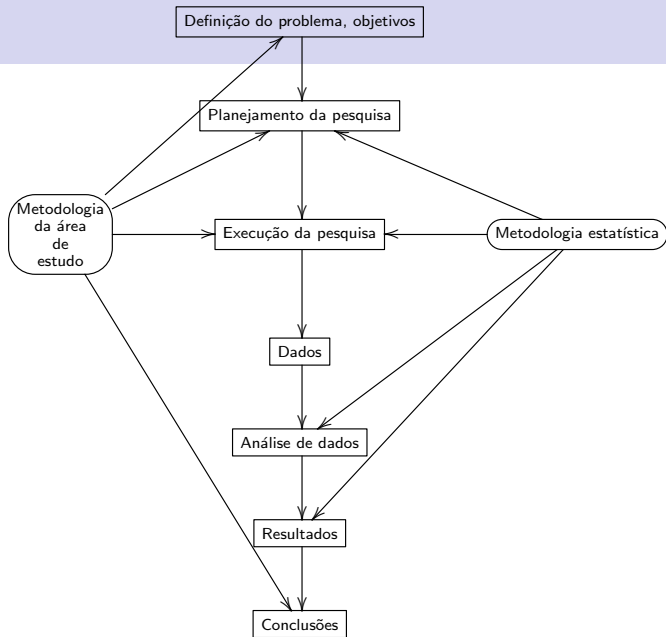


# O QUE É ESTATÍSTICA ?

“As Ciências Estatísticas, ou simplesmente Estatística, é um conjunto de técnicas e métodos de pesquisa que entre outros tópicos envolve o planejamento da pesquisa a ser realizada, a coleta qualificada dos dados, a inferência, a análise e a disseminação das informações”

Kutner et al. (2013) enfatizam que o desenho adequado de um estudo científico é muito mais importante do que as técnicas específicas usadas na análise.



# TIPOS DE PESQUISA: PESQUISA EXPERIMENTAL

- “Em **estudos experimentais** comparativos, a casualização é empregada para alocar um conjunto de tratamentos as unidades experimentais. Os resultados observados entre os grupos de tratamentos são comparados para verificar o efeito dos tratamentos. Os tratamentos são definidos por níveis de um ou mais fatores explicativos, chamados de fatores experimentais. Nestes tipo de estudo, as relações de causa e efeito entre os fatores experimentais e a variável resposta podem ser estabelecidas”.

(Kutner et al. (2013) *Applied linear statistical models*, McGraw-Hill.)

# TIPOS DE PESQUISA: PESQUISA EXPERIMENTAL

**Exemplo:** Um estudo é realizado para comparar três tipos de dieta. Seis indivíduos com características físicas e alimentares semelhantes são alocados aleatoriamente a cada tipo de dieta. Os indivíduos são submetidos as mesmas atividades durante um período de três meses e a perda média de peso em cada grupo é comparada no final dos três meses para verificar se existe uma dieta mais eficiente.

# TIPOS DE PESQUISA: PESQUISA DE LEVANTAMENTO

- “Em **estudos observacionais** comparativos, amostras são obtidas de duas ou mais populações e os resultados observados são comparados entre as populações. As populações são definidas pelos níveis de um ou mais fatores explicativos, denominados de fatores observacionais. Neste caso, se torna difícil estabelecer a relação de causa e efeito entre os fatores explicativos e o resultado ou variável resposta. Geralmente, evidências externas ao estudo observacional são necessárias para estabelecer uma possível explicação de causa e efeito.”

(Kutner et al. (2013) *Applied linear statistical models*, McGraw-Hill.)

# TIPOS DE PESQUISA: PESQUISA DE LEVANTAMENTO

**Exemplo:** Um pesquisador de um centro de reabilitação está interessado em examinar a relação existente entre exercício físico realizado antes de cirurgias de pessoas submetidas a correção no joelho e o tempo necessário de fisioterapia até a completa recuperação do paciente. São coletados dados sobre o número de dias necessário para recuperação e a condição física prévia (superior, regular, inferior) de pacientes submetidos a cirurgia.

