

APRESENTAÇÃO



Olá. É chegado o momento de darmos continuidade em nossos estudos sobre Metodologia da Pesquisa Científica. Através do texto desta unidade entenderemos quais são os tipos de pesquisa e quais são as metodologias aplicáveis para sua realização.





Através deste texto pretendemos levá-lo, também, a refletir sobre como a estatística pode ser usada na pesquisa científica, mais especificamente na pesquisa agropecuária.


Você encontrará, no decorrer deste texto, figuras, esquemas e exercícios que o ajudarão a sedimentar seus conhecimentos sobre a pesquisa científica.

Esta unidade foi escrita pelo professor Daniel Durante Pereira Alves, doutor em Filosofia, membro do Programa de Educação a Distância da Universidade de Uberaba.



Observe abaixo os símbolos que aparecem neste texto e seus significados. Esses serão usados para melhor apresentar a você, conceitos, exemplos, exercícios e observações.

	Exemplos ou Explicações.
	Atenção! Informação Importante.
	Exercícios escritos ou tarefas a cumprir.
	Conferindo suas respostas no texto.
Quadro cinza sem borda	Conceito ou definição relativos ao conteúdo.
Tarja cinza. Ex.: Café	Palavra que se encontra no glossário.
Quadro com borda dupla	Esquemas ou resumos.

 Caso tenha dúvidas em relação ao conteúdo ou sobre qualquer outro procedimento pertinente a esta Unidade de Estudo, não deixe de entrar em contato com nossa equipe de tutores, por meio das CARTAS-DÚVIDA, utilizando o endereço a seguir:

**UNIVERSIDADE DE UBERABA – Campus II
PROGRAMA DE EDUCAÇÃO A DISTÂNCIA**

Sala 2B - 32

**Av. Nenê Sabino, 1.801
38.055-500 – Uberaba, MG**

Telefones:

0xx34-3319-8841

0xx34-3319-8842

E-mail: ead.tutoria@uniube.br

Home Page: www.uniube.br/

Não perca esse endereço! É muito importante que você mantenha contato conosco!



Desejamos a você muito sucesso em seus estudos!

Equipe de Educação a Distância da Universidade de Uberaba

SUMÁRIO

1 - ENTENDENDO AS VÁRIAS POSSIBILIDADES	
DO MÉTODO CIENTÍFICO	5
1.1 - PESQUISA QUALITATIVA E PESQUISA QUANTITATIVA	5
2 - UMA METODOLOGIA QUANTITATIVA DA INVESTIGAÇÃO	
ADEQUADA À PESQUISA AGROPECUÁRIA	9
2.1 - PESQUISA OBSERVACIONAL E PESQUISA EXPERIMENTAL	11
2.2 - PLANEJANDO UMA PESQUISA EXPERIMENTAL	12
2.3 - TÉCNICAS DE AMOSTRAGEM	15
3 - A ESTATÍSTICA COMO FERRAMENTA DA PESQUISA	
QUANTITATIVA	17
3.1 - DISTRIBUIÇÃO ALEATÓRIA	18
3.2 - MEDIDAS DE TENDÊNCIA CENTRAL	19
3.3 - MEDIDAS DE DISPERSÃO	22
3.4 - INTERPRETANDO AS MEDIDAS DE TENDÊNCIA CENTRAL	
E DISPERSÃO	28
3.5 - ESTIMANDO O TAMANHO DAS AMOSTRAS	31
3.6 - INTERPRETANDO RESULTADOS:	
OS TESTES DE HIPÓTESES	34
4 - CONSIDERAÇÕES FINAIS	36
REFERENCIAL DE RESPOSTAS	37
BIBLIOGRAFIA.....	41

1 - ENTENDENDO AS VÁRIAS POSSIBILIDADES DO MÉTODO CIENTÍFICO

Uma vez que já vimos as características mais gerais da ciência, que a distinguem das outras formas de conhecimento, e que já vimos também as características mais gerais das diversas formulações do método científico, cabe agora nos aprofundarmos nas técnicas e métodos específicos ao tipo de pesquisa que nos interessa aqui, a saber, a pesquisa agropecuária em geral e, mais especificamente, a cafeicultura irrigada.



Neste contexto, o aspecto metodológico mais importante para o desenvolvimento da pesquisa corresponde aos métodos e técnicas da estatística, que tanto nos ajudam a projetar um experimento, como também a analisar e interpretar os dados obtidos nesse experimento.

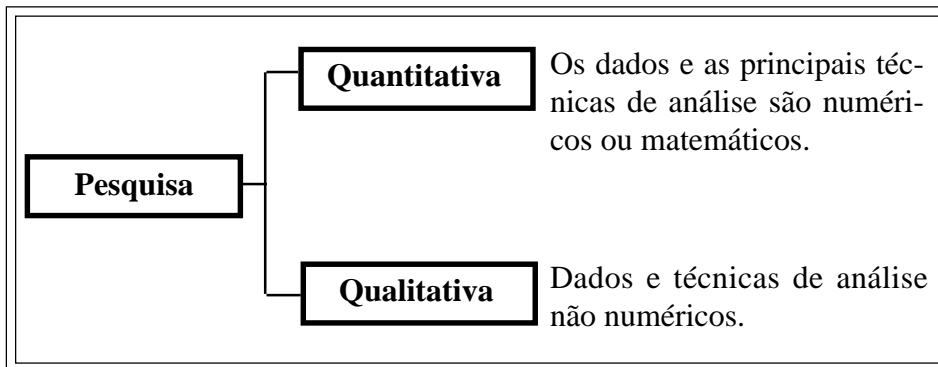
Mas, antes disso, vamos nos deter, ainda que de modo rápido, em uma questão fundamental, qual seja, a das diferentes possibilidades de desenvolvimento do método científico. Na medida em que são diferentes os tipos de problemas e questões para os quais buscamos respostas através da atividade científica, podemos distinguir duas abordagens do método científico.



Mesmo que concordem sobre os aspectos fundamentais do método científico, um antropólogo que deseje investigar, por exemplo, qual a auto-imagem dos menores abandonados, que dormem na Praça da Sé, em São Paulo, usará procedimentos e técnicas de pesquisa diferentes daqueles utilizados por um agrônomo interessado em compreender melhor os efeitos da adição de certos complementos minerais ao solo de uma cultura de café irrigado, no município de Catalão-GO.

1.1 - PESQUISA QUALITATIVA E PESQUISA QUANTITATIVA

Uma distinção fundamental que se faz na metodologia da pesquisa científica diz respeito à utilização da matemática, tanto para a produção quanto para a interpretação dos dados de uma pesquisa. Diz-se que uma pesquisa é *quantitativa*, ou utiliza uma *abordagem quantitativa*, quando a natureza dos dados dessa pesquisa e as principais técnicas de análise desses dados são numéricas ou matemáticas. Quando a natureza, tanto dos dados quanto das técnicas de análise destes não for numérica ou matemática, dizemos que se trata de uma *pesquisa qualitativa*. Observe o esquema a seguir que ilustra esses conceitos.



✓ **Vejamos alguns exemplos...**

Um engenheiro mecânico que pesquisa uma forma de diminuir o consumo de energia de um sistema de irrigação por pivô central, lida com dados tais como velocidade, tempo, pressão, escoamento, aceleração, entre outros. Tais dados são expressos numericamente e exigem técnicas numéricas de análise e interpretação, que caracterizam a sua atividade de pesquisa como *quantitativa*.

Um agrônomo, interessado em comparar a eficiência de dois sistemas distintos de irrigação de café, trabalha com dados tais como: volume de água gasto por hora, taxa de crescimento, produtividade por pé de café, relação custo/benefício, entre outros. Estes, apesar de serem dados menos abstratos que os do exemplo anterior, também são, em sua maioria, expressos numericamente e analisados por métodos matemáticos e estatísticos, caracterizando assim a pesquisa como *quantitativa*.

Um antropólogo, pesquisando a auto-imagem de menores abandonados, muito dificilmente utilizará, como ferramentas principais, dados e métodos numéricos de análise. Como medir numericamente auto-imagem? Muito provavelmente, o antropólogo utilizará em sua pesquisa, categorias da psicanálise, desenvolverá histórias de vida, estudos de caso. Poderá fazer análises comparativas com a auto-imagem de crianças em outras situações. Talvez até utilize, em um argumento ou outro, dados e informações numéricas. Mas, compreender a auto-imagem de um grupo social, exige técnicas e métodos não matemáticos, característicos de uma pesquisa *qualitativa*.



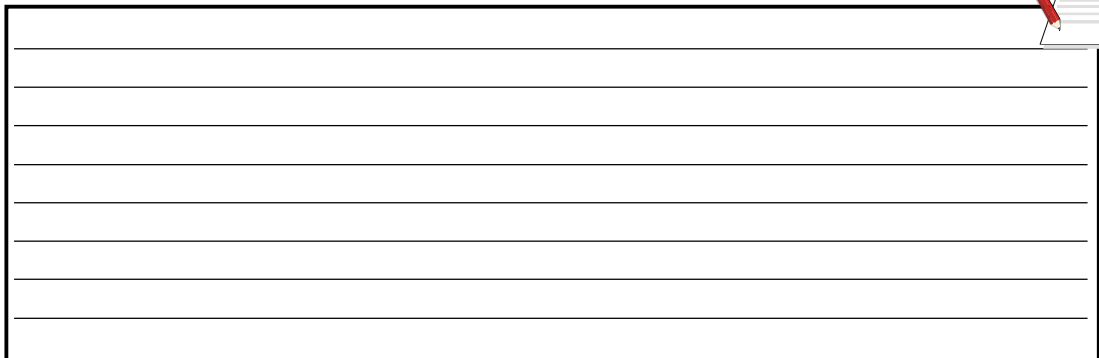
É importante notarmos que o que distingue a pesquisa quantitativa da qualitativa não é a disciplina científica na qual a pesquisa se insere. Mesmo em ciências humanas faz-se muita pesquisa quantitativa.

