

# Bases Teóricas da Estatística Experimental

---

*Francisco José Pfeilsticker Zimmermann*

## Introdução

Qualquer programa de melhoramento genético de uma espécie vegetal pode ser dividido em etapas, tais como, seleção de genitores, cruzamentos, desenvolvimento e avaliação das linhagens, ensaios de rendimento e recomendação. As fases de seleção de genitores, avaliação de linhagens e ensaios de rendimento são aquelas em que a estatística pode desempenhar um papel importante como um instrumental de auxílio ao melhorista na tomada de decisão. O número de tratamentos (linhagens ou cultivares), a disponibilidade de sementes e o grau de precisão requerido naquela etapa são os fatores mais importantes na definição dos delineamentos e dos testes estatísticos a serem empregados, quando necessários.

A avaliação e a seleção de genitores podem ser feitas a partir de entradas existentes em bancos de germoplasma, materiais introduzidos, de coleta ou linhagens. A avaliação de linhagens e os ensaios de rendimento, de uma forma genérica, referem-se aos materiais produzidos pelo programa de melhoramento. Para a identificação daqueles que apresentam potencial para tornarem-se futuras cultivares, os pesquisadores utilizam experimentos que recebem denominações especiais, tais como ensaios de observação, comparativos preliminares ou avançados de rendimento, e outros como ensaios estaduais, regionais ou nacionais e mais recentemente, por força de lei, ensaios de valor cultural e de uso (VCU). Em geral, esses ensaios podem ser divididos em três níveis: o inicial, que requer pouca precisão; o segundo

nível com repetições e multilocal; e o terceiro com o objetivo principal de identificar o melhor material. Várias denominações são utilizadas para descrever esses três níveis.

O ensaio de observação é conduzido tanto na fase de seleção de progenitores quanto na de avaliação de linhagens. É em geral executado em poucos locais, contrastantes, e contemplam um número elevado de tratamentos (200, 500 ou mais). Os ensaios comparativos preliminares de rendimento contemplam as linhagens selecionadas no ensaio de observação, são executados em pontos estratégicos e representam uma região macroecológica. Têm um número reduzido de tratamentos, entre 50 e 100, podendo chegar às vezes até 150. Finalmente, o ensaio avançado (VCU), que contém as melhores linhagens do ensaio preliminar, além de algumas cultivares já indicadas para cultivo (testemunhas), é conduzido em um grande número de locais e apresenta um pequeno número de tratamentos (abaixo de 50).

Para o preparo dos ensaios de observação, normalmente existe uma baixa disponibilidade de sementes das linhagens que serão avaliadas e não se faz necessário uma grande precisão, já que o objetivo é descartar aquelas de menor potencial para o programa. Portanto, nesses casos são mais frequentemente empregados os blocos aumentados de Federer (BAF) ou aqueles desenhos sem casualização com testemunhas intercalares. Nos ensaios comparativos preliminares, onde o objetivo principal é identificar as melhores linhagens para as avaliações multilocais mais precisas, são mais indicados os delineamentos reticulado quadrado e reticulado retangular. A definição do delineamento mais adequado se dará basicamente em função da disponibilidade de sementes associada ao número de ensaios a serem conduzidos. Finalmente, nos ensaios comparativos avançados (VCU), onde ocorre uma necessidade de grande precisão nos resultados experimentais, os delineamentos em blocos completos, casualizados ou mesmo os reticulados, tornam-se os mais recomendados. Nessa etapa, o número de tratamentos e o conhecimento da área experimental serão os fatores a definir o delineamento a ser empregado.

Somente o uso de desenhos experimentais adequados a cada etapa de desenvolvimento dos programas de melhoramento não garante o sucesso na identificação das melhores linhagens ou potenciais variedades. A escolha

do teste de médias correto é também um aspecto relevante na análise e interpretação dos resultados, pois a precisão e as exigências de cada teste são variáveis, fazendo com que para determinadas situações um tipo seja mais indicado que outro. Entre os testes de médias disponíveis na literatura, os mais usados são os de LSD (teste de t), Dunnett, Tukey, Duncan, Waller-Duncan e, mais recentemente, os de Ryan-Einot-Gabriel-Welsh e Scott-Knott.

