



Bioestatística

Distribuição de Frequência

[Parte 1 de 2]

Prof. Me. Yuri Albuquerque

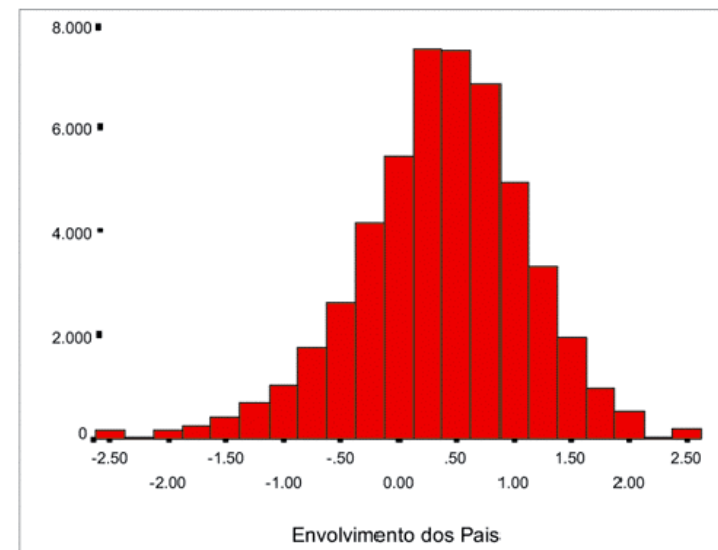
Distribuição de Frequência



Organização de Dados Estatísticos

Quando ao final de uma pesquisa todos os dados tiverem sido coletados, é provável que o pesquisador se depare com um grande número de planilhas numéricas ou de itens que, geralmente, tornam-se verdadeiros problemas a fim de que ele possa obter informações claras sobre o fenômeno estudado.

O primeiro passo em uma análise de dados é a organização e o resumo dos mesmos, para que eles possam ser rearranjados sob uma forma mais compreensível, o que, tradicionalmente, envolve o uso de uma **série estatística**, da **distribuição de frequência** e de sua **representação gráfica**.



Distribuição de Frequência



Série Estatística

A organização de uma série estatística é, usualmente, a primeira tarefa em qualquer tipo de análise de dados estatísticos, sendo está definida como uma **coleção de dados quantitativos**, sucessivamente dispostos de acordo com uma organização característica da variável em estudo, considerando-se, também, a disposição **temporal** ou **espacial dos dados**.

Desta forma, para que se possa diferenciar uma série estatística da outra, é necessário que se considere o fenômeno descrito, o local onde o fenômeno aconteceu e a época a que se refere o fenômeno estudado.

Produção de ovos de galinha por região geográfica do Brasil - 2010

Região	Ovos de galinha (mil dúzias)
Norte	116.843
Nordeste	497.904
Sudeste	1.433.915
Sul	845.751
Centro-Oeste	352.307

Fonte: IBGE, Produção da Pecuária Municipal, 2010.

Distribuição de Frequência



Série Estatística

Portanto, com base nestas características, as séries estatísticas podem ser subdivididas em quatro tipos, conforme mostrado no quadro ao lado:

Tipos de série estatística conforme a característica estudada

Tipo de série	Elemento variável	Elementos fixos	Definição
Cronológica(histórica)	Época	Local Fenômeno	• Formada por dados coletados ao longo do tempo (dia, mês, ano, hora, minuto etc.).
Geográfica(espacial)	Local	Época Fenômeno	• Constituída por dados provenientes de diferentes regiões geográficas (continente, país, cidade, localidade etc.).
Específica (categórica)	Fenômeno	Época Local	• Compreende um conjunto de dados coletados a partir de diferentes categorias de uma mesma variável (cores, produtos, forma etc.).
Distribuição de frequência	Local Época Fenômeno	Magnitude dos dados	• Constituída por dados referentes ao fenômeno estudado, variando a gradação entre eles.