✓ **Por exemplo...**

Um cientista político pode estar interessado em avaliar, numericamente, em termos de porcentagem de apoio e rejeição, o impacto que a adoção, pelo governo, de certo programa econômico, teria em uma determinada população. Uma pesquisa desse tipo exige **métodos matemáticos e estatísticos**, sendo, portanto, **quantitativa**.

Exercício 1

De acordo com o que vimos até aqui, o que você acha que pode definir se uma pesquisa será quantitativa ou qualitativa?




Será que o conhecimento científico sempre teve a mesma definição?
Vejam o quadro abaixo.

Costumava-se considerar científico apenas o conhecimento obtido através dos métodos quantitativos de pesquisa, por se caracterizarem pela uniformidade e por grande objetividade. Dizia-se, assim, que havia apenas um método para o conhecimento científico, o método quantitativo. Esta concepção tem por base a filosofia positivista de **Auguste Comte**, segundo a qual o método científico está “apoiado na experimentação, mensuração e controle rigoroso dos dados (fatos), tanto nas ciências naturais como nas ciências humanas. [...] Sua objetividade seria garantida pelos instrumentos e técnicas de mensuração e pela **neutralidade** do próprio pesquisador frente à investigação da realidade”(PÁDUA[1998], p.31).

Auguste Comte – Filósofo francês (1798-1857). É o filósofo do positivismo e fundador da Sociologia.

O que importa é termos em mente que os métodos e técnicas de pesquisa quantitativa, que veremos a seguir, não são os métodos e técnicas de toda e qualquer atividade científica. **São apenas os métodos mais adequados ao tipo de pesquisa que, normalmente, se faz na área agrícola.**



Há ciência fora dos procedimentos estatísticos padrões. O fator fundamental, que determina o tipo de metodologia que deve ser adotada em uma pesquisa científica, corresponde à identificação do tipo de problema ou questão que se deseja compreender melhor através da pesquisa. Além desse fator, podemos acrescentar, também, as convicções filosóficas do pesquisador. Este, no entanto, é um fator secundário, uma vez que a questão geradora e as hipóteses de uma pesquisa científica já carregam embutidas as convicções filosóficas de seus proponentes, que são relevantes para a pesquisa.

Exercício 2

De que forma uma pesquisa científica carrega embutidas as convicções filosóficas do pesquisador?

Exercício 3

Descreva, sucintamente, duas pesquisas científicas que possam ser caracterizadas, uma como pesquisa qualitativa e a outra como pesquisa quantitativa.

2 - UMA METODOLOGIA QUANTITATIVA DA INVESTIGAÇÃO ADEQUADA À PESQUISA AGROPECUÁRIA

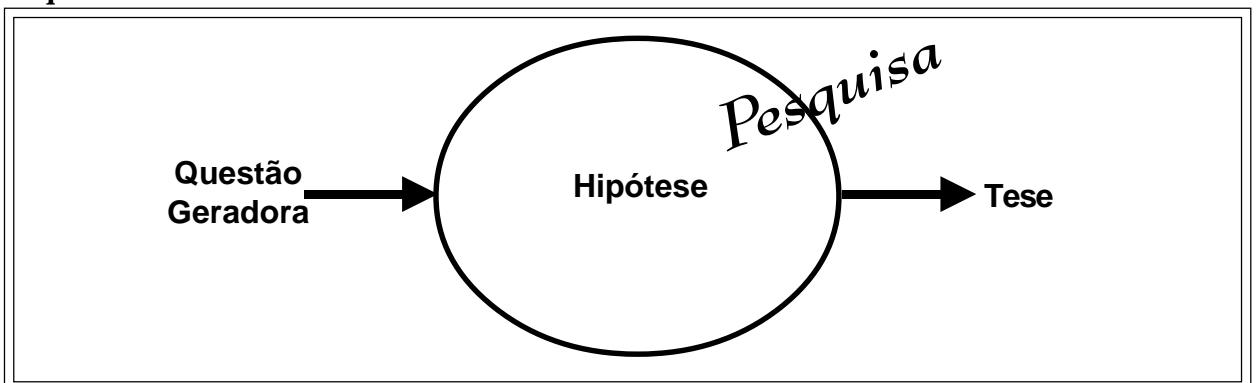
O ponto de partida para a atividade científica é um problema, uma lacuna, uma questão para a qual não temos uma resposta. É a chamada *questão geradora* da pesquisa. Mas a questão geradora não é suficiente. É necessário mais: uma *hipótese*. Uma hipótese é uma resposta provisória, uma estimativa, uma crença do pesquisador sobre qual seria a solução para a questão geradora. Mas, enquanto estimativa e crença, uma hipótese não tem valor científico. Não é um produto da ciência. Uma vez que o pesquisador possui uma hipótese, o próximo passo representa a elaboração de uma pesquisa que possa corroborar ou rejeitar a hipótese. O objetivo de toda a atividade de pesquisa é este: transformar hipóteses em teses. Em outras palavras, tornar crenças incertas, palpites, em resultados mais seguros e confiáveis que possam carregar o status de conhecimento científico, de teses científicas.

Questão geradora – ponto de partida para a pesquisa (problema).

Hipótese – resposta provisória, estimativa, crença do pesquisador sobre a solução da questão geradora.



Esquematizando...



Para a grande maioria das pesquisas em agropecuária, especialmente para as ligadas aos fatores físicos e biológicos dos cultivares, é conveniente adotarmos a postura metodológica defendida por **Karl Popper**. Segundo a sugestão de Popper, uma pesquisa bem projetada é uma tentativa de falseamento da hipótese.

Pesquisar é testar hipóteses. Se uma hipótese resiste a severos testes, severas tentativas de falseamento, então é porque ela é forte o suficiente para tornar-se uma tese e adquirir o status de conhecimento científico. Dessa forma, segundo **(REY [1998], p. 31)**, “o planejamento de uma pesquisa consiste na elaboração de um plano de observação, ou de experimentação, destinado a contestar determinada hipótese, por mais justa e sólida que possa parecer.”

Karl Popper – retome a Unidade 5, do Bloco Temático I, para se recordar do método de Popper.

✓ Considere a seguinte hipótese (1) a seguir:

(1) A adição de 5% de compostos orgânicos à água de irrigação de uma cultura jovem de mudas de café aumenta em 50% a velocidade de crescimento das mudas.



Chamemos de X , a variável que indica a presença de 5% de compostos orgânicos na água de irrigação, e de Y , a variável que mede a taxa de crescimento das mudas. Nossa hipótese, então, representa uma relação causal entre X e Y , que pode ser reescrita como:



(1) na presença de X , Y deve aumentar 50%



Hipótese: $X \Rightarrow Y$ aumenta 50%

X = 5% de compostos orgânicos na água de irrigação
 Y = taxa de crescimento das mudas



Elaborar uma pesquisa que teste a hipótese (1), é elaborar uma experiência que possa verificar que esta relação entre X e Y sempre se dá. Deve-se realizar experimentos onde há ausência de X e onde há a presença de X sob circunstâncias diversas. O pesquisador deve tentar imaginar alguma possibilidade em que esta relação não se daria e verificar, testando-a.

Apenas como notação, as variáveis como X , que precedem a outra e que devem ser manipuladas pelo pesquisador para verificar a hipótese, são chamadas de **variáveis independentes**, e as variáveis como Y , que podem ter seus valores alterados conforme a alteração de X , **variáveis dependentes**.

Variáveis Independentes (X):
variáveis que devem ser manipuladas pelo pesquisador.



Variáveis Dependentes (Y):
variáveis que podem ter seus valores alterados conforme alteração das variáveis independentes.

Exercício 4

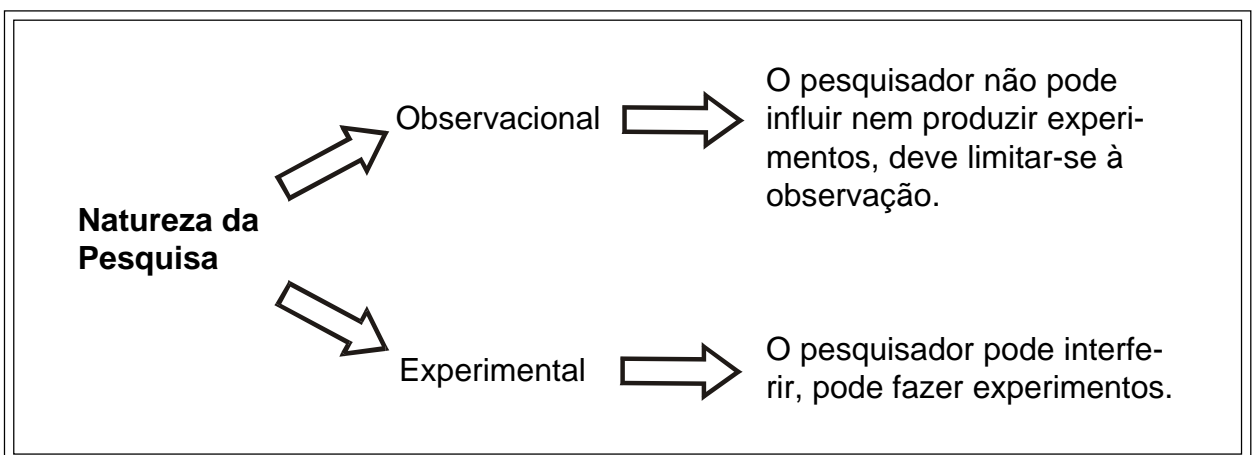
Atribua variáveis às hipóteses abaixo, identificando as variáveis independentes e as dependentes.