## Desenhos Experimentais

A estatística intervém na pesquisa e/ou no método científico através da experimentação e da observação. O uso da estatística como ferramenta deve estar presente desde o planejamento geral do projeto de pesquisa. Se um projeto de pesquisa vai produzir dados que devem ser analisados, um método estatístico deve fazer parte do delineamento do ensaio.

Algumas definições devem ser feitas para facilitar o entendimento geral dos processos:

### *Estatísticas*

- Pesquisa e experimentação: para muitos, pesquisa e experimentação são sinônimos; e em certo sentido, assim podem ser considerados mas, se quisermos ser precisos, o termo pesquisa deve sempre ser empregado quando se investigam coisas novas e experimentação ao se verificar a adaptação de conhecimentos ou tecnologias a situações diversas àquelas nas quais foram criadas ou desenvolvidas. Assim, a criação de novas cultivares deve ser considerada como pesquisa, mas a realização de um ensaio de competição de linhagens e/ou cultivares em ambiente diverso àquele no qual foram criadas é uma experimentação.
- Fator: aquilo que se aplica em um ensaio de forma não homogênea, por exemplo, cultivar, quando se testam várias delas; adubação, ao se comparam diversas formulações; etc;
- Nível: as diferentes manifestações de um fator, por exemplo; as doses de adubação empregadas, os espaçamentos utilizados, as linhagens que se testam, diferentes temperaturas de cocção, marcas de pneus, etc;
- Tratamento: cada um dos níveis de um fator ou cada uma das combinações dos níveis dos fatores, quando testado mais de um fator;
- Testemunha: tratamento padrão de comparação. Pode ser a ausência do fator (dose zero de um adubo, por exemplo), ou a aplicação usual do fator

(cultivar recomendada para cultivo na região, espaçamento adotado pelos agricultores, etc);

- Ensaio ou experimento: o conjunto de todos os tratamentos, aplicados de forma repetida. Quando mais de um fator estiver sendo estudado, o ensaio é chamado de ensaio fatorial;
- Delineamento: o esquema adotado para a distribuição dos tratamentos;
- Unidade experimental: sujeito ao qual se aplica um dos tratamentos. Pode também ser chamada de parcela ou canteiro. Pode ser uma área de solo, um vaso, um animal, uma placa de Petri, um indivíduo, a posição de montagem de um pneu, etc;
- Área útil: porção da unidade experimental efetivamente utilizada na avaliação do tratamento.
- Bordadura: parte da unidade experimental não coletada para avaliação do efeito do tratamento. É empregada para evitar efeito de competição ou de contaminação entre parcelas vizinhas. Normalmente é constituída pelo mesmo material da área útil. Em alguns casos, pode ser montada com materiais distintos, como no caso de alguns experimentos em fitopatologia, em que as bordaduras de blocos são semeadas antecipadamente para servirem de fonte de inóculo para o ensaio, como forma de se garantir maior uniformidade de infecção.
- Repetição: cada uma das aplicações de um tratamento;
- Bloco: conjunto ambiental homogêneo que contém todos os tratamentos ou parte deles. No delineamento em quadrado latino, em que se apresentam dois tipos de blocos, eles recebem as denominações de colunas e linhas.

Em qualquer ensaio ou experimento conduzido, medidas são tomadas para representar o efeito dos tratamentos em estudo. Essas medidas são os diversos valores que as variáveis, associadas aos objetos que vêm sendo testados, podem assumir. Elas estão sempre sujeitas a erro ou variações, que são devidas à atuação de fatores não controlados sobre as unidades experimentais como, por exemplo, pequenas variações no ambiente ou material ambiental e que passam despercebidas ou não são conhecidas pelo pesquisador. Além disso, erros podem ser cometidos na tomada dos dados, como na leitura de um equipamento, balança ou trena, por exemplo. Nesse último caso, os chamados erros grosseiros. Não importa o tipo de erro que se cometa, o seu efeito sobre as variáveis representa o erro experimental. Os delineamentos experimentais então devem ser escolhidos de forma a minimizar esses erros associados às condições ambientais, enquanto que em

relação aos erros grosseiros, apenas um maior cuidado ou treinamento na tomada dessas medidas é requerido.