# **Delineamento e análise de Experimentos**



# UNIDADE EXPERIMENTAL OU PARCELA

- Segundo Barbin (2003), unidade experimental ou parcela são os indivíduos (plantas, animais, aparelho eletrônico, vaso, etc) aos quais será aplicado um tratamento. Elas devem ser tão homogêneas quanto possível, para que, quando submetidas a diferentes tratamentos, sejam os efeitos facilmente detectados.

# UNIDADE EXPERIMENTAL OU PARCELA

- A unidade experimental ou parcela pode ser composta por apenas 1 indivíduo ou por mais de 1 indivíduo. Exemplos:
  - Em café, usam-se de duas a quatro linhas com 7 a 10 plantas cada uma;
  - Com gado de corte, 1 ou mais animais, de acordo com a disponibilidade de encontrar animais com características iguais para compor a parcela;
  - Com animais de pequeno porte (coelhos, frangos, etc) podem-se usar vários indivíduos para compor a parcela;
  - Com árvores frutíferas, uma a duas plantas são o suficiente, dependendo do grau de homegeneidade dessas árvores.

# QUAL DEVE SER O TAMANHO DA UNIDADE EXPERIMENTAL OU PARCELA?

- Verificar se já foi realizado ensaio com a espécie em estudo;
- Caso sim, verificar se os valores do coeficiente de variação e as conclusões foram satisfatórias;
- Caso contrário, deve-se conversar com o pesquisador para saber sobre a homogeneidade do material em estudo, disponibilidade de material, tamanho da área experimental disponível, número de pessoas que irão conduzir o experimento, etc.
- Barbin (2003) afirma que é sempre preferível aumentar o número de repetições a aumentar o tamanho da parcela.

# PRINCÍPIOS BÁSICOS EM ESTUDOS EXPERIMENTAIS

1. Casualização

2. Repetição

3. Controle Local (Bloco)

# 1. CASUALIZAÇÃO

1. Casualização - é o processo de designar tratamentos às unidades experimentais por processo aleatório. Com isso, todos os tratamentos têm a mesma chance de ocupar determinada unidade experimental.

## 2. REPETIÇÃO

2. Repetição - consiste em se ter várias unidades experimentais com o mesmo tratamento. Dessa forma, é possível confirmar a resposta que a unidade experimental dá a um determinado tratamento.

### 3. CONTROLE LOCAL - BLOCO

3. Controle Local (Bloco) - é usado quando a área experimental é heterogênea. Então, a área experimental é subdividida em áreas menores e homogêneas. Em cada sub-área deve-se colocar todos os tratamentos, de preferência em igual número.

# TIPOS DE DELINEAMENTOS EXPERIMENTAIS

1. Experimentos Inteiramente Casualizados
2. Experimentos em Blocos Casualizados
3. Experimentos em Quadrados Latinos
4. Experimentos Fatoriais
5. Experimentos em Parcelas Subdivididas



# Experimentos Inteiramente Casualizados

# INTRODUÇÃO

- Considere uma área experimental plana, de solo bem homogêneo;
- Em seguida, selecione sementes de um híbrido simples de milho;
- Agora faça a semeadura de maneira que as sementes sejam colocadas no solo, na mesma posição e na mesma profundidade;
- As sementes irão germinar, as plantas crescerão e no momento em que emitirem o pendão, mede-se as alturas das plantas, do solo até a inserção da última folha.

# INTRODUÇÃO

**As plantas terão as mesmas alturas?**

# INTRODUÇÃO

- não!!!
- Dificilmente, terá duas plantas com a mesma altura.
- Então, se tudo que estava ao alcance do pesquisado foi controlado, o que explica as plantas terem alturas diferentes?