- O cruzamento da espécie (A) de café com a espécie (B) produz um café híbrido (AB) com produtividade (grãos por planta) 20% superior à produtividade do café da espécie (A) e 30% superior à do café (B).
- O tratamento de uma lavoura de milho com o composto (C) diminui, em 2 meses de aplicação, a incidência da praga (P) em 85%.
- Para manter a produtividade de lavouras irrigadas do café (A), o método de gotejamento necessita de 40% menos água que o método de pivô central e 70% menos que o método de aspersão.

2.1 - PESQUISA OBSERVACIONAL E PESQUISA EXPERIMENTAL

Podemos caracterizar uma pesquisa, quanto à sua natureza, sob duas categorias: **pesquisa observacional** e **pesquisa experimental**. A primeira é própria de fenômenos em que o pesquisador, devido à natureza do objeto de estudo, não tem condições de influir nem produzir experimentos, mas tem que se limitar a observações. Já na pesquisa experimental, o pesquisador, para verificar sua hipótese, pode interferir nos valores da variável independente.

Resumindo...



Exemplificando...

✓ Se estivermos interessados em verificar como o regime de chuvas afeta a incidência de pragas nas culturas de café de uma certa região, não há nada que possamos fazer a não ser observar. Não temos o poder de manipular a variável independente (quantidade de chuva). Esta será, portanto, uma pesquisa observacional. Já a verificação da aceleração de crescimento de mudas, devido à adição de componentes orgânicos à água de irrigação, pode e deve ser desenvolvida como uma pesquisa experimental, pois é possível manipular a variável independente (presença de compostos orgânicos na água).

2.2 - PLANEJANDO UMA PESQUISA EXPERIMENTAL

Como vimos, o que caracteriza a pesquisa experimental, ou **experimentação**, é a possibilidade que o pesquisador tem de manipular as variáveis independentes.

Assim, a primeira tarefa do pesquisador é descobrir, através de sua hipótese, quais parâmetros devem ser medidos, reescrever sua hipótese em termos de variáveis, e identificar, dentre essas, quais serão submetidas à manipulação, representando as variáveis independentes, e quais serão as variáveis dependentes.



! É fundamental que o projeto da pesquisa permita que se façam comparações, a fim de sabermos como as variáveis dependentes reagiram às manipulações impostas às variáveis independentes. Essas comparações serão cruciais para testarmos nossa hipótese.

A diversidade dos fenômenos naturais e das condições de experimentação exige uma grande variedade de tipos de planejamento.

✓ Apenas a título de ilustração, seguem alguns dos **exemplos** mais comuns:

2.2.1

Pesquisas com um só grupo experimental

Nesse tipo de pesquisa, há um único grupo de sujeitos pesquisados, no qual se fará uma observação ou um conjunto de repetidas observações preliminares. Em seguida, aplica-se um **incremento** à variável independente e faz-se novas observações no grupo para avaliar os efeitos de tal manipulação. **Esse tipo de pesquisa é muito comum quando se deseja avaliar a resposta de determinados indivíduos a alguma droga ou tratamento.**

Incremento: Ato ou efeito de crescer, de aumentar.

Do dicionário Michaelis em <http://www.uol.com.br/michaelis/>

Pesquisas com grupos experimentais independentes

Aqui são dois os grupos pesquisados, independentes, que devem se submeter a condições distintas de alteração da variável independente. Por exemplo, podemos projetar uma pesquisa para testar a **hipótese 1** separando dois grupos experimentais, submetidos às mesmas condições, com a exceção de que, no primeiro, em um dado momento, adicionaremos componentes orgânicos à água da irrigação e, no outro, não. Ao final de um período, comparamos as alturas das mudas dos dois grupos. A hipótese estará confirmada se as mudas do primeiro grupo forem em média 50% maiores que as mudas do segundo. Caso isso não ocorra, a hipótese está refutada.

2.2.2

2.2.3

Pesquisas com tratamentos múltiplos e uma só variável independente

Suponha que queiramos verificar como varia o crescimento de mudas de café irrigado conforme a concentração de componentes orgânicos na água de irrigação. Para isso, podemos separar, por exemplo, 4 grupos pesquisados, sendo que a eles aplicamos, respectivamente, concentrações de 5%, 10%, 15% e 0% de componentes orgânicos na água de irrigação. Fazendo observações preliminares, e, após o período das aplicações, temos meios de avaliar como a concentração de componentes orgânicos influencia no crescimento das mudas. **Esse é um caso de pesquisa com múltiplos tratamentos e uma só variável independente.**

2.2.4

Pesquisas com múltiplas variáveis independentes

Nesse caso, mede-se, em um mesmo experimento, os efeitos de múltiplas variáveis independentes. É claro que, quanto maior o número de variáveis independentes, maior o número de combinações possíveis para os casos a serem analisados. Considere um experimento com duas variáveis independentes A e B , sendo que nos interessam duas condições de análise (valores) para cada uma (A_1, A_2 e B_1, B_2). Temos, dessa forma, quatro possibilidades diferentes para o conjunto dos valores de A e B . São elas:

- ocorrem simultaneamente A_1 e B_1 ;
- ocorrem simultaneamente A_1 e B_2 ;
- ocorrem simultaneamente A_2 e B_1 ;
- ocorrem simultaneamente A_2 e B_2 .

Esse tipo de pesquisa permite a identificação de possíveis interações de fatores diversos em um dado fenômeno. Se, por exemplo, quisermos verificar as interações dos fatores **componentes orgânicos** e **quantidade de água**, sobre o crescimento de mudas de café irrigado, podemos planejar uma pesquisa com as variáveis independentes: **quantidade de água** e **porcentagem de componentes orgânicos na água**, e analisar os resultados que as várias combinações possíveis desses fatores têm no crescimento das mudas.

Exercício 5 

Quais, dentre os quatro tipos de modelagem experimental, apresentados anteriormente, são mais adequados para as seguintes situações:

1. Deseja-se verificar como varia a produtividade de uma lavoura de café de acordo com a quantidade de água disponível para irrigação.
2. Deseja-se investigar os efeitos que a aplicação da droga (X) tem em uma população bovina.
3. Deseja-se verificar qual dos métodos de irrigação (gotejamento ou pivô central) é mais eficiente (necessita de menos água).
4. Deseja-se verificar quais as opções mais produtivas com relação à quantidade de água e porcentagem de componentes orgânicos na água de irrigação de uma lavoura de café.

2.3 - TÉCNICAS DE AMOSTRAGEM

Chamamos de **população** ou **universo**, o **total de elementos de um conjunto**, como, por exemplo, todos os pés de café de uma lavoura. Uma **amostra** representa um **subconjunto qualquer de uma população**.

Veja o exemplo...

✓ 5 pés de café desta lavoura, tomados ao acaso, representam uma amostra. Os 10 pés de café mais altos da lavoura representam outra amostra.

! Quando fazemos pesquisa, em muitos casos não é possível obter dados de toda uma população. Quando isso ocorre, temos que trabalhar com uma amostra.

Escolher uma amostra que seja representativa de toda a população, para o desenvolvimento da pesquisa, é tarefa muito importante. Para tal, existem certas **técnicas de amostragem**, que dependem da natureza da população e dos dados que se quer obter.



✓ **Aqui estão algumas dessas técnicas para você conferir:**

2.3.1 Amostra Casual Simples

Composta por elementos selecionados ao acaso na população, de modo que todo elemento da população tem igual probabilidade de compor a amostra. É recomendada quando a população a ser estudada não se encontra organizada sob nenhum aspecto.

2.3.2 Amostra Sistemática

Composta por elementos selecionados de acordo com algum sistema. Por exemplo, um sistema pode ser o seguinte: em uma lavoura de milho, toma-se para amostra todos os pés das filas pares cuja ordem, da esquerda para a direita, é um número múltiplo de 3. Esquemáticamente, teríamos:

fila 1: 01 02 03 04 05 06 07 08 09 10 11 12 13
 fila 2: 01 02 03 04 05 06 07 08 09 10 11 12 13
 fila 3: 01 02 03 04 05 06 07 08 09 10 11 12 13
 fila 4: 01 02 03 04 05 06 07 08 09 10 11 12 13
 fila 5: 01 02 03 04 05 06 07 08 09 10 11 12 13

Essa técnica de amostragem é recomendada quando a população a ser estudada encontra-se organizada de alguma forma.



2.3.3 Amostra Estratificada

Composta por elementos pertencentes a todos os estratos da população. Os estratos são subgrupos da população, reunidos de acordo com certos valores de uma variável. Por exemplo, podemos estratificar uma população de pés de milho de acordo com o número de espigas que eles apresentam. Assim, podemos considerar que os pés que têm de 0 a 2 espigas estão no estrato 1, os que têm de 3 a 4 estão no estrato 2 e assim por diante. Uma amostra estratificada dessa população deve conter pés de milho de todos os estratos. Essa técnica é recomendada quando a característica que define os extratos é importante em nossa pesquisa.

2.3.4 Amostra por Conveniência

Composta por elementos selecionados de acordo com critérios não objetivos de conveniência. Suponha que se queira pesquisar a população de pássaros que está atacando uma determinada lavoura. Prepara-se, então, algumas armadilhas ao longo da lavoura. Os pássaros capturados representam uma amostra da população que ataca a lavoura, composta por conveniência, ou seja, pelo fato de terem sido capturados. Esse tipo de amostra é bastante criticado pelos estatísticos, pelo risco de não ser representativa de toda a população, devendo ser evitada sempre que possível. No entanto, em casos como o do exemplo, não há escolha.

3 - A ESTATÍSTICA COMO FERRAMENTA DA PESQUISA QUANTITATIVA

Os métodos e técnicas estatísticos ocupam importante papel na atividade científica em geral, sendo fundamentais para a pesquisa quantitativa.



