Centro de Ciências Agrárias e Ambientais da UFBA
Departamento de Engenharia Agrícola

Disciplina: AGR116 – Bioestatística
Professor: Celso Luiz Borges de Oliveira
Assunto: Estatística Descritiva
Tema: Medidas de Posição e Medidas de Dispersão

RESUMO E NOTAS DA AULA Nº 02

A Estatística desempenha três grandes funções: Descritiva, Indutiva ou Inferencial e a de Planejamento.

a) **Descritiva** – descreve um conjunto de dados variáveis, reduzindo-os a um pequeno número de medidas que contém toda a informação relevante. Utiliza número para descrever fatos. Somente descreve e avalia certo grupo (amostra), sem tirar quaisquer conclusões ou *inferências* sobre um grupo maior (população).

b) **Indutiva ou Inferencial** – diz respeito à análise e interpretação de dados amostrais. Consiste me obter e generalizar conclusões sobre a população a partir de uma amostra. Utiliza-se da estimação de *parâmetros* e verificação de hipóteses, esta por meio, da aplicação dos *testes de significância*.

c) **Planejamento** – auxilia no delineamento de experimentos e levantamento para, dentro de uma precisão estipulada, obter-se a informação desejada livre da influência de fatores perturbadores. A Estatística fornece os preceitos da *casualização*, *repetição*, *controle local*, os delineamentos experimentais e os métodos de amostragem, ou seja, normas lógicas que garantam a validez das comparações entre *tratamentos* e aumentem a precisão dessas comparações.

ESTATÍSTICA DESCRITIVA

Os dados são descritos por meio de medidas que os representem de forma sumária, utilizando-se as medidas de posição ou de tendência central e as medidas de dispersão ou de variação.

1. Medidas de Posição $\begin{cases} Média \\ Mediana \\ Moda \end{cases}$

1.1 Média

1.1.1 Média Aritmética ($\mu, \bar{x}$)

A média aritmética representa o valor central em torno do qual os dados se distribuem. Fornece-nos uma síntese da grandeza, ou seja, da magnitude dos dados de uma população.

Seja o conjunto de **N** variantes de uma variável X. Define-se a média aritmética para a população por:

$\overline{X} = (x_1 + x_2 + ... + x_N) \therefore \overline{X} = \sum x_i / N$

Selecionando-se variantes desta mesma população poderia ser obtida a estimativa da média da população, para tanto, calcula-se:

$\bar{x} = (x_1 + x_2 + x_n)/n \therefore \bar{x} = \sum x_i / n$

Exemplo 1:

Calcular a média aritmética para o conjunto de valores que a variável X (peso de cinco alunos, em kg/aluno) pode assumir:

X = {50, 70, 45, 60,75}

$\bar{x} = \sum x_i / n$ = 300/5 = 60kg/aluno

Propriedades da Média Aritmética

a) A soma dos desvios de um conjunto de dados em relação à sua média é nula.

$(x_1 - \bar{x}) + (x_2 - \bar{x}) + ... + (x_n - \bar{x}) = 0 \qquad \sum (x_i - \bar{x}) = 0 \therefore \sum x_i - n\bar{x} = 0$

Lembrando que, $\bar{x} = \sum x_i / n$,

Logo; $n\bar{x} - n\bar{x} = 0 \therefore 0 = 0$

b) A média é o valor que torna mínima a “soma do quadrado dos desvios” (SQD)

$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2 = mínima$

Façamos, $(x_1 - m)^2 + (x_2 - m)^2 + ... + (x_n - xm)^2 = \sum (x_i - m)^2$

Z = $\sum (x_i - m)^2 \therefore d_z / d_m = 0$ logo, 2$\sum (x_i - m) = 0 \therefore \sum x_i - nm = 0$

$\frac{\sum x_i}{n} = m$ (esta é a fórmula da média)

1.1.2 Média Ponderada ($\bar{x}_p$)

Se as variantes de um conjunto de dados estiverem associadas às freqüências ou pesos, obtém-se a média ponderada da seguinte forma:

$$\bar{x}_p = \frac{y_1 f_1 + y_2 f_2 + ... + y_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum x_i f_i}{\sum f_i} \qquad \bar{x}_p = \frac{\sum x_i f_i}{\sum f_i}$$

Exemplo 2 :

Produção de leite (kg/c)	nº de vacas
15	10
10	6
8	4

$$\bar{x}_p = \frac{(15)(10) + (10)(6) + (8)(4)}{10 + 6 + 4} = 12{,}1 kg / c$$

1.2 Mediana (Md)

Colocados os valores de um conjunto de dados em ordem crescente, mediana é o elemento que ocupa a posição central, ou seja, existem números iguais de elementos antes e depois da mediana, ou seja, é o valor abaixo e acima do qual se tem a metade dos dados, ou ainda, é o valor central desse conjunto de dados.

