

Qual o delineamento e quantas observações devo considerar em meu projeto?

Ivan Barbosa Machado Sampaio*

Em qualquer área da experimentação, um projeto de pesquisa depende de planejamento prévio que permita a consecução de seus objetivos, obtidos de um ensaio viável, tanto para as fontes financiadoras como para a ciência (**Slide 1**).

O dimensionamento físico da pesquisa (instalações, animais, entrevistas, etc) refletirá os custos do projeto e conseqüentemente sua viabilidade econômica. Já a viabilidade científica dependerá de um grande número de fatores que se apresentam na experimentação e que precisa ser controlado adequadamente, para que as conclusões decorrentes do processo de investigação não contenham confundimentos que as invalidem. Esse tipo de ação é proporcionado pela Estatística, que, por meio de um planejamento prévio e uma análise adequada de resultados, fornece conclusões robustas e acuradas.

Antes de discutir a ação desses fatores e como controlá-los, vamos comentar a respeito da variação de respostas experimentais (**Slide 2**). O exemplo aqui citado utiliza a resposta animal, mas pode ser estendido para qualquer tipo de variável. Se selecionarmos coelhos machos de uma mesma raça com praticamente o mesmo peso à desmama (28 dias), julgamos obter uma amostra uniforme que será mantida em um mesmo ambiente sob o mesmo manejo alimentar. O objetivo do ensaio é verificar o efeito do manejo alimentar sobre o peso final dos animais aos 45 dias. Para nossa surpresa, àquela idade os animais mostram uma grande variação de peso. Para a ciência, a média desses pesos definirá o efeito provável daquela ração, mas a variação observada entre as respostas individuais (desvio padrão) refletirá na instabilidade daquele tipo de resposta. Se utilizarmos grupos de diferentes rações, suas médias variarão, dependendo da qualidade de cada uma, mas as instabilidades dentro de cada grupo serão as mesmas, pois a capacidade de variar é inerente à resposta estudada.

Mesmo quando estudamos uma mesma espécie, digamos, vacas leiteiras de terceira parição, respostas diferentes podem apresentar instabilidades distintas. Quando estudamos a produção de leite desses animais, a produção média é de 22 L diários, com uma variação média (desvio padrão) de 8 L/dia. Já se avaliarmos o período de gestação delas, a média será 278 dias, com um desvio de 5 dias (**Slide 3**).

*Professor Emérito da Escola de Veterinária da UFMG

Seminário patrocinado pela Pró-Reitoria de Pesquisa e Inovação da Universidade Federal de Goiás, em 28/05/2015

Então percebemos que a variação de resposta em um grupo de animais é uma característica dessa variável que está sendo medida. O período de gestação, sendo uma resposta fisiológica pouco variável, apresenta uma medida de instabilidade relativa (Coeficiente de variação, $CV = \text{desvio}/\text{média}$) igual a 2%, enquanto a produção de leite apresenta um $CV = 36\%$. Assim sendo, o pesquisador deve estar preparado para reconhecer que existem respostas muito instáveis (geralmente aquelas dependentes de hormônios ou de resposta imunológica) e pouco instáveis. Ele saberá reconhecê-las por sua experiência ou pelo cálculo do CV. Respostas animais geralmente apresenta CV entre 20 e 30%. Outras áreas (**Slide 4**) podem estar associadas à variação menor (Agronomia) ou maior (Ciências Humanas). Precisamos da Estatística porque as inferências dela advindas precisam respeitar as características individuais de cada variável. Todas as conclusões estão sujeitas a um erro provável de 5%, nível máximo admitido pela pesquisa. Para que todos esses mecanismos funcionem plenamente, precisamos acatar os princípios básicos da experimentação, a saber:

- a) Número de repetições (**Slide 5**): ponto crucial no planejamento, pois além de permitir o cálculo da instabilidade (ponto básico para a estatística), refletirá em custos. Quanto mais instável (maior CV) for a resposta, maior será a amostragem exigida. Se a confiabilidade em uma média precisar ser aumentada, maior será o número de observações que gerarão aquela média.
- b) Casualização das unidades experimentais aos grupos testados (**Slide 6**): a variação observada nas respostas indica que há uma grande frequência de valores em torno da média, frequência esta que vai diminuindo simetricamente à medida que os valores se distanciam da média. Esse tipo de distribuição de frequência é chamado de distribuição normal. Quando a variável é pouco instável, a amplitude de variação é menor. Como não temos domínio sobre a variabilidade de respostas, a ciência considera como respostas típicas aquelas que se encontram no centro de distribuição (95%). Os valores muito baixos (2,5%) e muito altos (2,5%) são considerados atípicos e ignorados nas discussões de resultados. Daí o erro probabilístico de 5% presente em nossas conclusões.
- c) Uniformidade de amostra (**Slide 7**): para que os tratamentos possam ser avaliados corretamente, a amostra inicial deve ser constituída de unidades experimentais uniformes. Sempre ocorrerá uma variação individual dentro de cada grupo, mas como já vimos, ela será igual em cada tratamento e, ao final, se houver um grupo melhor que os outros, esse destaque será revelado na comparação entre as médias dos tratamentos. Se o grupo inicial não for uniforme, deveremos separá-lo em subgrupos uniformes e cada subgrupo (bloco) deverá conter equitativamente todos os tratamentos. No slide 7, por exemplo, há animais que, embora na mesma idade fisiológica, apresentam

pesos diferenciados: um grupo mais pesado e outro mais leve. Em cada tratamento há três animais mais pesados e apenas um leve. Posteriormente, em uma análise adequada, o efeito de bloco (peso inicial) será controlado e as médias dos três tratamentos poderão ser comparadas sem o concurso daquele efeito. O efeito de blocos, quando existente, não deve interagir com o efeito de tratamentos. Se a amostra inicial contivesse seis machos e seis fêmeas para cada tratamentos, sexo não poderia ser bloqueado, pois a indicação de tratamento poderia diferir entre machos e fêmeas. Neste caso sexo seria um fator e os grupos formados seriam definidos por um fatorial 3x2 (três rações e dois sexos), com seis repetições.

- d) Uniformidade de meio (**Slide 8**): se as unidades experimentais forem uniformes, o meio sob o qual elas serão criadas também deve ser uniforme (como em uma estufa com temperatura e umidade controladas). Só assim poderemos atribuir aos tratamentos as eventuais diferenças entre suas médias. Quando o meio não for uniforme, como no slide 8, apenas as médias dos tratamentos A e B poderiam ser comparadas porque provieram de amostragem inicialmente uniforme e conduzidas em meio uniforme até o resultado final. O tratamento C foi instalado sob condições distintas das de A e B e, portanto, a média de C contém um efeito confundido com o efeito do segundo galpão. A solução para este problema seria permitir que os três grupos experimentais participassem igualmente de cada galpão (bloco) e, desta maneira, o efeito de blocos seria retirado em uma análise posterior, permitindo uma comparação com menor erro entre as médias dos tratamentos.
- e) Uniformidade de aplicação de tratamento (**Slide 9**): cada tratamento deve permitir que os animais nele presentes desfrutem igualmente de seus efeitos. Se um cocho de comprimento reduzido propiciar mais acesso ao alimento ali contido a animais mais fortes e com hierarquia superior sobre os demais em um mesmo piquete, haverá uma variação maior naquele piquete resultante da ação hierárquica (além da variação individual que certamente ocorre, mesmo sem hierarquia). A solução desse problema, (que pode inutilizar um ensaio) só pode ser preventiva: livre acesso ao alimento em cochos individuais. Injeções hormonais devem ser volumetricamente precisas, pois variações mínimas causam variações dramáticas de respostas.