É, de fato, a utilização desses métodos e técnicas que caracteriza uma atividade de pesquisa como quantitativa.

Segundo (REY [1998], p. 45), “(...) a aplicação da estatística ao tratamento dos resultados de um experimento ou de uma observação científica compreende dois aspectos. O primeiro, é a apresentação dos dados de forma sistematizada, clara e sintética, a fim de que se possa entender ou visualizar o comportamento das variáveis observadas, e, eventualmente, alguma relação que haja entre elas. Este é o aspecto descritivo da estatística (...). O segundo, é aquele que permite analisar os resultados e fazer inferências, para a tomada de decisões quanto à aceitação ou rejeição das hipóteses formuladas: essa é a função da inferência estatística”.

A estatística é útil tanto para apresentar os dados de uma pesquisa quanto para ajudar na decisão sobre a validade de uma hipótese. A estas duas funções podemos acrescentar também a importância da estatística para o projeto dos experimentos de uma pesquisa.



Nosso objetivo aqui é apresentar uma breve descrição de certas técnicas estatísticas que auxiliam a atividade do pesquisador, tendo sempre como base a pesquisa agropecuária. Não temos a pretensão de fazer uma exposição completa, a qual, como dissemos, tem apenas caráter introdutório.

Três textos foram muito utilizados durante a produção desse material e devem ser consultados por você em caso de necessidade de aprofundamento do que aqui expomos.

São eles: GOMES[1987], REY[1998] e VIEIRA[1980].

3.1 - DISTRIBUIÇÃO ALEATÓRIA

Em uma lavoura de café, na qual todos os pés são da mesma espécie, nascidos de sementes idênticas, plantados em solo com as mesmas características e sujeitos às mesmas condições climáticas, poderíamos logicamente supor que todos esses pés de café deveriam ter um desenvolvimento idêntico, possuindo todos, por exemplo, a mesma altura. No entanto, sabemos que isso não ocorre.

As variações, em relação à altura, entre os pés de café dessa lavoura, são devidas a um enorme conjunto de fatores que não podemos controlar nem ao menos perceber.



Esse tipo de **variação, que não possui uma causa definida**, é chamado pelos estatísticos de **variação aleatória** e possui propriedades específicas bastante estudadas, que são de muita utilidade na atividade do pesquisador. Por mais que a ciência avance na compreensão dos fenômenos, suas causas e relações, sempre haverá espaço para a aleatoriedade, e praticamente todos os tipos de medidas que os cientistas realizam estão sujeitas à variação aleatória.



No caso das pesquisas agropecuárias, onde estão envolvidos fatores geográficos, biológicos, físicos, químicos, e muitos outros, a variação aleatória é sempre significativa e deve ser considerada com cuidado.

Quando, para testar a **hipótese 1**, por exemplo, utilizamos dois grupos experimentais e dizemos que o valor da variável Y (altura das mudas de café) no grupo A é, por exemplo, 40 cm e no grupo B é 30 cm, isso significa que estes números 40 e 30 foram obtidos através de medidas em muitos pés de café do grupo. São números que representam uma tendência de cada grupo àquela altura média. As alturas reais de cada uma das mudas do grupo variam aleatoriamente em torno desses números. Além disso, junto com os números 40 e 30, precisamos informar quão longe ou perto deles estão as alturas reais dos pés. Qual o desvio médio dos pés de café com relação a estes números, ou seja, quão homogêneas são as alturas das plantas de cada grupo.

A **estatística descritiva** aborda aspectos da pesquisa sobre os quais estaremos estudando a partir deste momento. Através destes, aprenderemos como:

- Atribuir valores a variáveis;
- Estimar a confiança e a representatividade dos dados obtidos;
- Definir quão grande deve ser uma amostra, a fim de que ela seja suficientemente representativa de uma população.

3.2 - MEDIDAS DE TENDÊNCIA CENTRAL

Uma **medida de tendência central** representa, como o nome sugere, **um valor em torno do qual os dados de um conjunto se distribuem**. A mais comum delas é a **média aritmética**. Também importantes são a **mediana** e a **moda**.



Quando, em uma pesquisa, utilizamos uma **amostra** (ou **toda a população**) para **atribuir valor a uma variável**, tal como a altura das mudas de café de uma lavoura, o **valor** que a variável deve assumir **corresponde a uma medida de tendência central dos valores dessa amostra**. Em geral utilizamos a média aritmética.

3.2.1. Média Aritmética

A **média aritmética** ou, simplesmente **média**, representa a **soma de todos os valores de um conjunto numérico dividido pelo total de elementos do conjunto**.



Considere por exemplo uma amostra casual de 6 pés de café de uma lavoura, com as seguintes alturas em centímetros: 176, 214, 267, 198, 235, 212. A média aritmética desta amostra é:

$$m = \frac{176 + 214 + 267 + 198 + 235 + 212}{6} = 217$$



Utilizaremos a letra maiúscula M para representar a média de toda uma população N . A letra minúscula m , em itálico, indicará a média de uma amostra extraída da população.

A média aritmética é então obtida através da seguinte fórmula:

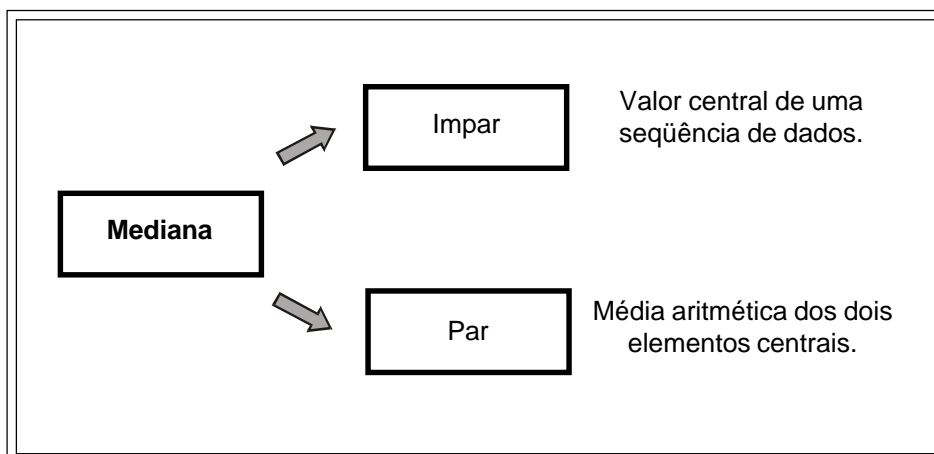
$$M = \frac{\sum X}{N},$$

onde: $\sum X = X_1 + X_2 + \dots + X_N$

3.2.2. Mediana

A **mediana** representa o **valor central de uma seqüência de dados**, quando esta seqüência tem uma **quantidade ímpar de elementos**. **Caso a seqüência tenha uma quantidade par de elementos, a mediana representa a média aritmética dos dois elementos centrais**.

Indicaremos a mediana por Me .



✓ **Exemplificando...**

Considere, por exemplo, as duas seqüências seguintes:
 X: 123, 214, 175, 91, 188 e
 Y: 214, 123, 214, 174, 91, 188

Colocando-as em ordem crescente temos:
 X: 91, 123, 175, 188, 214 e
 Y: 91, 123, 174, 188, 214, 214

5 Elementos				
91	123	175	188	214

↑

6 Elementos					
91	123	174	188	214	214

↑ ↑

⇒ **Mediana Ímpar**
175

⇒ **Mediana Par**
 $\frac{174 + 188}{2} = 181$
média aritmética dos dois valores centrais

3.2.3. Moda

A *moda* é o valor que ocorre com maior freqüência em uma seqüência de números. Indicaremos a moda por *Mo*.

Exemplificando...

Na seqüência 3, 4, 4, 6, 6, 6, 8, 9, 9, **a moda é 6.**

No entanto, na seqüência X, do exemplo anterior sobre Mediana: 91, 123, 175, 188, 214 **não existe Moda**, pois nenhum valor se repete.

Já na seqüência 3, 4, 4, 6, 6, 6, 8, 9, 9, 9, tem **duas modas, 6 e 9.**

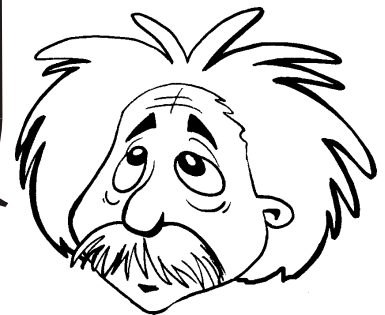
Exercício 6 

Considere a tabela, a seguir, como correspondente a uma pequena lavoura de café, onde os números representam as alturas dos pés em centímetros. Calcule a média aritmética, a mediana e a moda das duas amostras sistemáticas, uma delas constituída pelos números marcados com um círculo e a outra pelos números marcados com um hexágono em tom cinza.

098	123	112	136	087	099	107	124	085	102	114	097
096	114	135	128	109	097	136	127	110	094	102	107
085	103	118	129	127	136	142	126	112	135	107	103
101	100	145	127	114	104	116	122	098	099	103	111
081	096	092	113	121	132	111	127	097	100	120	098
079	087	101	099	103	110	125	111	103	098	101	085

Além de informarem os pontos em torno dos quais os valores de uma amostra se distribuem, as **medidas de tendência central** nos dão uma outra informação bastante importante. Quando tomamos uma amostra razoavelmente grande, a estatística nos diz que, em uma **distribuição aleatória**, os valores da média aritmética M , mediana Me e moda Mo devem coincidir. Ou seja, devemos ter $M = Me = Mo$. Não se trata de obter uma igualdade absoluta, mas a diferença entre eles não pode ser significativa.