Mediana para dados não-agrupados em classes:

1º caso : O número de elementos é ímpar.

Seja uma variável aleatória X assumindo os seguintes valores:

X={14,8,10,5,7}, Calcule a Mediana.

A Mediana é o elemento que ocupa a posição $\frac{n+1}{2}$, ou seja,

Md= $X_{\frac{n+1}{2}}$, em que, Md = 8

2º caso: O número de elementos é par. X={14,8,10,5,7,15},

A Mediana é a media aritmética dos valores centrais da ordem $\frac{n}{2} e \frac{n}{2} + 1$

Md= $\frac{n_{\frac{n}{2}} + n_{\frac{n}{2}+1}}{2}$, em que, Md = 9

1.3 Moda (Mo)

Em algumas situações, a distribuição das observações é tal que as freqüências são maiores nos extremos. Nesses casos, a utilização da media e da mediana é contra indicada, pois são valores pouco representativos do conjunto e o uso da moda poderá então ser considerado.

A moda é o valor de maior ocorrência dentro de um conjunto de observações. Um serie de dados, com relação à moda, pode ser classificada em:

- Amodal - não possui moda;
- Unimodal - possui apenas uma moda;
- Bimodal - possui duas modas;
- Multimodal - possui mais de duas modas.

Moda para dados não-agrupados em classes:

Exemplo:

1) Identificar a moda para os dados da Tabela 1 e Tabela 2, abaixo:

Tabela 1

x_i	f_i
0	4
1	5
2	7
3	3
4	2
5	1
	22

Mo= 2

Tabela 2

x_i	f_i
3	1
5	4
9	2
10	4
15	3

Bimodal

Mo_1= 5 Mo_2= 10

Relação entre média, mediana e moda

a) Distribuição simétrica: $\bar{x} = Mo = Md$

b) Distribuição assimétrica: a média e a mediana se deslocam.

   i. Assimetria positiva: $\bar{x} > Md > Mo$
   ii. Assimetria negativa: $\bar{x} < Md < Mo$

2. Medidas de Dispersão

Estas medidas avaliam o grau de variação ou heterogeneidade dos dados. Caracterizar um conjunto de dados por medidas de posição é inadequado, pois conjuntos com medidas de posição semelhantes podem apresentar características muito diferentes, p.ex.; com relação à variabilidade do conjunto de valores.

As medidas de dispersão são *estatísticas* (*estimadores*) descritivas, que quantificam de algum modo a variabilidade dos dados, geralmente utilizando como referência uma medida de posição.

Exemplo 3:

Amostra A: {4, 8, 3, 9, 7, 5}
Amostra B: {1, 5, 2, 14, 3, 11}

Notamos que $\bar{x}_A = 6$ e $\bar{x}_B = 6$, porém, a dispersão dos valores na amostra B é maior, portanto precisamos de alguma informação a mais que permita diferenciá-las.

$$\text{Medidas de Dispersão} \begin{cases} \textit{Amplitude Total} \\ \textit{Variância} \\ \textit{Desvio Padrão ou Erro Padrão} \\ \textit{Coeficiente de Variação} \\ \textit{Variância da Média} \\ \textit{Desvio ou Erro Padrão da Média} \\ \textit{Coeficiente de Variação da Média} \end{cases}$$

2.1. Amplitude Total (A_t)

É a diferença entre o maior e menor valor de uma série de dados. É uma medida de dispersão não muito informativa, pois levam em conta apenas os dados extremos de uma amostra. Tem importância para os estudos dos valores máximos e mínimos.

Para os dados do Exemplo1: X = {50, 70, 45, 60,75} temos;
A_t = 75 – 45 = 30kg/aluno

2.2. Variância (s^2)

Esta medida leva em conta todas as observações da amostra e mede a dispersão desses valores em torno da média. É dada pela soma dos quadrados dos desvios (SQD) em relação a media aritmética, dividida pelo número de graus de liberdade (G.L)*. Por definição é a média dos quadrados dos desvios.