Devemos ressaltar que as respostas medidas (aqui denominadas “variáveis”) em qualquer área têm naturezas diferentes e que todo o planejamento e análise de um ensaio deverão respeitá-las, para que os objetivos da pesquisa sejam alcançados sem questionamentos. Uma variável é dita quantitativa quando for expressa em magnitude numérica. Ainda assim, elas podem se apresentar com características específicas, a saber (**Slide 10**):

- a) Variáveis normais: valores quantitativos, contínuos e com distribuição de frequência normal. Exemplos: altura na cernelha em equinos adultos, produção de leite, tempo de cicatrização, peso ao nascer, etc. A grande maioria das respostas biológicas se enquadra nesta condição e demanda estratégia de análise paramétrica (análise de variância).
- b) Variáveis não normais: valores quantitativos, contínuos, mas com distribuição não normal. Exemplos: cortisol em vacas, concentração de anticorpos, basófilos na crase sanguínea. Geralmente resultados imunológicos e parasitológicos, se quantitativos e contínuos, se enquadram nesta condição, a qual inclui variáveis com coeficiente de variação acima de 50%. Para esses casos, a estratégia de análise será a não paramétrica.
- c) Variáveis subjetivas: valores quantitativos, mas não contínuos, logo, sem normalidade. Advêm de uma avaliação subjetiva do pesquisador. Exemplos: vigor espermático, escore corporal de vacas pós-parto, avaliação sensorial de vinhos, etc. Demandam análise não paramétrica.
- d) Variáveis discretas: valores quantitativos, mas descontinuados, logo sem normalidade, geralmente produto de contagem. Exemplos: número de ectoparasitas, número de oócitos, ovos de helmintos por grama de fezes, etc. Esse tipo de resposta demanda estratégia não paramétrica para sua análise.

Uma resposta é dita categórica quando não existe magnitude para caracterizá-la, e sim uma determinada classificação. Neste caso, o resultado que a Estatística considera para análise é a frequência que cada categoria de resposta ocorre. Dependendo da existência de grupos experimentais, o estudo exigido, sempre não paramétrico, será:

- e) Estudo de prevalência: quando há apenas um grupo experimental (geralmente um levantamento), onde será estudada a dispersão de frequência das respostas obtidas em levantamentos. Exemplos: diagnóstico de doenças (positivo, suspeito, negativo), APGAR de recém-nascido em humanos (baixo, médio, alto), prenhez (sim e não), displasia coxo-femural em cães da raça Labrador, etc. O tratamento estatístico requer a definição da prevalência média com seu respectivo intervalo de confiança.
- f) Estudo de distribuição de frequência segundo a resposta categórica e os grupos experimentais utilizados, formando os quadros chamados de tabelas de contingência. Exemplos: estudo de prenhez em éguas (positiva e negativa) segundo o diluente de sêmen utilizado na inseminação (gema de ovo, água de coco e citrato), formando uma tabela 3x2. Sobrevivência de camundongos (sim e não) segundo o grupo estudado (controle e vacinado), gerando uma tabela 2x2. Essas tabelas serão analisadas pelo teste de χ^2 (qui-quadrado).

As situações de “a” a “d” acima citadas indicam apenas parcialmente o tipo de estratégia exigida (paramétrica ou não paramétrica). Dependendo da presença de eventuais fatores experimentais que estiverem atuantes no ensaio, haverá sempre um delineamento específico para controlá-los, deixando que apenas os tratamentos que estão sendo testados sejam devidamente comparados pelo pesquisador. A confiabilidade dos resultados dependerá da correta definição do delineamento (acurácia) e do número de repetições (animais) por tratamento (robustez), de modo que as conclusões finais sejam válidas e incontestáveis. As apresentações a seguir caracterizam a contribuição da Estatística à experimentação nas ciências em geral.

Quando a experimentação exigir o sacrifício de animais, a Estatística poderá intervir, de modo que um número mínimo deles conduza a um resultado robusto e contemple as exigências da Ética (**Slide 11**).