Não deixe de ver o conceito de **Medidas de Tendência Central**, no início da seção 3.





Se os valores das médias M , Me e Mo diferirem significativamente, isso quer dizer que a **distribuição dos valores da amostra não é aleatória**, devendo haver alguma causa identificável que provoca essa distorção.

Refleta sobre o conceito de distribuição aleatória, que se encontra no item 3.1.



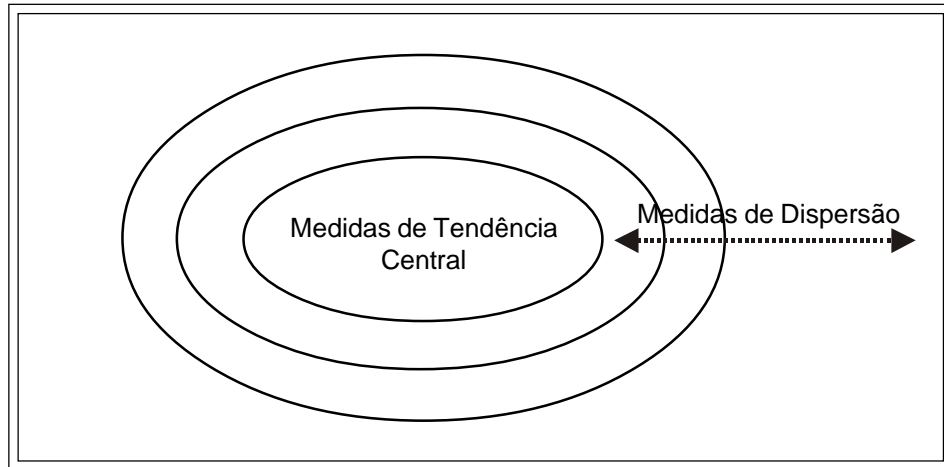
Exercício 7

Qual a importância de se descobrir se houve aleatoriedade de uma distribuição em uma pesquisa quantitativa?

3.3 - MEDIDAS DE DISPERSÃO

Apesar de as **medidas de tendência central** corresponderem a valores em torno dos quais os demais se distribuem, elas não têm poder de expressar quão agrupados ou dispersos em torno deste valor central os demais estão. Olhando apenas para a média aritmética das alturas de pés de café de uma lavoura, não temos condição de **saber se a maioria dos pés de café tem altura próxima à média ou se a altura dos pés da lavoura varia muito**. Este outro tipo de informação é obtido através das **medidas de dispersão**, sendo o **desvio-padrão**, a **variância** e o **coeficiente de variabilidade** as mais utilizadas.

Esquematizando, temos...



Assim, quando, em uma **pesquisa**, utilizamos a **média para atribuir valor a uma variável**, temos também que informar qual a **dispersão dos valores medidos em torno desta média**. Essa informação, veremos mais adiante, será útil para calcularmos a **confiança e representatividade dos dados da pesquisa**. A **variância**, o **desvio padrão** e o **coeficiente de variabilidade** são as ferramentas estatísticas de que dispomos para calcularmos essa dispersão.

3.3.1. Variância, Desvio-Padrão e Coeficiente de Variabilidade

Como os **dados de um conjunto se distribuem em torno da média**, nada mais natural do que medir o **grau de dispersão de um conjunto de dados através do desvio destes em relação à média**, ou seja, através da **diferença entre estes dados e a média**. A **variância** nada mais é do que a média dos quadrados dos desvios individuais, e o **desvio padrão** corresponde à raiz quadrada da variância.



Variância = média dos quadrados dos desvios individuais.
Desvio padrão = raiz quadrada da variância.

Considerando X uma seqüência de N números temos que a variância, que indicaremos por S^2 , é definida pela seguinte fórmula:

$$S^2 = \frac{(X_1 - M)^2 + (X_2 - M)^2 + \dots + (X_N - M)^2}{N}$$

Ou, mais resumidamente:

$$S^2 = \frac{\sum_{i=1}^N (X_i - M)^2}{N}$$

Logo, o desvio padrão é dado por:

$$S = \sqrt{S^2}$$

✓ **Vejamos um exemplo...**

Considere o seguinte conjunto de dados:

X : 123, 214, 175, 91, 188

Note que: $M = \frac{123 + 214 + 91 + 175 + 188}{5} \Rightarrow M = 158,2$

Assim, os desvios de cada um dos valores correspondem a:

$$\begin{aligned} M - 123 &= 158,2 - 123 = 35,2 \\ M - 214 &= 158,2 - 214 = -55,8 \\ M - 91 &= 158,2 - 91 = 67,2 \\ M - 175 &= 158,2 - 175 = -16,8 \\ M - 188 &= 158,2 - 188 = -29,8 \end{aligned}$$

Portanto, a variância S^2 é dada por:

$$\begin{aligned} S^2 &= \frac{35,2^2 + (-55,8)^2 + 67,2^2 + (-16,8)^2 + (-29,8)^2}{5} \\ S^2 &= \frac{1239,04 + 3113,64 + 4515,84 + 282,24 + 888,04}{5} \\ S^2 &= 2007,76 \end{aligned}$$

O desvio padrão é, portanto, a raiz quadrada de 2007,76, ou seja:

$$S = \sqrt{2007,76} = 44,80$$

Note, pela fórmula que define o desvio padrão, que este representa uma grandeza expressa na mesma unidade dos valores do conjunto de dados. Para que possamos comparar o grau de dispersão de diferentes conjuntos de dados, lançaremos mão do *coeficiente de variabilidade (CV)*, que será definido como uma grandeza absoluta, ou seja, independente da unidade.

Definimos o **coeficiente de variabilidade CV** como o **valor percentual** que o **desvio padrão S** representa em relação à **média M**.



$$CV = \frac{100 \times S}{M}$$

✓ **Vejamos um exemplo da importância do coeficiente de variabilidade:**

Considere os **dois seguintes grupos** de dados que representam a idade, em anos, de 6 indivíduos: 3, 1, 5 e 53, 51, 55

Acontece que as diferenças de idade, no primeiro grupo, são muito mais significativas do que no segundo, pois representam, nos indivíduos do primeiro grupo, um percentual muito maior em relação às suas idades do que no segundo. É claro que a diferença de idade, entre um bebê de 1 ano e uma criança de 5, é muito mais significativa do que a diferença de idade entre dois adultos de 51 e 55 anos. É esse tipo de informação que o coeficiente de variabilidade capta. Veja:

No primeiro grupo, a **média de idade** é de **3 anos**, e no segundo, de **53**.
Os dois grupos têm a mesma dispersão de dados, com **desvio padrão S = 2**.



O coeficiente de variabilidade dos dados do primeiro grupo, é muito maior do que no segundo, apesar de os dois grupos possuírem desvios-padrão com valores idênticos e expressos na mesma unidade (anos).



No primeiro grupo temos:

$$CV = \frac{100 \times 2}{3} = 66,67\%$$

Já no segundo grupo:

$$CV = \frac{100 \times 2}{53} = 3,77\%$$

Exercício 8

Elabore um resumo ou um esquema do que você aprendeu sobre Medidas de Tendência Central e Medidas de Dispersão, enfatizando a importância do uso das mesmas.



3.3.2 - MEDIDAS DE DISPERSÃO SOBRE AMOSTRAS E GRAUS DE LIBERDADE



Note que, para os cálculos do desvio-padrão, da variância e do coeficiente de variabilidade (respectivamente S , S^2 e CV), apresentados acima, utilizamos a média absoluta M , que, como definimos em 3.2.1, é calculada a partir de todos os indivíduos de uma população.

Acontece que, na prática das pesquisas, raramente se tem o valor da **média** M . Normalmente se trabalha com a **média amostral** m , obtida de uma amostra que tomamos como representativa da população. Neste caso, os cálculos da variância e do desvio padrão se alteram um pouco, em função do que chamaremos de *graus de liberdade*, para refletir a diminuição da confiabilidade que o trabalho com a amostra representa. Estes novos desvios-padrão e variância calculados através de amostras são indicados por s e s^2 (letras minúsculas).

Assim, considerando uma amostra X com N elementos e média m , temos que as novas fórmulas para **variância** e **desvio-padrão**, são:



$$s^2 = \frac{\sum (X_i - m)^2}{N - 1}$$

$$s = \sqrt{s^2}$$

Note que, além de utilizarmos a média amostral m no lugar de M , no denominador da expressão de s^2 , diferentemente do cálculo de S^2 , fazemos a subtração de uma unidade do tamanho da amostra.

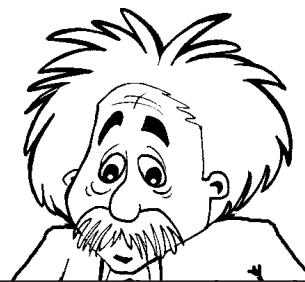
Sabemos que a diminuição de uma unidade, no denominador de uma fração, produz um aumento no valor da fração. Isso significa que, quando calculados com base em dados de uma amostra, os valores do desvio padrão e da variância tendem a ser ligeiramente maiores do que quando calculados com base nos dados de toda a população. Quanto maior for uma amostra (N), menos significativa será a diferença de N para $N-1$ e, portanto, os valores da variância e desvio-padrão amostrais (s e s^2) estarão mais próximos dos valores reais (S e S^2).

O conceito de *graus de liberdade* diz respeito a esta “distância” entre o tamanho da amostra e o da população. Dizemos que um sistema baseado em uma amostra com N elementos, possui $(N-1)$ *graus de liberdade*. Em relação a um sistema que não é baseado em amostra, ou seja, cujos cálculos são feitos sobre todos os elementos da população, dizemos que há infinitos graus de liberdade.