Os delineamentos, para a minimização dos efeitos das variações que ocorrem no ambiente em que se conduz um ensaio, são estruturados segundo alguns os princípios básicos da experimentação que são a casualização, a repetição e o controle local.

Casualização: é a atribuição dos tratamentos às unidades experimentais de forma aleatória, sem nenhuma interferência por parte do pesquisador. Para cada tipo de delineamento experimental existe uma forma adequada de se fazer essa alocação dos tratamentos, que será vista adiante à medida que se for estudando cada um dos desenhos experimentais.

Repetição: refere-se à implantação de cada tratamento em mais de uma unidade experimental. Sua utilização está vinculada à necessidade de avaliação da variabilidade do material experimental.

Controle local: está associado ao conhecimento do ambiente experimental e consiste na divisão das parcelas experimentais em subconjuntos homogêneos, quando é sabido que o total das unidades experimentais não possui a homogeneidade exigida. No geral, o conjunto de parcelas homogêneas constitui o que se convencionou chamar de bloco.

## **Blocos Completos Casualizados**

Este delineamento que é, até hoje, o mais utilizado pelos pesquisadores, não apresenta teoricamente qualquer restrição quanto ao número de repetições ou de tratamentos. De uma maneira geral, o delineamento é recomendado para ensaios com até 25 tratamentos, mas esse número pode ser muito maior ou menor. A restrição no número de tratamentos é em função do tamanho de área contínua que pode ser considerada homogênea, que dependerá do tipo do solo, da vegetação original, do uso e cuidados que ela vem tendo ao longo do tempo, do tipo de cultura a ser usada, do tamanho útil da parcela a ser empregada, da existência ou não de bordaduras, etc.

A forma e a distribuição dos blocos dependerá da área disponível, e das mesmas informações necessárias dadas acima, para a definição do número

de tratamentos (Figura 1). Podem ser deixadas ruas ou espaços livres entre blocos para facilidade de circulação; as parcelas dentro de cada bloco podem ser dispostas das mais diversas formas, buscando-se sempre a melhor homogeneidade da área para cada bloco.

7	3	2	6	4	10	9	8	1	5
9	1	8	3	6	10	7	2	5	4
1	7	3	5	8	4	9	10	2	6

**Fig. 1.** Exemplo de um ensaio em blocos completos casualizados, com dez tratamentos e três repetições.

Comparado com os outros delineamentos até aqui discutidos, apresenta inúmeras vantagens, tais como: simplicidade de planejamento, de instalação e de análise. Permite ainda que seus blocos sejam inclusive instalados em locais não contíguos, permitindo uma maior exploração do ambiente, que terá suas diferenças medidas pelo efeito desses mesmos blocos. Acidentes como a perda de um ou mais tratamentos ou de um bloco não afetam a sua análise de variância, que será conduzida como se o ensaio tivesse sido planejado com um ou mais tratamentos a menos ou com menor número de repetições, no caso de perda de um bloco. A perda de uma ou algumas poucas parcelas é remediada pela estimação matemática de seus possíveis valores, por meio de fórmulas já conhecidas.

Para a análise de experimentos nesse delineamento, sugere-se, entre outros, a leitura do livro "Curso de Estatística Experimental", de Pimentel Gomes, publicado em 1990.

## Testes para Comparação de Médias

A partir do resultado da análise de variância e constatada a existência de efeitos significativos dos tratamentos (cultivares e/ou linhagens) é de interesse do pesquisador verificar onde se encontram efetivamente essas diferenças. Em outras palavras, o pesquisador deve agora descobrir onde estão as diferenças indicadas, por ter sido o teste F significativo, entre as linhagens em teste, e/ou entre as linhagens e as cultivares já em uso pelos agricultores e aqui utilizadas como material testemunha.