Se os princípios básicos da experimentação não forem atendidos, a Estatística poderá controlar os efeitos eventualmente existentes (desuniformidade amostral em termos de peso e sexo, de ambiente e temporal) por meio de um delineamento adequado (**Slide 12**), gerando:

- 1) Ensaios pareado: quando há apenas dois grupos experimentais que podem ser testados sobre o mesmo animal, em um lapso reduzido de tempo ou ao mesmo tempo (no caso de alíquotagem). Exemplos: aferição de pressão diastólica em cavalos antes e depois de exercício em esteira, mensuração do vigor em duas alíquotas de sêmen submetidas cada uma a um diluente (gema e água de coco), 24h depois do resfriamento a 4°C (**Slide 13**). É um delineamento poderoso por contornar a variabilidade entre animais, mas seu uso é limitado por testar apenas dois tratamentos e exigir amostras idênticas (repicagem de um mesmo tecido, alíquotas de sangue ou sêmen, situações de antes e depois em um mesmo indivíduo).
- 2) Delineamento Inteiramente Casualizado: quando a amostra for uniforme, os animais forem sorteados aos tratamentos, e o ensaio for realizado sob as mesmas condições de meio. Se, ao final, ficar comprovada uma diferença entre os tratamentos, ela só poderá ser atribuída aos tratamentos envolvidos na comparação, já que a amostra e o meio eram uniformes para todos os grupos experimentais (**Slide 14**). Em termos de amostragem, precisaremos de pelo menos 10 graus de liberdade para o erro experimental, para garantir uma confiabilidade (acurácia) desejável.
- 3) Delineamento em Blocos ao Acaso: quando houver desuniformidade (de meio, de amostragem, de tempo), será sempre possível submeter todos os tratamentos a cada uma dessas condições formando então os blocos (que seriam subcondições uniformes para os tratamentos. O fator implícito em

blocos não poderá interagir com os tratamentos. Essa fonte de variação é então controlada na análise pertinente, permitindo a avaliação do erro experimental sem o concurso dessa fonte. Exemplo: testar o desempenho de frangos de corte em cinco rações com diferentes níveis de proteína em galpões que possuem apenas sete boxes cada. Se precisarmos de quatro repetições (boxes), então serão necessários quatro galpões (blocos), cada um com os cinco grupos experimentais. Também, se desejarmos testar cinco diluentes para verificar o teor de frutose no mesmo sêmen após 24h de conservação a 4°C: como um ejaculado de um reprodutor suíno pode ser aliqüotado em até 40 amostras, poderemos designar cada diluente a uma alíquota do mesmo varrão (**Slide 15**). Se usarmos quatro reprodutores (blocos), obteremos 12 graus de liberdade para o erro experimental, como planejado.

- 4) Delineamento em Quadrado Latino: quando existirem dois fatores circunstanciais que possam ser bloqueados. Exemplo: mais comuns em agricultura, as gradações de fertilidade em dois sentidos exigem a formação de blocos em dois sentidos (em relação aos dois sentidos de declividade). Com animais, a mensuração de degradabilidade de cinco rações para suínos, oferecidas sequencialmente ao mesmo animal (**Slide 16**), de modo que todos os animais (blocos horizontais) recebam os cinco tratamentos e em cada tempo sequencial, todos os tratamentos estejam presentes em cada coluna (blocos verticais). Note a preocupação do planejador de garantir o mínimo de 10 graus de liberdade para o erro experimental.
- 5) Delineamento em Parcelas Subdivididas: em ensaios fatoriais, quando não for possível casualizar todas as combinações possíveis para formar um bloco, por motivos operacionais ou de restrição amostral. Por exemplo: para testar a ação do fatorial 4x3 (quatro diluentes e três tempos de conservação do sêmen) sobre os valores de frutose remanescentes, não seria possível obter 4x3=12 alíquotas de um mesmo ejaculado equino (**Slide 17**). O máximo de alíquotas para um ejaculado equino é de quatro amostras aliqüotadas. A solução seria obter as quatro alíquotas por garanhão (blocos), designar um diluente (parcela) para cada uma, mantê-las a 4°C e no tempo oportuno (subparcelas) analisar a quantidade de frutose. Os tratamentos (diluentes) e os subtratamentos (tempos) podem interagir. Note que a análise apresenta dois erros experimentais. É desejável que os dois erros apresentem graus de liberdade superiores ou próximos de 10, principalmente o erro(b). Nos ensaios com animais, geralmente a subparcela representa o fator tempo.