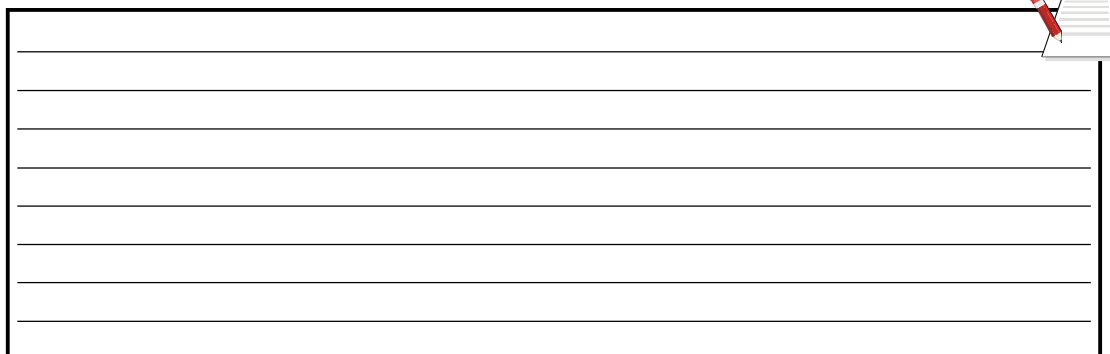
Exercício 9

Considere a mesma tabela do exercício anterior, que representa uma pequena lavoura de café, onde os números correspondem às alturas dos pés em centímetros. Considerando as médias das amostras indicadas por círculos e as amostras indicadas por hexágonos, calculadas anteriormente, calcule os valores do desvio-padrão, da variância e do coeficiente de variabilidade para cada uma das amostras.

Observe a tabela abaixo e utilize o espaço da próxima página, para desenvolver este exercício.



098	123	112	136	087	099	107	124	085	102	114	097
096	114	135	128	109	097	136	127	110	094	102	107
085	103	118	129	127	136	142	126	112	135	107	103
101	100	145	127	114	104	116	122	098	099	103	111
081	096	092	113	121	132	111	127	097	100	120	098
079	087	101	099	103	110	125	111	103	098	101	085



3.4 - INTERPRETANDO AS MEDIDAS DE TENDÊNCIA CENTRAL E DISPERSÃO

Veremos, agora, como as medidas de tendência central e de dispersão são úteis para a atividade de pesquisa, informando sobre a **confiança e representatividade dos dados**. Novamente alertamos que não entraremos em maiores detalhes estatísticos.

Para os que desejem se aprofundar no assunto, além da literatura já recomendada, sugerimos **LITTLE & HILLS[1976]**.



✓ Considere uma amostra de 11 plantas escolhidas de uma lavoura de café. Seja 176cm a altura média de pés de café da amostra, com desvio-padrão de 21,3cm. Assim, temos: $m=176\text{cm}$, $s= 21,3$ e $N=11$. Este exemplo será usado, neste tópico, para nos auxiliar no entendimento da importância das medidas de tendência central e dispersão.



Considerando estes dados, como responder às seguintes perguntas:
Dentro de que limites variam as alturas dos pés de café da lavoura, em sua maioria?
Qual a probabilidade de haver nesta lavoura pés de café com alturas superiores a 230cm? Ou inferiores a 86cm?

As respostas a estas perguntas são obtidas estatisticamente, utilizando-se as propriedades matemáticas da famosa *curva de distribuição normal*, ou *curva de Gauss*.

A **Curva de Gauss** corresponde a uma função matemática que descreve, em termos ideais, o comportamento dos fenômenos com distribuição aleatória.

A **primeira pergunta**, feita anteriormente, está procurando por um intervalo de confiança. Podemos refazê-la, mais especificamente, da seguinte maneira: **Quais seriam os limites de altura de 95% dos pés de café da lavoura?**



A resposta a esta pergunta corresponde ao que chamamos de *intervalo de confiança de 95%* de uma amostra. Ela é obtida através de uma equação simples, envolvendo a média, o desvio-padrão e os graus de liberdade. Chamemos de L_i o limite inferior e L_s o limite superior.

Assim Temos:

$$\begin{aligned} L_i &= m - (t_0 \times s) \\ L_s &= m + (t_0 \times s) \end{aligned}$$



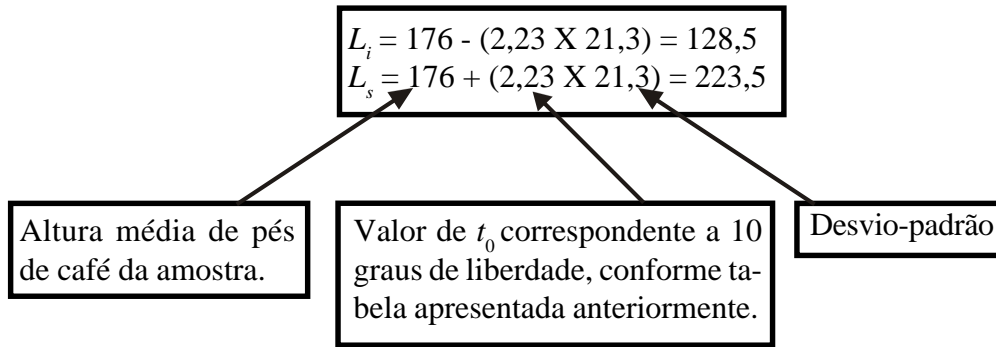
Onde t_0 é obtido através da consulta a uma tabela que depende do número de graus de liberdade do sistema, do grau de confiança desejado e está presente na maioria dos livros de estatística.

N.º de graus de liberdade	Valores de t_0
2	4,30
4	2,78
10	2,23
15	2,13
20	2,09
30	2,04
40	2,02
60	2,00
120	1,98
∞	1,96

Esta é uma versão resumida da tabela, retirada de (GOMES[1987], p. 19).

A tabela anterior vale apenas para um **grau de confiança** de 95%, ou seja, quando queremos calcular os limites que englobam 95% dos indivíduos de uma amostra. Para outros graus de confiança, deve-se consultar outras tabelas.

Como, em nosso exemplo $N=11$, temos 10 graus de liberdade e, portanto, $t_0=2,23$ e os valores para L_i e L_s são:



Dessa forma, respondendo à nossa primeira pergunta, diríamos:

95% dos pés de café da lavoura têm altura superior a 128,5cm e inferior a 223,5cm. Ou, dito de outra forma, a **probabilidade** de um **pé de café** desta lavoura ter **altura entre 128,5cm e 223,5cm é de 95%.**

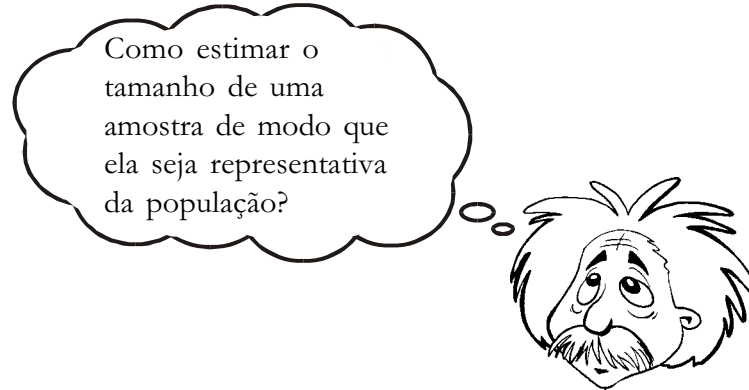
Esta informação já nos diz algo sobre a **segunda pergunta** (as probabilidades de haver pés de café com alturas superiores a 230cm ou inferiores a 86cm). Como estes valores estão além dos limites do intervalo de confiança de 95% ($86 < 128,5$ e $230 > 223,5$), certamente há menos de 5% de pés de café nestas condições, ou seja, há menos de 2,5 % de probabilidade de haver pés de café menores que 86cm e menos de 2,5% de probabilidade de haver pés de café maiores que 230cm.

Exercício 10

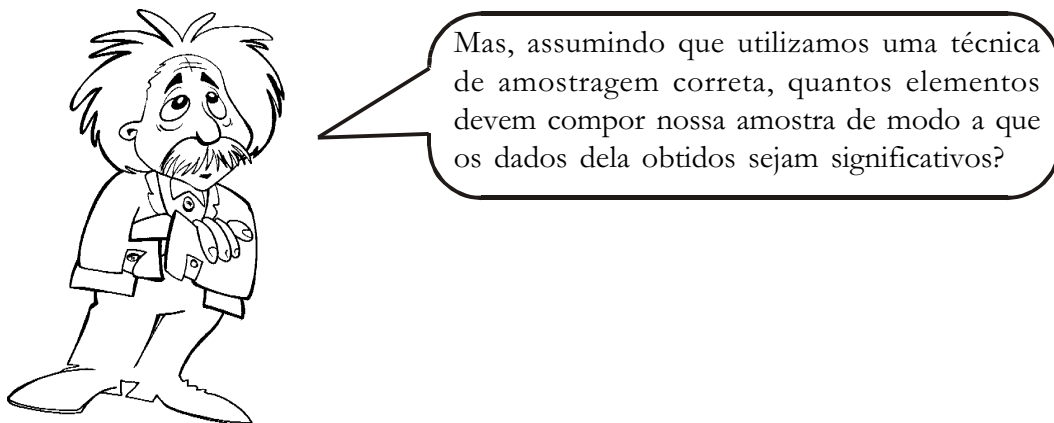
Considere novamente a tabela dos exercícios anteriores, que traz as alturas de uma pequena lavoura de café (tabela da página 27). Arredondando o tamanho das amostras indicadas por círculos e hexágonos para 10 graus de liberdade, obtenha o valor de t_0 da tabela da página anterior e calcule os limites inferior e superior do intervalo de confiança de 15% para essas duas amostras. Qual o significado desse intervalo de confiança?