# INTRODUÇÃO

- Pode-se admitir que a variação da altura das plantas foi devido a fatores impossíveis de serem controlados, ou seja, foi devida à variação do acaso;
- Sendo assim, o estatístico deve sempre alertar o pesquisador sobre todas as fontes controláveis em seu experimento, evitando que fatores possíveis de serem controlados venham influenciar a variação do acaso.
- Segundo Barbin (2003), talvez esteja aí o sucesso ou fracasso de um experimento.

# INTRODUÇÃO

**Como seria o modelo matemático para representar essa situação?**

# INTRODUÇÃO

Neste caso, o modelo matemático é definido por:

$$y_j = \mu + \varepsilon_j, \quad (1)$$

em que  $y_j$  são os valores observados,  $\mu$  é a média verdadeira,  $\varepsilon_j$  os desvios em relação à média com  $j = 1, 2, \dots, n$ .

# INTRODUÇÃO

- Agora considere que sejam aplicadas 4 doses distintas de nitrogênio no experimento de semente de milho. Sendo elas:
  - 0 N
  - 1 N
  - 2 N
  - 3 N
- As doses serão distribuídas ao acaso nas unidades experimentais;
- A área experimental total será dividida em 12 unidades experimentais;
- O número de repetições de cada dose é 3.



# INTRODUÇÃO

1	2	3
4	5	6
7	8	9
10	11	12

Croqui do ensaio.

# INTRODUÇÃO

A aleatorização das 4 doses de Nitrogênio foi feita do seguinte modo:

- Enumeram-se as parcelas de 1 a 12, como indicado no croqui do ensaio,
- Com o uso de um computador, foi selecionado números aleatórios na faixa de 1 a 12 produzindo a seguinte listagem: {12, 6, 9, 2, 10, 8, 11, 3, 1, 7, 5, 4},
- As parcelas 12, 6 e 9 receberam a dose 3 N; as parcelas 2, 10 e 8 receberam a dose 2 N, as parcelas 11, 3 e 1 receberam a dose 1 N e as parcelas 7, 5 e 4 receberam a dose 0 N. A ordem das doses também foi obtida por sorteio.

# INTRODUÇÃO

Dose 1 N	1	Dose 2 N	2	Dose 1 N	3
Dose 0 N	4	Dose 0 N	5	Dose 3 N	6
Dose 0 N	7	Dose 2 N	8	Dose 3 N	9
Dose 2 N	10	Dose 1 N	11	Dose 3 N	12

Croqui de campo, indicando o processo de aleatorização do experimento.

# INTRODUÇÃO

**Como seria o modelo matemático para representar essa situação?**

# MODELO

- Neste caso, o experimento representa um Experimento Inteiramente Casualizado, em que os Princípios da Repetição e Casualização foram utilizados.
- O modelo matemático para esse experimento é definido por:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (2)$$

em que  $y_{ij}$  são os valores observados (para o exemplo, a altura das plantas),  $\mu_i$  é a média do  $i$  tratamento,  $\varepsilon_{ij}$  componente de erro aleatório com  $j = 1, 2, \dots, n..$

- A componente de erro aleatório incorpora todas as outras fontes de variabilidade do experimento, incluindo medição, variabilidade decorrente de fatores não controlados, diferenças entre as unidades experimentais, etc.

# MODELO

- O modelo (2) é conhecido como modelo de médias e pode ser reescrito por:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, n \quad (3)$$

em que  $\mu$  é a média geral,  $\tau_i$  é a média ou efeito dos tratamentos e  $\varepsilon_{ij}$  componente de erro aleatório.

- Observe que os tratamentos são indexados por  $i$  e os indivíduos por  $j$ .
- O modelo (3) é conhecido como modelo de efeitos.