Para a população: $$\sigma^2 = \frac{SQD}{N} = \frac{\sum (x-\mu)^2}{N} = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{N}}{N}$$

Para a amostra: $$s^2 = \frac{SQD}{GL} = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}$$

Considerando os dados do Exemplo 1: X = {50, 70, 45, 60,75} e n= 5 temos;

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{18650 - \frac{(300)^2}{5}}{5-1} = \frac{18650 - \frac{90000}{5}}{4} = \frac{18650 - 18000}{4} = 162{,}5(kg/aluno)^2$$

(G.L)*: De uma maneira geral, o número de graus de liberdade associados a uma estatística é o número de elementos da amostra, n , menos o número de parâmetros já estimado. Existem n-1 desvios independentes, e, é possível demonstrar que, utilizando-se o denominador n-1, obtém-se um estimador de melhor qualidade da variância populacional σ^2 pela variância da amostra s^2.

Considerações sobre a Variância

1. O sinal dos desvios é irrelevante, uma vez que na fórmula os desvios são elevados ao quadrado.
2. A magnitude da diferença é que vai determinar maior ou menor peso para a variância.
3. A variância de um único dado é indeterminada. Isso revela obviamente, de que nada podemos dizer da variação de dados sem repetição.
4. Somando-se ou subtraindo-se uma mesma constante (k) a todas as observações de um conjunto de dados a variância não se altera.
5. Multiplicando-se ou dividindo-se cada observação de um conjunto de dados por uma mesma constante (k≠0), a variância fica multiplicada ou dividida pelo quadrado dessa constante.

Como medida de dispersão a variância tem a desvantagem de apresentar unidade de medida igual ao quadrado da unidade dos dados observados, p.ex.; se os dados são medidos em metros (m), a variância é dada em metros ao quadrado $(m)^2$. Para voltarmos à unidade de medida original, precisamos de uma outra medida de dispersão, o Desvio Padrão.

2.3. Desvio Padrão ou Erro Padrão (s)

É a raiz quadrada positiva do valor da variância, avalia a variação dos dados e nos permite discutir os resultados na mesma unidade de mensuração dos dados. O desvio padrão da população e da amostra é representado, respectivamente por σ e s.

$$\sigma = \sqrt{\sigma^2} \qquad \text{e} \qquad s = \sqrt{s^2} \text{ ,}$$

Considerando os dados do cálculo da variância no exemplo anterior, tem-se;

$s = \sqrt{s^2} \therefore s = \sqrt{162{,}5(kg/aluno)^2}$ = 12,75 kg/aluno

2.4. Coeficiente de Variação (CV)

Frequentemente se tem o interesse em comparar variabilidades de diferentes conjuntos de valores. A comparação se torna difícil em situações onde as medias são muitos desiguais ou as unidades de medidas são diferentes.

O Desvio Padrão é uma medida absoluta da dispersão e o Coeficiente de Variação é uma medida de dispersão que expressa percentualmente o Desvio Padrão por unidade de média, ou seja, o CV representa o Desvio Padrão que seria obtido se a média fosse igual a 100.

$$CV = \frac{s}{\bar{x}}.100$$

Utilizando-se o valor obtido para a média ($\bar{x} =$60kg/aluno) e o desvio padrão (s = 12,75kg/aluno), temos que;

$$CV = \frac{12{,}75}{60}.100 = 21\%$$

Segundo vários autores, os coeficientes de variação nos experimentos podem ser considerados:

- Baixo: quando o CV for menor do que 10%;
- Médio: entre 10%-20%;
- Alto: quando for superior a 20% e menor ou igual a 30%;
- Muito alto: superior a 30%.

APLICAÇÃO:

- Utilizado para avaliação da precisão de experimentos;
- Para analisar qual amostra é mais homogênea (menor variabilidade). Na situação em que as amostras possuem a mesma media, a conclusão pode ser feita a partir da comparação de suas variâncias. Para amostras com médias diferentes, aquela que apresentar menor CV, é a mais homogênea.

2.5. Variância da Média ($s^2(\bar{x})$)

Se tivéssemos varias amostras provenientes de uma mesma população, obteríamos também diversas estimativas da media, e provavelmente, distintas entre si. A partir dessas diversas estimativas da média, poderíamos estimar uma variância; considerando os desvios de cada média, em relação a média de todas elas, então teríamos uma variância da média.

