais e os intervalos de confiança são largos, em função do baixo número de graus de liberdade.

Professional 2: $NR = 8$ réplicas é muito! A maior parte do esforço experimental é dedicada a repetir o que já se sabia, aumentando ainda mais o custo e o tempo do estudo experimental.

3- Sobre o número total de experimentos:

Professional 1: Como discutido antes, pode ser que o número de dados seja insuficiente.

Professional 2: $NE = NL \times NR = 64$ é um absurdo!

4- Sobre a escolha das condições experimentais:

Professional 1: Escolher as condições experimentais de forma aleatória foi uma boa idéia, mas NL deveria ser maior.

Professional 2: A escolha aleatória das condições experimentais não faz qualquer sentido! Primeiramente, não é possível reproduzir os experimentos com a precisão pretendida para a concentração de anti-oxidante. E em segundo lugar, quantidade apreciável de condições experimentais está concentrada em regiões extremamente estreitas da região experimental de interesse, como mostram a Tabela 1.1 e a Figura 1.8.

5- Sobre o ordenamento experimental:

Professional 1: O ordenamento aleatório da seqüência experimental ajuda e filtra fatores externos.

Professional 2: O ordenamento aleatório da Figura 1.9 é muito ruim! Não é possível preparar pequenas quantidades de mistura de plástico e anti-oxidante na extrusora. Portanto, todo o material que contenha a mesma quantidade de anti-oxidante deve ser preparado de uma só vez; caso contrário, os custos do estudo experimental são excessivamente grandes, tornando o projeto inviável.

6- Sobre a natureza dos erros experimentais:

Professional 1: Quais são os erros experimentais associados à concentração de anti-oxidante no material? Como esses erros estão correlacionados às flutuações experimentais

observadas no tempo de amarelecimento? É possível garantir que os erros experimentais são verdadeiramente independentes?

Professional 2: Sabe-se com grande precisão a concentração de anti-oxidante no produto final e como a extrusora é limpa entre um experimento e outro, pode-se admitir que as operações de formulação das misturas são independentes. E não há por que imaginar que os erros experimentais mudem de ponto para ponto, já que as mesmas técnicas e equipamentos são utilizados sempre.

E agora ? Como o pesquisador deve se posicionar diante das críticas levantadas por cada um dos profissionais consultados, uma vez que as opiniões são em todos os casos tão díspares? É importante que o leitor perceba que as críticas apresentadas anteriormente são frequentemente encaminhadas a técnicos que militam na área de planejamento experimental e análise de dados. Poderiam ambos os profissionais consultados estarem certos? Poderiam ambos estarem errados?

O Exemplo 1.7 apresentado acima mostra que o experimentador é instado permanentemente a procurar algum tipo de compromisso entre o ideal estatístico e a realidade prática. No Exemplo 1.7, o *Professional 1*, com provável formação teórica no assunto, força o pesquisador a se posicionar em relação a pontos fundamentais, como a natureza dos erros experimentais. O *Professional 1* reconhece a importância da caracterização apropriada dos erros experimentais para a análise quantitativa dos dados e sabe que o grau de incerteza das conclusões eventualmente tiradas cai lentamente com o aumento do número de dados experimentais. O *Professional 2* provavelmente tem larga experiência em trabalhos de chão de fábrica, onde o custo e o tempo são fundamentais para definir o sucesso de uma empreitada. O *Professional 2* força o pesquisador a tomar atitudes mais ousadas em relação ao plano experimental, que podem talvez comprometer a qualidade da análise mas podem também permitir uma tomada de decisão mais rápida. E aí? Como se posicionar diante desse problema?

Independentemente do posicionamento individual de cada um frente ao problema apresentado no Exemplo 1.7, uma coisa é certa: poucos estudos experimentais podem efetivamente realizar número tão grande de medidas experimentais ($NE = NL \times NR = 64$) para o estudo de um único efeito (entender como o aumento da concentração de aditivo afeta o desempenho óptico do material), como o proposto no Exemplo 1.7. Além disso, a quarta crítica levantada pelo *Professional 2*, a respeito da cobertura desuniforme da região de experimentação, será provavelmente acolhida pela grande maioria dos experimentadores. Há, portanto, que se repensar o problema e avaliar conseqüências. Isso será feito a partir de agora, nos capítulos que se seguem.