Distribuição de Frequência



Série Estatística – Exemplo

Para representar uma série estatística, a *Tabela* abaixo mostra os valores do consumo de suco de frutas, coletados a partir de uma amostra de 60 alunos de uma escola pública, em uma determinada cidade.

*Valores do consumo diário de suco de frutas, por 60 estudantes
de uma escola pública (mL/dia)*

192	196	173	193	195	204	235	190	194	184
194	218	204	210	214	198	196	215	202	201
203	197	195	229	207	203	208	198	194	193
195	198	189	204	202	215	199	217	195	209
181	205	183	195	183	190	182	219	170	197
207	211	206	166	194	186	227	185	201	186

Distribuição de Frequência



Série Estatística – Exemplo

Mas o que pode ser feito para tornar as informações, contidas nos dados, mais compreensíveis? A primeira tarefa é ordená-los de forma **crescente** ou **decrescente** para transformá-los em uma **série estatística** conhecida como **rol** (dados ordenados).

Essa ordenação propicia algumas vantagens, em termos de inferências, em relação à sua forma bruta, tal como:

- Alguns estudantes apresentam consumos iguais;
- Os valores do consumo de suco de frutas variam de estudante para estudante;
- É possível observar de forma mais ampla a variação dos dados;
- Os valores dos consumos mínimo e máximo são facilmente percebidos;
- É possível observar uma tendência para a concentração dos valores na faixa de 193 à 199 mL.



Distribuição de Frequência



Série Estatística – Exemplo

COLOQUEM A TABELA EM ORDEM

*Valores do consumo diário de suco de frutas, por 60 estudantes
de uma escola pública (mL/dia)*

192	196	173	193	195	204	235	190	194	184
194	218	204	210	214	198	196	215	202	201
203	197	195	229	207	203	208	198	194	193
195	198	189	204	202	215	199	217	195	209
181	205	183	195	183	190	182	219	170	197
207	211	206	166	194	186	227	185	201	186

Distribuição de Frequência



Série Estatística – Exemplo

ROL – Valores do Consumo diário de suco de frutas, por 60 estudantes de um escola pública (mL/dia)

166	170	173	181	182	183	183	184	185	186
186	189	190	190	192	193	193	194	194	194
194	195	195	195	195	195	196	196	197	197
198	198	198	199	201	201	202	202	203	203
204	204	204	205	206	207	207	208	209	210
211	214	215	215	217	218	219	227	229	235



Série Estatística – Exemplo

Portanto, é interessante notar que, embora o **rol** possa fornecer mais informações e com menor esforço de busca, em relação aos dados brutos, ainda assim persiste a questão da análise de 60 observações, fato este agravado quando a quantidade de dados for extremamente grande.

Por outro lado, se os dados formam um conjunto relativamente constituído por poucos valores, muitos dos quais são repetidos, pode-se, simplesmente, contar quantas vezes cada valor ocorre e, então, apresentar o resultado sob a forma de uma tabela de frequência.

Distribuição de Frequência



Distribuição de Frequência

Quando trabalhamos com um grande número de dados, e até quando eles não são tão numerosos assim, podemos encontrar dificuldades em obter uma apresentação clara desses dados, de tal forma que eles possam nos transmitir todas as informações que eles contêm.

Assim, como no exemplo, a organização dos dados requer, inicialmente, que eles sejam rearranjados e dispostos, em sua forma bruta, em algum formulário especial. Posteriormente, esses dados devem ser organizados em intervalos bem definidos chamados de categorias ou classes, as quais são dispostas em tabelas conforme a frequência de cada classe, formando uma organização estatística conhecida como **distribuição de frequência**.

Distribuição de Frequência



Distribuição de Frequência

É representada por uma tabela que mostra os intervalos de classe referentes à organização dos valores dos dados coletados, com as suas respectivas frequências (f) com que esses valores ocorrem em cada uma das classes.

- Quando os dados são agrupados de acordo com um tamanho numérico, ou seja, quando são provenientes de uma variável quantitativa, a tabela é chamada de **distribuição quantitativa** ou **numérica**.
- Porém, quando eles são agrupados em categorias não-numéricas, ou seja, provenientes de uma variável qualitativa, a tabela é dita **distribuição qualitativa** ou **categórica**.