3.5 - ESTIMANDO O TAMANHO DAS AMOSTRAS

Quando trabalhamos com amostras, como quase sempre ocorre na pesquisa agropecuária, uma questão importante que surge é:



As técnicas de amostragem que vimos em 2.3, são úteis para nos ajudar a produzir amostras não viciadas, cujos elementos sejam realmente obtidos da população aleatoriamente.



A resposta a esta pergunta depende de **quão próxima da média verdadeira** queremos que a nossa **média amostral** esteja.

Antes de obtê-la, vejamos como calcular o **desvio padrão da média amostral**:

Suponha uma lavoura de café da qual tomamos 4 amostras, com 11 elementos cada, e calculamos a média das alturas dos pés de cada uma delas. Podemos agora calcular o desvio-padrão dessas médias. ✓

Há, no entanto, uma maneira direta de calcular o *desvio padrão da média*, o qual chamaremos de S_m , baseada nos dados de apenas uma amostra.

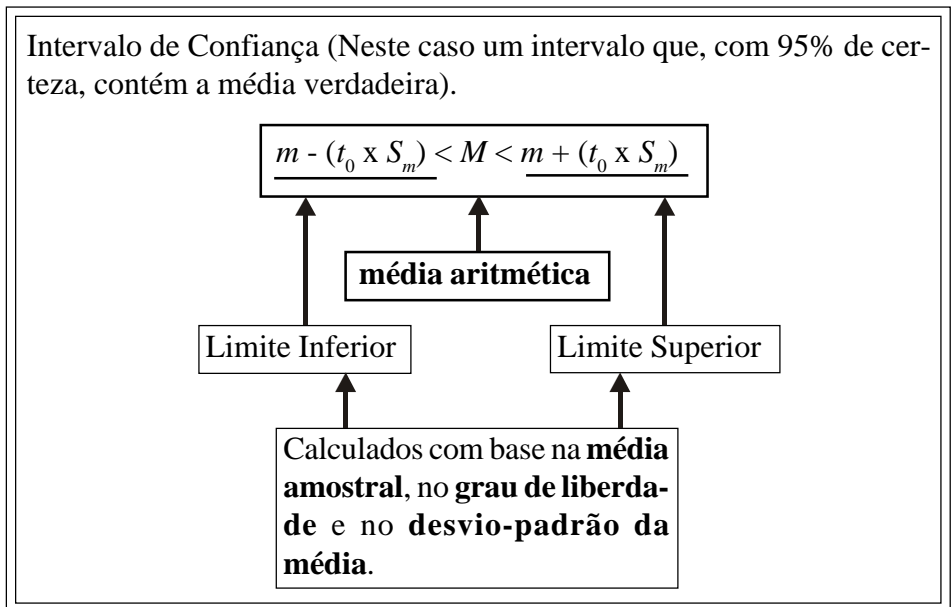
Seja N o número de elementos da amostra e s seu desvio-padrão. O desvio-padrão da média é dado por:

3.5.1. $S_m = \frac{s}{\sqrt{N}}$

Com esse desvio-padrão saberemos calcular o intervalo de, por exemplo, 95% de confiabilidade para a média, ou seja, um intervalo que, com 95% de certeza, conterá a média verdadeira. Assim, de acordo com o que vimos no item 3.5, em 95% dos casos:

$$m - (t_0 \times S_m) < M < m + (t_0 \times S_m)$$

Relembrando o que foi visto no item 3.4, entenderemos melhor:



Note que $(t_0 \times S_m)$ corresponde ao incremento/decremento que, quando aplicado a m , define o intervalo de confiança da média. Este incremento/decremento é chamado de *erro padrão* e denotado por D .

Temos: $D = (t_0 \times S_m) = \left(t_0 \times \frac{s}{\sqrt{N}} \right)$

Podemos agora manipular algebricamente esta equação de modo a isolar N .

Temos: **3.5.2.** $N = \left(t_0 \times \frac{s}{D} \right)^2$

Um procedimento, muito utilizado para **estimar o tamanho que uma amostra** deve ter, consiste em trabalharmos, preliminarmente, com uma **amostra piloto**, para que possamos obter os **valores de s e t_0** da equação 3.6.2.



Mas, para que possamos **calcular N** , resta saber o valor de **D** . Como dissemos, o **erro padrão D** representa o **incremento/decremento** que, **aplicado à média amostral m** , produz um **intervalo que com 95% de certeza** conterá a **média verdadeira**.


O pesquisador determina, então, **que tipo de erro padrão** seria **aceitável** em sua pesquisa, **definindo assim o valor de D** . Com isso, ele obtém todas as **informações necessárias** para **calcular, através da equação 3.6.2, o tamanho da amostra** mais adequado à sua pesquisa.

Relembrando temos:

“ s ” é o desvio padrão e “ t_0 ” é o valor da tabela 3.5 para 10 graus de liberdade e 95% de confiança.

Exercício 11

Considerando $D=5\text{cm}$, arredondando t_0 para 2,23 (dez graus de liberdade em um intervalo de confiança de 95%) e tomando o desvio padrão s obtido da amostra indicada por hexágonos de nossa tabela de alturas de pés de café (página 27), calcule uma estimativa para o tamanho N de uma amostra que garanta, com 95% de certeza, que a média verdadeira M da população esteja compreendida no intervalo $(m-5, m+5)$, onde m representa a média das alturas desta amostra.



3.6-INTERPRETANDO RESULTADOS: OS TESTES DE HIPÓTESES

Vamos agora dar uma rápida olhada nas possibilidades de utilização da estatística como ferramenta para interpretar e tomar decisões com relação aos dados de uma pesquisa.

✓ **Vejamos um exemplo:**

Há 3 meses atrás verificamos que 20% dos pés de uma lavoura de café estavam infectados com uma doença A. Hoje, suspeitamos que este número aumentou e queremos verificar.

São duas as hipóteses a verificar:

- H_0 : a proporção de pés de café doentes continua 20%.
- H_1 : a proporção de pés de café doentes é maior do que 20%.

A primeira hipótese é conhecida como *nulidade*, por razões óbvias, enquanto que a segunda é chamada de *hipótese alternativa*. Podemos rescrevê-las, matematicamente, como:



$H_0: p = 0,2$
 $H_1: p > 0,2$

Para verificar se o tratamento deu algum resultado, basta tomarmos uma amostra da lavoura de, digamos, 100 plantas e verificarmos quantas delas apresentam a doença A.

Se, por exemplo, 67 dos pés da amostra estiverem doentes, parece bastante razoável concluir que a doença se alastrou e, portanto, devemos nos decidir pela hipótese H_1 . Mas, e se encontrarmos 21, 22 ou mesmo 24 pés doentes? O que dizer? Estes valores certamente dariam um percentual acima de 20% em relação à amostra, mas, e em relação à lavoura toda? Será que são suficientes para explicar uma rejeição de H_0 e escolha H_1 ? Será que são suficientes para concluir que a doença se alastrou?

Temos que estabelecer um limite, a partir do qual consideraremos a hipótese H_0 rejeitada e optaremos por H_1 .



Mas, baseado em quê escolheremos este limite?
 O que significa, em termos estatísticos, dizer que só vamos abandonar a hipótese H_0 se encontrarmos, por exemplo, 40 ou mais pés de café doentes em nossa amostra de 100?
 Qual seria um bom limite para que a escolha de H_1 , baseada neste limite, nos desse 95% de certeza de que é de fato H_1 que ocorre na população total?

Esse é o tipo de questionamento que o teste de hipóteses coloca. Ou seja, trata-se de saber qual o melhor modo de interpretar os dados numéricos obtidos de uma pesquisa, a fim de rejeitarmos ou corroborarmos uma hipótese. Aceitar H_0 , a hipótese de nulidade, significa rejeitar a hipótese H_1 que está sendo testada. Rejeitar H_0 corresponde, neste caso, a aceitar H_1 , ou seja, corresponde a corroborar a hipótese que está sendo testada.



Vimos que o objetivo de uma pesquisa científica é por a prova a uma hipótese, para verificar se ela pode assumir o status de conhecimento científico. Tendo isso em vista, devemos projetar um teste de hipótese que evite, ao máximo, o erro de rejeitar a nulidade quando é, de fato, isto que ocorre.

Em geral, o que se faz é escolher um *nível de significância* para o teste. **O nível de significância representa a probabilidade máxima que admitimos para o erro de rejeitar H_0 quando H_0 for verdadeira (o que, neste caso, é o mesmo que escolher H_1 quando H_1 é falsa).**

Assim, no nosso exemplo, ao invés de atribuímos arbitrariamente o número 40, ou qualquer outro, como limite entre a escolha das duas hipóteses, escolhemos um nível de significância para nosso teste, digamos 5%, e calculamos o limite numérico a partir deste valor. O valor mais adequado para este limite será, então, aquele que garantir o nível de significância de 5% que escolhemos.

✓ Temos então o seguinte...

O número de pés de café doentes, que obtemos em nossa pesquisa, corresponde a nossa média amostral m . Calculamos então o desvio padrão de nossa média amostral S_m (de acordo com a equação 3.6.1) e, em seguida, os limites de 95% de confiabilidade dessa média ($L_i = m - (t_0 \times S_m)$ e $L_s = m + (t_0 \times S_m)$).

Se $L_i = m - (t_0 \times S_m)$ for menor ou igual a 20% dos pés de café da amostra,

ou seja, se $L_i \leq \frac{20 \times N}{100}$, então saberemos, com 95% de certeza, que a

doença não se alastrou. Caso $L_i > \frac{20 \times N}{100}$, então saberemos, também com

95% de certeza, que a doença se alastrou, ou seja, que mais de 20% dos pés de café da lavoura apresentam a doença A.