Entre os testes para comparação de médias, devem ser destacados o de Dunnett, que é específico para comparação entre materiais em teste e uma testemunha, e os de Tukey, de Duncan, de Waller-Duncan, de Ryan-EinotGabriel-Welsh, além do de Scott-Knott, indicados para comparações gerais entre médias de tratamentos.

O teste de Dunnett apresenta como uma restrição ao seu uso o fato de fazer comparações com uma única testemunha, quando nos trabalhos de melhoramento normalmente são utilizados várias cultivares como testemunhas.

Os demais são indicados nas comparações entre todos os pares de média, não se fazendo distinção entre materiais em teste e as testemunhas. Todos têm vantagens e desvantagens do ponto de vista estatístico, mas o maior defeito dos quatro primeiros, no aspecto prático, é o fato de usualmente não darem uma definição muito clara dos grupos de médias, pelo fato de permitirem a superposição desses grupos, isto é, um determinado tratamento pode pertencer a mais de um agrupamento de médias. O teste de Scott-Knott, ainda pouco conhecido, faz a divisão dos tratamentos em grupos homogêneos e sem interseção, já que se baseia nos mesmos princípios da análise de conglomerados ("cluster").

Como exemplo de aplicação de alguns desses testes em um ensaio comparativo avançado de arroz de terras altas, conduzido em Machadinho, MT, no ano agrícola de 1997/98, no delineamento de blocos completos com quatro repetições, selecionaram-se três cultivares e três linhagens. A análise de variância revelou um CV de 19% e teste F igual a 3,36, significativo a 5% de probabilidade. Os testes de Duncan, Tukey e outros, são facilmente encontrados nos livros textos de Pimentel-Gomes, e Cochran e Cox, "Diseños experimentales", publicado em 1971 e disponibilizados na maioria dos softwares estatísticos, SAS, MSTATC, GENES, SISTAT, SOC, entre outros, o que não ocorre com o Scott-Knott.

Os resultados obtidos com a aplicação dos testes de LSD, Duncan, Tukey e Scott-Knott encontram-se na Tabela 1.

**Tabela 1.** Comparação de médias por meio dos testes LSD, Duncan, Tukey e Scott-Knott, todos a 5% de probabilidade. (DMS = diferença mínima significativa entre 2 médias; e DXS = diferença mínima significativa entre seis médias).

<i>Tratamento</i>	<i>Produção (kg/ha)</i>	<i>LSD</i>	<i>DUNCAN</i>	<i>TUKEY</i>	<i>SCOTT-KNOTT</i>
CNA8548(1)	2991,5	A	A	A	A
XINGU (2)	2499,5	AB	AB	AB	A
PROGRESSO (3)	2258,3	B	B	AB	B
CNA 8674 (4)	2070,5	B	B	AB	B
MARAVILHA (5)	1995,5	B	B	AB	B
CNA 8676 (6)	1937,3	B	B	B	B
DMS	-	655,9	655,9	999,8	-
DXS	-	-	730,2	-	.

De um ponto de vista prático, o teste de Scott-Knott permite claramente a tomada de decisão para a recomendação de cultivares, mostrando que uma das linhagens e uma cultivar, já recomendada, são claramente iguais entre si e superiores aos demais tratamentos. Isso não se verifica nos outros testes aplicados nesse exemplo, e que representa uma situação extremamente comum na prática diária de um pesquisador. O teste de Tukey, por exemplo, mostra que a única diferença significativa é aquela entre as linhagens CNA 8548 e CNA 8676, sendo todas as demais diferenças não significativas, o que torna impossível qualquer decisão a seu respeito.

O teste Scott-Knott encontra-se disponível nos softwares SAEG e Genes, da Universidade Federal de Viçosa, SCOTT, da Embrapa Milho e Sorgo, e SEIVAR, da Universidade Federal de Lavras.