Distribuição de Frequência



Distribuição de Frequência

As tabelas abaixo mostram exemplos de distribuição quantitativa e qualitativa. Ambas mostram a apuração das notas e respectivas frequências, obtidas por 40 alunos de uma turma do curso de bioestatística. Na Tabela a esquerda, as classes estão organizadas por nota e, na Tabela a direita, pelos respectivos conceitos.

*Tabela – Exemplo de
distribuição quantitativa*

Classe	Frequência
0,0 - 0,9	2
1,0 - 3,9	8
4,0 - 6,9	16
7,0 - 8,9	10
9,0 - 10,0	4
Total	40

*Tabela – Exemplo de
distribuição qualitativa*

Classe	Frequência
Semceito	2
Insuficiente	8
Regular	16
Bom	10
Excelente	4
Total	40

Distribuição de Frequência



Distribuição de Frequência

De qualquer maneira, quando a variável estudada for contínua, é conveniente agrupar os valores coletados em classes. Do mesmo modo, se a variável em estudo for discreta, e o número de valores coletados for muito grande, é recomendado o agrupamento dos dados em categorias (classes), o que reduz o tamanho da tabela e evita o aparecimento de valores com frequência nula, melhorando consideravelmente a compreensão do fenômeno em estudo.

Distribuição de Frequência



Amplitude total (At)

Assim, para que se possa entender a organização de uma distribuição de frequência, é necessário o conhecimento prévio de algumas definições.

Amplitude total (At) – em um conjunto de dados, corresponde à diferença entre o maior e o menor valor observado para a variável em estudo.

Exemplo: O conjunto de dados representa o consumo diário de suco de frutas, de uma amostra de 60 estudantes matriculados em uma escola pública.



Distribuição de Frequência



Amplitude total (At)

166	170	173	181	182	183	183	184	185	186
186	189	190	190	192	193	193	194	194	194
194	195	195	195	195	195	196	196	197	197
198	198	198	199	201	201	202	202	203	203
204	204	204	205	206	207	207	208	209	210
211	214	215	215	217	218	219	227	229	235

$$At = 235 - 166 = 69$$

Distribuição de Frequência



Classe

Corresponde, em uma série ou num conjunto de dados, a um grupo ou divisão que apresenta características semelhantes, da mesma categoria. Ou seja, é um dos grupos em que se divide a amplitude total do conjunto de valores observados para a variável em estudo.

Classes (nº de mL)	Frequência (nº de alunos)
166 – 174	3
175 – 183	4
184 – 192	8
193 – 201	21
202 – 210	14
211 – 219	7
220 – 228	1
229 – 235	2
	$\Sigma f = 60$

Classes de frequência

Frequência de classes

Limites de classe

Soma das frequências é igual ao número de indivíduos da amostra

Distribuição de Frequência



Limites de classe

Correspondem aos **valores extremos** de uma determinada classe. A primeira classe do exemplo mostrado na tabela anterior tem como limites os valores 166 e 174. O valor 166 é denominado **limite inferior**, enquanto o valor 174 é denominado **limite superior**. Em uma distribuição de frequência para dados contínuos, é desejável que os limites de classe sejam números que separem as classes sem haver lacunas entre elas.

Distribuição de Frequência



Intervalo de classe

Corresponde ao conjunto de observações contidas entre os dois valores limites de uma classe. Em uma distribuição de frequência, os intervalos de classe devem ser mutuamente exclusivos, ou seja, um indivíduo não pode ser alocado em dois intervalos ao mesmo tempo e, tampouco podem existir indivíduos sem alocação. O quadro abaixo mostra a notação para os intervalos de classe.