1.4. Conclusões

Neste capítulo foram introduzidos e discutidos alguns conceitos preliminares importantes a respeito do procedimento de planejamento de experimentos, como os objetivos perseguidos durante o planejamento experimental, a importância da aleatorização e as limitações impostas pelo cronograma e orçamento do projeto. Esses conceitos devem ser compreendidos antes do desenvolvimento das técnicas matemáticas e numéricas que caracterizam o procedimento de planejamento experimental propriamente dito. Também foi largamente discutido e mostrado que não existe receita universal para o planejamento de experimentos, uma vez que cada

processo estudado apresenta características que devem ser levadas em consideração durante a construção do planejamento de experimentos. Particularmente, a discussão apresentada no Exemplo 1.7 colocou frente a frente opiniões de dois profissionais que apresentam pontos de vista muito distintos a respeito das expectativas de avaliação experimental. Como discutido nos próximos capítulos, a construção de um bom planejamento de experimentos requer o adequado balanceamento entre as opiniões de ambos os profissionais – ao mesmo tempo que a realização de um grande número de experimentos (preferencialmente de forma aleatória) é muito importante para a análise quantitativa dos dados, um número muito grande de experimentos pode também inviabilizar a execução do estudo científico. Por isso, nos próximos capítulos são abordados os procedimentos clássicos de planejamento de experimentos e os procedimentos avançados, que envolvem a técnica de planejamento seqüencial de experimentos baseadas na modelagem matemática dos processos.

1.5. Leitura Adicional

A literatura relacionada aos procedimentos de planejamento de experimentos é muito vasta, já que estas técnicas encontram aplicações nas mais variadas áreas relacionadas à investigação tecnológica e/ou científica. Assim, não cabe aqui uma revisão detalhada de todos os livros que tratam deste importante tema.

O leitor interessado numa revisão dos conceitos básicos de estatística e de estimação de parâmetros, fundamentais para a correta compreensão dos princípios que fundamentam os procedimentos de planejamento experimental apresentados nos próximos capítulos, sugere-se a leitura do Volume I desta série de publicações:

“Análise de Dados Experimentais I. Fundamentos de Estatística e Estimação de Parâmetros”, M. Schwaab e J.C. Pinto, E-papers, Rio de Janeiro, 2007.

O leitor interessado também pode encontrar motivação para analisar dois textos clássicos relacionados ao uso de técnicas estatísticas para a análise de dados e processos e para o planejamento de experimentos:

“Process Analysis by Statistical Methods”, D.M. Himmelblau, John Wiley & Sons, New York, 1970.

“Statistics for Experimenters. Design, Innovation, and Discovery”, 2ª Edição, G.E.P. Box, W.G. Hunter e J.S. Hunter, John Wiley & Sons, New York, 2005.

1.6. Exercícios Sugeridos

1- Construa um planejamento experimental para o problema que tem apenas uma única variável manipulável (x_1) e uma única variável de resposta (y_1), relacionadas na forma:

$$y_1 = \text{sen}(2 \pi x_1)$$

Admita que $x_1^{\min} = 0$ e que $x_1^{\max} = 1$. Execute as seguintes tarefas:

- a) Verifique os resultados obtidos quando os pontos são distribuídos regularmente na região experimental, admitindo a existência de $NL = 3, 4, 5$ e 10 condições de experimentação. O que pode ser concluído?
- b) Repita o exercício proposto no item anterior, mas distribua os pontos de forma aleatória. Que diferenças podem ser observadas?
- 2- Suponha que um investigador (cozinheiro, nesse caso) pretende avaliar a influência que um tempero (por exemplo, a quantidade de sal) exerce sobre a qualidade de um prato comestível (por exemplo, uma fornada de arroz). De acordo com a lista de perguntas apresentadas ao final da Seção 1.2, como o experimentador deveria se posicionar diante do problema?
- 3- Repita o problema proposto no exercício anterior, mas admita que o prato comestível analisado é uma fornada de camarão refogado. Será que seriam necessárias mudanças de procedimento e de medição? Discuta.
- 4- Proponha um planejamento experimental para o problema proposto no Exercício 1.2, de acordo com o Algoritmo 1.1, admitindo que uma única variável independente é manipulada (por exemplo, o teor de sal). Refaça o planejamento, admitindo que três variáveis independentes podem ser manipuladas (por exemplo, quantidade de água, teor de sal e tempo de cozimento).
- 5- Analise o comportamento da função

$$y_1 = \frac{1}{1 + \exp(-\alpha x_1)}$$

e de sua derivada

$$y_2 = \alpha \left[\frac{1}{1 + \exp(-\alpha x_1)} \right]^2 \exp(-\alpha x_1)$$

no intervalo limitado por $x_1^{\min} = -1$ e que $x_1^{\max} = 1$. Admita que o parâmetro α pode assumir os valores iguais a $0, 1, 10$ e 100 . Se a amostragem de x_1 for regular no intervalo analisado, o que pode ser concluído?