Assim sendo, a estratégia de análise de uma resposta depende da natureza dela e das condições experimentais do ensaio (**Slide 18**). Há sempre uma correspondência de cada delineamento, registradas por seus autores, entre as estratégias paramétrica e a não paramétrica, exceto para os delineamentos em Quadrado

Latino e em Parcelas Subdivididas, como veremos na Tabela 1. Variáveis categóricas exigem maiores amostragens.

Tabela 1: Estratégia de análise segundo a natureza da variável e as condições experimentais vigentes.

Condições experimentais	Natureza da variável	
	Paramétrica	Não paramétrica
Pareamento	Ensaio pareados	Teste de Wilcoxon
Inteira- Casualizado	$t^*=2$ Inteira-ao Acaso	Teste de Mann-Whitney
	$t>2$ Inteira-Casualizado	Teste de Kruskal-Wallis
Blocos ao Acaso	Blocos ao Acaso	Teste de Friedman
Quadrado Latino	Quadrado Latino	Teste de Kruskal-Wallis
Parcelas Subdivididas	Parcelas Subdivididas	Teste de Kruskal-Wallis
Associação de respostas	Correlação de Pearson	Correlação de Spearman

* t =número de tratamentos

Existem vários testes estatísticos para a tarefa final de comparação de médias entre os diversos grupos experimentais (Slide 19). A abordagem não paramétrica apresenta testes específicos para cada delineamento.

Já nas situações paramétricas, os testes dependem do número de tratamentos e dos coeficientes de variação do ensaio. Quando temos apenas dois tratamentos, o teste universal é o de Student, que na realidade pode ser aplicado até um total de quatro tratamentos. A partir de cinco tratamentos a recomendação é pelo teste de Duncan, independentemente do CV do ensaio.

Outros testes, mais rigorosos, como Student-Newman-Keuls e Tukey não devem ser alternativas válidas para ensaios com altos valores de coeficiente de variação ($CV > 30\%$), pois tendem a apresentar resultados inconclusivos devido aos seus grandes intervalos de confiança.

Os pesquisadores, por sua vez, possuem um julgamento crítico que os permitem antecipar algumas conclusões a partir dos resultados obtidos. Se essas conclusões

não forem confirmadas pela análise estatística, certamente este fato será produto de uma estratégia de análise inadequada, de uma amostragem insuficiente ou de um teste inapropriado.

Quando um ensaio apresenta várias respostas medidas, o pesquisador pode se interessar pela associação entre cada par delas. Se as duas variáveis estão dentro da abordagem paramétrica, a correlação entre elas é medida pelo coeficiente de correlação de Pearson. Quando pelo menos uma delas é de natureza não paramétrica, o coeficiente de correlação de Spearman é o método de avaliação indicado.

É preciso alertar que em um mesmo ensaio, embora as condições experimentais sejam as mesmas, podem existir variáveis de diferentes naturezas, e, portanto, haverá diferentes estratégias de análise, dependendo da abordagem em pauta.

Considerações finais (Slide 20)

As principais contribuições da Estatística para o dimensionamento de projetos em qualquer área da ciência são:

- Definir um número de repetições que se adeque à robustez da pesquisa, às exigências da Ética e à exequibilidade do projeto.
- Contornar a desuniformidade da amostra, do meio e do tempo, pela utilização de delineamentos experimentais que controlem tais efeitos.
- Definir um teste estatístico adequado segundo a natureza e a instabilidade da variável estudada
- Permitir que o pesquisador confirme suas percepções técnicas com acurácia, sem a intervenção de confundimentos.