Notação para os intervalos de classe

Notação	Significado	Características
166 – 174	Intervalo aberto nos limites inferior e superior.	Não contém os valores 166 e 174.
166 — 174	Intervalo fechado no limite inferior e aberto no limite superior.	Contém o valor 166, mas não contém o valor 174.
166 — 174	Intervalo aberto no limite inferior e fechado no limite superior.	Não contém o valor 166, mas contém o valor 174.
166 — 174	-Intervalo fechado nos limites inferior e superior.	Contém os valores 166 e 174.



Distribuição de Frequência



Amplitude do intervalo de classe (h)

Corresponde ao comprimento da classe, sendo definida como a diferença entre os limites superior e inferior. Pode ser calculada, com aproximação, pela fórmula:

$$h = \frac{At}{K}$$

Distribuição de Frequência



Frequência simples absoluta (f)

Para uma classe ou um valor individual, a frequência simples corresponde ao número de observações encontradas para essa classe ou esse valor.

Por exemplo: a frequência da quarta classe (193 – 201) na tabela é igual a 21. Note que a soma das frequências é igual ao número de valores observados (n).

$$\sum_{i=1}^k f_i = n$$

Distribuição de Frequência



Construindo uma distribuição de frequência

Uma distribuição de frequência caracteriza-se por apresentar os dados em uma forma relativamente compacta, fornecendo uma boa visão geral do fenômeno estudado, além de conter informações adequadas para muitos propósitos, mesmo que alguma dessas informações possa ser perdida.

Por exemplo, alguma informação que pode ser determinada a partir da análise dos dados brutos, pode não ser obtida da distribuição, tal como ocorre na tabela, onde a menor e a maior nota da turma não podem ser observadas, assim como não se pode saber quantos alunos obtiveram a nota 7,0.

De qualquer forma, é interessante notar que a distribuição de frequência mostra as informações de uma maneira mais fácil de manusear, e qualquer possível perda de informação é um preço justo a pagar.

Tabela – Exemplo de distribuição quantitativa

Classe	Frequência
0,0 - 0,9	2
1,0 - 3,9	8
4,0 - 6,9	16
7,0 - 8,9	10
9,0 - 10,0	4
Total	40

Distribuição de Frequência



Construindo uma distribuição de frequência

Assim sendo, a construção de uma tabela de distribuição de frequência numérica, ou seja, para dados contínuos, consiste, essencialmente, de quatro passos:

- a. **Determinação do número de classes (K)**
- b. **Determinação do intervalo de classe (Ic)**
- c. **Determinação dos limites de classe**
- d. **Contagem da frequência (f)**

Distribuição de Frequência



a. Determinação do número de classes (K)

É extremamente importante que a tabela da distribuição seja apresentada com um número adequado de classes, sendo o ideal de cinco a 15, porque, de outra maneira, pode ser difícil detectar algum padrão de tendência para os dados apresentados, pois se esse número for reduzido, os dados ficarão muito condensados, de tal modo que pouca informação poderá ser extraída da tabela.

Por outro lado, se forem utilizadas muitas classes, poderá ocorrer que algumas delas apresentem frequências nulas ou muito pequenas, resultando em uma distribuição irregular, que pode comprometer a interpretação do fenômeno estudado.

Distribuição de Frequência



a. Determinação do número de classes (K)

Para a determinação do número de classe, existem vários métodos, porém os mais tradicionalmente utilizados são:

- **Método tradicional:** $K = \sqrt{n}$, onde: K é o número de classes
 n é o número de observações
- **Fórmula de Sturges:** $K = 1 + 3,3 \times \log_{10} n$

Distribuição de Frequência



b. Determinação do intervalo de classe (Ic)

A determinação do intervalo de classe é calculada dividindo-se a **amplitude (At)** dos dados pelo número de classes. Se necessário, arredonde o valor encontrado para o próximo número inteiro.

$$Ic = \frac{At}{K}$$

Em uma distribuição de frequência, é conveniente que cada classe tenha o mesmo tamanho de intervalo, porém se pode utilizar o menor valor encontrado nos dados como o limite inferior da primeira classe, assim como o maior valor para o limite superior da última classe. Algumas vezes é mais conveniente escolher um valor um pouco menor que o valor mínimo da distribuição. O bom senso é a melhor solução.