# MODELO

- Segundo Montgomery (2009), tanto o modelo de médias quanto o modelo de efeitos são modelos estatísticos lineares, ou seja, a variável resposta  $y_{ij}$  é uma função linear dos parâmetros do modelo;
- A notação do modelo de efeitos é mais utilizada em pesquisas experimentais devido ao apelo intuitivo de que  $\mu$  é uma constante e o efeito dos tratamentos  $\tau_i$  representam o desvio da constante quando um tratamento específico é aplicado.
- O modelo definido em (3) também é chamado de análise de variância unidirecional ou de fator único (ANOVA) porque apenas um fator é investigado.

# ANOVA

- Só é possível realizar a análise de variância se certas condições, ou seja, certas exigências do modelo matemático forem satisfeitas:
  - Os erros devem ter distribuição normal;
  - Os erros devem ser independentes;
  - Os erros deve ter a mesma variância, ou seja, deve existir homocedasticidade.
- Pode-se representar essas condições por:

$$\varepsilon_{ij} \sim N(0, \sigma^2) \quad (4)$$



# ANOVA

- Para o modelo de efeitos (ANOVA):

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, n \quad (5)$$

em que  $\mu$  é a média geral,  $\tau_i$  é a média ou efeito dos tratamentos e  $\varepsilon_{ij}$  componente de erro aleatório.

- As hipótese de interesse são definidas por:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0, & (\text{O efeito de tratamento é nulo}) \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

- Ou de forma equivalente:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_a \\ H_1 : \exists \mu_i \neq \mu_j, \quad i \neq j, \end{cases}$$

em que  $\mu_i = \mu + \tau_i$ .

# ANOVA

TABLE: Tabela de análise de variância

Fonte de variação	SQ	g.l.	QM	F
Tratamentos	$SQ_{Trat}$	$a - 1$	$QM_{Trat}$	$QM_{Trat} / QM_{Res}$
Resíduo	$SQ_{Res}$	$an - a$	$QM_{Res}$	
Total	$SQ_T$	$an - 1$		

# ANOVA

- O nome análise de variância é derivado de uma partição da variabilidade total em suas partes componentes. A soma total dos quadrados:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 \quad (6)$$

é usada como uma medida da variabilidade geral dos dados;

- Intuitivamente, isso é razoável porque se dividirmos  $SS_T$  pelo número apropriado de graus de liberdade (neste caso,  $an - 1 = N - 1$ ), teríamos a variância amostral dos  $y$ 's;
- A variância da amostra é, obviamente, uma medida padrão de variabilidade.

# ANOVA

- Observe que a  $SS_T$  pode ser escrita como:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad (7)$$

- A equação (7) é conhecida com a identidade fundamental da ANOVA;
- Ela afirma que a variabilidade total dos dados, medida pela soma total de quadrados, pode ser dividida em uma soma de quadrados das diferenças entre as médias de tratamento e a média geral mais uma soma de quadrados das diferenças das observações com a média do respectivo tratamento;

# ANOVA

- Sendo assim, a equação (7) também pode ser escrita como

$$SS_T = SS_{T_{rat}} + SS_{Res}. \quad (8)$$

- $SS_{T_{rat}}$  - soma de quadrados devido aos tratamentos - representa a diferença entre as médias de cada tratamento e a média geral, ou seja, é uma medida de diferenças entre as médias de tratamento,
- $SS_{Res}$  - soma dos quadrados devido ao erro (ou seja, dentro dos tratamentos) - representa que as diferenças das observações em relação a média do respectivo tratamento pode ser devido à um erro aleatório.

# ANOVA

- Há um total de  $an = N$  observações. Então, a  $SS_T$  tem  $an - 1 = N - 1$  graus de liberdade;
- Existem  $a$  níveis de um fator ( $a$  médias de tratamentos). Então, a  $SS_{Trat}$  tem  $a - 1$  graus de liberdade;
- E existem  $n$  repetições dentro de qualquer tratamento fornecendo  $n - 1$  graus de liberdade para estimar o erro experimental. Como existem  $a$  tratamentos, tem-se  $a(n-1) = an - a = N - a$  graus de liberdade para o erro.