Este é um exemplo simples, mas bastante típico, do tipo de situação em que a inferência estatística propiciada pelos testes de hipótese auxilia os pesquisadores na interpretação dos resultados de suas pesquisas.



Existem inúmeros tipos de testes de hipóteses diferentes, utilizados em variadas situações. A escolha e aplicação do teste de hipótese mais adequado em uma determinada circunstância, nem sempre é tarefa simples e, em muitos casos, exige a ajuda de um estatístico profissional. Dentre os testes mais conhecidos, estão o teste do qui-quadrado, o teste de variância, o teste do sinal, o teste de Walsh, e muitos outros.

Não vamos nos aprofundar neste tópico, mas fica aqui nossa recomendação de leitura ao estudante interessado: **VIEIRA[1980]** e **MILLER[1977]**.



Exercício 12

O que a expressão *nível de significância* quer dizer?

<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>



4 - CONSIDERAÇÕES FINAIS

Há outros tópicos da estatística que deixamos de mencionar e que são muito importantes para a atividade do pesquisador agropecuário, como é o caso das **noções de correlação e regressão**, mas que, não estão diretamente relacionadas com o estudo deste curso de Cafeicultura Irrigada.

Nosso intuito, ao apresentar, neste curso de metodologia da pesquisa, alguns rudimentos das ferramentas estatísticas, úteis à pesquisa agropecuária, não foi o de instrumentalizar o estudante para realizar o trabalho estatístico das pesquisas, o que seria esperado de um curso de estatística, mas foi o de evidenciar o valor, a força, a importância e o caráter fundamental da estatística para o desenvolvimento da pesquisa científica agropecuária. Esperamos, sinceramente, tê-lo atingido e recomendamos, ao estudante interessado, que consulte a bibliografia indicada.



REFERENCIAL DE RESPOSTAS

A seguir, você encontrará as respostas dos exercícios. Se errou algum, não desanime, volte ao tópico referente à dúvida que você teve durante o exercício e releia-o atentamente. Se ainda assim persistir alguma dúvida, entre em contato com seu tutor!
Estaremos juntos na próxima unidade. Parabéns pelo seu esforço!



Exercício 1

O que distingue, fundamentalmente, a forma de pesquisar quantitativa da qualitativa, é o tipo de pergunta que o pesquisador quer responder e como ele encaminha seu trabalho para obter as respostas. Questões relacionadas ao significado dos fenômenos sociais que, por exemplo, levam em conta motivações, crenças, valores, são inadequadas para serem tratadas por métodos matemáticos e estatísticos. Elas representam o universo onde se insere a pesquisa qualitativa.

Exercício 2

Retome a Unidade de Estudo 5 do Bloco Temático I, para comprovar sua resposta.

Exercício 3

Aqui vão apenas dois exemplos:


1. Pesquisa Qualitativa: um levantamento comparativo sobre as principais características das divindades em que acreditam as diversas nações indígenas que habitam a Amazônia brasileira.
2. Pesquisa Quantitativa: um levantamento sobre qual a quantidade ideal de compostos orgânicos na água para a irrigação de uma certa lavoura de café, no município de Monte Carmelo-MG.

Exercício 4


Existem várias possibilidades de atribuição de variáveis, que dependem de como se deseja projetar a pesquisa. O fundamental é que saibamos separar as diversas categorias de informação em variáveis e que saibamos diferenciar as que serão manipuladas na pesquisa (as variáveis independentes), das variáveis em que verificaremos as alterações provocadas por essas manipulações (as variáveis dependentes). Segue uma sugestão para cada hipótese.


- a) x_1 : produtividade do café da espécie A (var. Independente)
 x_2 : produtividade do café da espécie B (var. Independente)
 y : produtividade do café da espécie AB (var. Dependente)
- b) x : aplicação do defensivo C na lavoura, durante 2 meses (var. Independente)
 y : incidência da praga P na lavoura (var. Dependente)

- c) x_1 : quantidade de água por planta, utilizada em irrigação por gotejamento, em lavouras do café A (var. Independente)
 x_2 : quantidade de água por planta, utilizada em irrigação por pivô, em lavouras do café A (var. Independente)
 x_3 : quantidade de água por planta, utilizada em irrigação por aspersão, em lavouras do café A (var. Independente)
 y_1 : produtividade de lavouras irrigadas por gotejamento (var. Dependente)
 y_2 : produtividade de lavouras irrigadas por pivô (var. Dependente)
 y_3 : produtividade de lavouras irrigadas por aspersão (var. Dependente)

Exercício 5 

1. Tratamentos múltiplos e uma só variável independente (2.3.3)
2. Um só grupo experimental (2.3.1)
3. Grupos experimentais independentes (2.3.2)
4. Múltiplas variáveis independentes (2.3.4)


Exercício 6 

 Amostra indicada por círculos:

$$m = \frac{112 + 99 + 85 + 97 + 118 + 136 + 112 + 103 + 92 + 132 + 97 + 98}{12} = 106,75$$

$$me = \frac{99 + 103}{2} = 101$$


$$mo_1 = 97 \text{ e } mo_2 = 112$$

 Amostra indicada por hexágonos:


$$m = \frac{114 + 109 + 127 + 102 + 100 + 114 + 122 + 103 + 87 + 103 + 111 + 101}{12} = 107,75$$

$$me = \frac{103 + 109}{2} = 106$$


$$mo_1 = 103 \text{ e } mo_2 = 114$$

Exercício 7 

Tal fato é bastante útil para identificarmos certos fatores que podem estar afetando os dados de uma pesquisa. Identificados esses fatores, temos duas opções: ou os isolamos, mantendo-os inalterados, ou, caso não seja possível, associamos uma variável independente a eles e medimos sua variação, no decorrer da pesquisa.

Exercício 8 

A criação de um resumo ou esquema é muito pessoal, no entanto, você deve rever cada parte ou item deste resumo ou esquema, comparando com os textos e fórmulas apresentados neste material, para reforçar seu aprendizado e para eliminar possíveis erros.

Exercício 9 

Amostra indicada por círculos:

$$s^2 = \frac{(112 - 106,75)^2 + (99 - 106,75)^2 + \dots + (98 - 106,75)^2}{12 - 1} = \frac{2706,25}{11} = 246,0227$$

$$s = \sqrt{246,0227} = 15,685$$


$$cv = \frac{100 \times 15,685}{106,75} = 14,693\%$$

Amostra indicada por hexágonos:

$$s^2 = \frac{(114 - 107,75)^2 + (109 - 107,75)^2 + \dots + (103 - 107,75)^2}{12 - 1} = \frac{1278,25}{11} = 116,2045$$

$$s = \sqrt{116,2045} = 10,779$$

$$cv = \frac{100 \times 10,779}{107,75} = 10,003\%$$

Exercício 10 

Amostra indicada por círculos:

$$L_i = m - (t_\alpha \times s) = 106,75 - (2,23 \times 15,685) = 71,773$$


$$L_s = m + (t_\alpha \times s) = 106,75 + (2,23 \times 15,685) = 141,727$$

Amostra indicada por hexágonos:

$$L_i = m - (t_\alpha \times s) = 107,75 - (2,23 \times 10,779) = 83,713$$

$$L_s = m + (t_\alpha \times s) = 107,75 + (2,23 \times 10,779) = 131,787$$


Dizer que L_i e L_s representam os limites inferiores e superiores para o *intervalo de confiança* de 95% de uma amostra de alturas de pés de café de uma lavoura, significa dizer que, com 95% de certeza, qualquer pé de café da lavoura tem altura superior a L_i e inferior a L_s . Ou seja, L_i e L_s definem uma faixa de valores que compreende 95% dos dados de toda a população.

Exercício 11 

Como vimos, para estimar o tamanho N de uma amostra, trabalhamos com uma amostra-piloto a fim de obtermos uma estimativa do desvio padrão para, utilizando a equação 3.6.2, calcularmos o valor de N . Neste caso, nossa amostra piloto é a amostra amarela dos exercícios anteriores. Temos:

$$N = \left(t_0 \times \frac{s}{D} \right)^2 = \left(2,23 \times \frac{10,779}{5} \right)^2 = 4,807^2 = 23,111 \approx 24 \text{ (arrendondado para cima)}$$

Dessa forma, temos 95% de certeza que a média real M estará compreendida no intervalo $(m-5, m+5)$, quando m é a média calculada através de uma amostra de 24 pés de café.

Exercício 12 

A expressão **nível de significância** se relaciona aos testes de hipóteses. Um teste de hipótese, como vimos, representa um método estatístico de análise dos dados obtidos em uma pesquisa, com fins de decidir sobre a corroboração ou rejeição da hipótese definidora da pesquisa. O **nível de significância** de um teste de hipótese representa a probabilidade máxima que se admite para o erro de corroborar uma hipótese que, na verdade, é falsa. Assim, quando afirmamos que o nível de significância de um teste de hipótese é 5%, isso significa que há 5% de probabilidade do teste apontar, como verdadeira, uma hipótese falsa.

BIBLIOGRAFIA

- VIEIRA, S. **Introdução à bioestatística**. 3.ed. Rio de Janeiro: Ed. Campus, 1980.
- REY, L. **Planejar e redigir trabalhos científicos**. 2.ed. São Paulo: Edgard Blücher, 1998.
- LITTLE, T. & HILLS, F. J. **Métodos estadísticos para la investigación en la agricultura**. México: Ed. Trillas, 1976.
- MILLER, S. **Planejamento experimental e estatística**. Rio de Janeiro: Zahar, 1977.
- GOMES, F. P. **A estatística moderna na pesquisa agropecuária**. 3.ed. Piracicaba: POTAFOS, 1987.
- PÁDUA, E. M. M. de. **Metodologia da pesquisa: abordagem teórico-prática**. 3.ed. Campinas: Papyrus, 1998.