Distribuição de Frequência



c. Determinação dos limites de classe

Cada classe tem um limite inferior e um superior, os quais deverão ser determinados. Como dito anteriormente, o limite inferior da primeira classe pode ser determinado pelo menor valor encontrado nos dados.

Para determinar os demais limites inferiores, adicione a amplitude do intervalo de classe ao limite inferior da classe precedente. Então encontre o limite superior da primeira classe e, de igual maneira, determine os limites superiores das demais classes.

Lembre-se: as classes não devem se sobrepor. Portanto, tenha certeza de que cada valor encontra-se dentro de uma única classe, assim como o menor e o maior valores estão dentro da distribuição, e que nenhum dos valores situa-se dentro de uma “fenda” entre duas classes consecutivas.



Distribuição de Frequência



d. Contagem da frequência (f)

Inicialmente, identifica-se a classe de cada um dos valores da série e, a seguir, contam-se quantos valores pertencem a cada classe, ou seja, determina-se a frequência de cada classe. Uma maneira prática é fazer marcas na linha da classe apropriada e, então, contá-las, como será mostrado tomando-se como base o exemplo anterior.

$$K = \sqrt{n} \rightarrow \sqrt{60} = 7,74 \therefore K = 8 \text{ (método tradicional)}$$

$$K = 1 + 3,3 \times \log_{10}(60) = 1 + 3,3 \times (1,778) = 6,867 \therefore K = 7 \text{ (fórmula de Sturges)}$$



Distribuição de Frequência



b. Determinação do intervalo de classe (Ic)

O menor valor é igual a 166, e o maior é igual a 235, então a amplitude é igual a 69. Dividindo-se a amplitude pelo número de classes (8), e arredondando-se o valor, encontramos um **Intervalo de classe** igual a 9.

$$Ic = \frac{\text{Amplitude}}{\text{Número de classes}} \rightarrow \frac{69}{8} \cong 8,62 \therefore Ic = 9$$

Distribuição de Frequência



Construindo uma distribuição de frequência

O menor valor encontrado, no caso, 166, corresponde ao limite inferior da primeira classe. Para encontrar o limite inferior das 7 classes subsequentes, adiciona-se o intervalo de classe calculado, ou seja, o valor 9, ao limite inferior da classe anterior. O limite superior da primeira classe é 174, que corresponde a 1 menos o limite inferior da segunda classe. Os limites superiores das demais classes são $174 + 9 = 183$, $183 + 9 = 192$, e assim, sucessivamente.

Classes (nº de mL)	Frequência (nº de alunos)
166 – 174	3
175 – 183	4
184 – 192	8
193 – 201	21
202 – 210	14
211 – 219	7
220 – 228	1
229 – 235	2
	$\Sigma f = 60$



Bioestatística

Elementos da Distribuição de Frequência

[Parte 2 de 2]

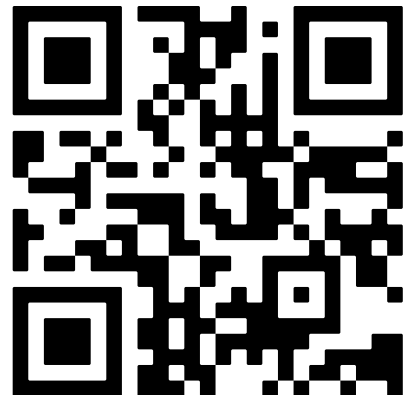
Prof. Me. Yuri Albuquerque

REFERÊNCIAS

- Fontelles, Mauro José. Bioestatística: aplicada à pesquisa experimental. São Paulo: Livraria da Física, vol 1, 2012.
- Vieira, Sonia. Introdução à Bioestatística. 6º Ed. Rio de Janeiro: GEN, 2021.

DOWNLOAD DO
CONTEÚDO DA AULA

<https://yurialb.github.io>



CONTATOS



E-mail: yuri.albuquerque@outlook.com