Entretanto, demonstra-se que a partir de uma única amostra podemos estimar essa variância pela fórmula:

$\hat{V}(\bar{x}) = s^2(\bar{x}) = \frac{s^2}{n}$, o caso em que a população for infinita ou se a amostragem for feita com reposição, ou ainda em termos práticos, quando a fração de amostragem $f = \frac{n}{N} < 0{,}05$.

Quando a população for finita ou a amostragem for sem reposição, ou ainda, quando $f = \frac{n}{N} \geq 0{,}05$, então a variância da média será:

$\hat{V}(\bar{x}) = s^2(\bar{x}) = \frac{s^2}{n} \cdot \frac{N-n}{N-1}$, em que; **s²** é a estimativa da variância de **n** dados, e N é o tamanho da população.

Ao termo acrescido à fórmula da variância, chamamos de Fator de Correção $\left(\frac{N-n}{N-1}\right)$.

2.5. <u>Desvio ou Erro Padrão da Média</u> (s($\bar{x}$))

Dá-nos uma idéia da precisão com a média foi obtida, ou seja, o quanto o valor da estimativa da média ($\bar{x}$) se aproxima ou se afasta da média verdadeira (μ). É calculada extraindo-se a raiz quadrada da variância da média.

$s(\bar{x}) = \sqrt{\hat{V}(\bar{x})} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$, para a média de uma amostra retirada de uma população infinita.

$s(\bar{x}) = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$, para a média de amostra de população finita.

Verifica-se pela fórmula, que quanto maior for o tamanho da amostra menor será a variância da média, e consequentemente o desvio ou erro padrão da média, pois este é inversamente proporcional ao tamanho da amostra.

2.5. <u>Coeficiente de Variação da Média</u> (CV($\bar{x}$))

Quando se torna necessário, para fins comparativos, expressar o desvio padrão da média em termos de percentagem das suas respectivas médias, calcula-se o Coeficiente de Variação da Média, pela equação:

$$CV(\bar{x}) = \frac{s(\bar{x})}{\bar{x}} \cdot 100$$

GLOSSÁRIO

Casualização - para evitar que os tratamentos sejam favorecidos por quaisquer fatores externos, procede-se à casualização, ou seja, as unidades experimentais são designadas de forma totalmente casual. Tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer unidade experimental (a casualização dá validade ao teste de hipótese).

Controle Local - tem por finalidade dividir um ambiente heterogêneo em sub-ambientes homogêneos e tornar o delineamento experimental mais eficiente pela redução do *erro experimental*.

Erro Experimental - é a medida das variações existentes entre os dados ou observações que se apresentam nas unidades experimentais que recebem tratamentos iguais, ou seja, é a causa da variação que reflete os efeitos do acaso.

Estatísticas (Estimador)- são medidas descritivas das características das amostras. ($\bar{x}, s^2, s$)

Estimativa - são os valores assumidos pelos estimadores ou estatísticas.

Inferência (estimação)- processo pelo qual se obtém informações sobre a população a partir de amostras.

Parâmetro **-** são medidas descritivas das características da população

Repetição - este princípio se refere ao uso de mais de uma parcela por tratamento. Tem por finalidade propiciar uma estimativa do erro experimental, melhorar a precisão do experimento e fazer com que o teste de hipótese seja possível.

Testes de Significância - são regras ou normas que nos permitem decidir se aceitamos ou rejeitamos uma determinada hipótese, ou se a amostra observada difere significativamente dos valores esperados.

Tratamento - são alternativas de um fator a ser estudado, devendo seu efeitos serem medidos e comparados, além de classificados em quantitativos e qualitativos.

BIBLIOGRAFIA CONSULTADA

Barbetta, P. A. *Estatística Aplicada às Ciências Sociais*. Florianópolis. 3ª ed. Editora DA UFSC. 1999.284p.

Costa, J. A. *Notas e Resumos de aulas da disciplina AGR116-Bioestatística*. DEA/AGRUFBA.

Costa, S. F. *Introdução Ilustrada à Estatística*. São Paulo. 3ª ed. Editora Harbra Ltda. 1998. 313p.

Costa Neto, P. L. de O. *Estatística*. Ltda. São Paulo 17ª ed. Editora Edgard Blücher. 1999. 264p.

Regazzi, A. J. *Curso de Iniciação Estatística*: Roteiro de Aulas. Viçosa. DPI/UFV. 1997. 138p.

Santos, J. W. dos; Gheyi, H. R. (Eds.) *Estatística Experimental Aplicada*. Campina grande: Editora Gráfica Marcone Ltda, 2003. 213p. Tópicos de Engenharia Agrícola e Agronômica.

Wonnacott, R. J. e Wonnacott, T. H. *Fundamentos de Estatística*; tradução de Alfredo Alves de Farias. Rio de Janeiro. LTC – Livros Técnicos e Científicos Editora S.A., 1985. 357p.

Centro de Ciências Agrárias e Ambientais da UFBA
Departamento de Engenharia Agrícola

Disciplina: AGR116 – Bioestatística
Professor: Celso Luiz Borges de Oliveira
Assunto: Estatística Descritiva
Tema: Estudo de População e Aplicação da Estatística Descritiva

EXERCÍCIO PRÁTICO DE CAMPO Nº 01

Selecione uma população para ser estudada utilizando o plano de amostragem aleatória simples, e:

1. Defina inicialmente:

a) Objetivo da amostragem
b) variável
c) parcela
d) unidade de peso ou medida
e) tamanho da população

2. Analise a variável respondendo as seguintes questões:

   a) Após o cálculo das estimativas da média, variância, desvio padrão e o erro padrão da média de uma amostra prévia ou piloto, dimensione o tamanho da amostra definitiva, com 5% de significância, admitindo um grau de precisão de 10% para a média.
   b) Obtenha o intervalo de confiança para a média da população ao nível de 10% de significância.
   c) Estime o intervalo de confiança para o total da população, ao nível de 5% de significância. Caso não tenha o tamanho da população, idealize um valor aproximado.
   d) Caso fosse requerida uma precisão de 5% ou 1% na estimativa da média, quais seriam os novos tamanhos das amostras? Discuta os resultados.
   e) Pressupondo que a estimativa da variância para a característica da sua população fossem 10%, 20% e 30% maior do que a encontrada para a amostra definitiva, o quanto esses dados influenciariam no tamanho da amostra.
   f) Resumam em um quadro os resultados dos estimadores e estimativas para a sua população e a de outra equipe e responda:
      f.1) Qual a população que apresentou a maior variabilidade? Por quê?
      f.2) Caso os tamanhos das amostras definitivas entre as populações tenham sido diferentes, a que se atribui o resultado da questão anterior? Discuta.
      f.3) Qual a média mais precisa?
   g) Caso a amostra prévia fosse a definitiva, qual o grau de precisão estimado para a média com 95% de confiança?

“A educação é a caminhada do homem da ignorância excessivamente confiante para a incerteza refletida” **Don Clark**

Tabela de *t* – Amplitude Total Estudentizada

Distribuição de t (*Student*)

GL	α				
	10%	**5%**	**2%**	**1%**	**0,1%**
01	6,31	12,71	31,82	63,66	636,62
02	2,92	4,30	6,97	9,93	31,60
03	2,35	3,18	4,54	5,54	12,92
04	2,13	2,78	3,75	4,60	8,61
05	2,02	2,57	3,37	4,03	6,87
06	1,94	2,45	3,14	3,71	5,96
07	1,90	2,37	2,37	3,50	5,41
08	1,86	2,31	2,90	3,36	5,04
09	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,44
12	1,78	2,18	2,68	3,06	4,32
13	1,77	2,16	2,65	3,01	4,22
14	1,76	2,15	2,62	2,98	4,14
15	1,75	2,13	2,60	2,95	4,07
16	1,75	2,12	2,58	2,92	4,02
17	1,74	2,11	2,57	2,90	3,97
18	1,73	2,10	2,55	2,88	3,92
19	1,73	2,09	2,54	2,86	3,88
20	1,73	2,09	2,53	2,85	3,85
21	1,72	2,08	2,52	2,83	3,82
22	1,72	2,07	2,51	2,82	3,79
23	1,71	2,07	2,50	2,81	3,77
24	1,71	2,06	2,49	2,80	3,75
25	1,71	2,06	2,49	2,79	3,73
26	1,71	2,06	2,48	2,78	3,71
27	1,70	2,05	2,47	2,77	3,69
28	1,70	2,05	2,47	2,76	3,67
29	1,70	2,05	2,46	2,76	3,66
30	1,70	2,04	2,46	2,75	3,65
40	1,68	2,02	2,42	2,70	3,55
60	1,67	2,00	2,39	2,66	3,46
120	1,66	1,98	2,36	2,62	3,37
∞	1,65	**1,96**	2,33	**2,58**	3,29

α = nível de significância (10%, 5%, 1%)
1 – α = grau de confiança (90%, 95%, 99%)