

# AMOSTRAS, POPULAÇÕES E HISTOGRAMAS

## META

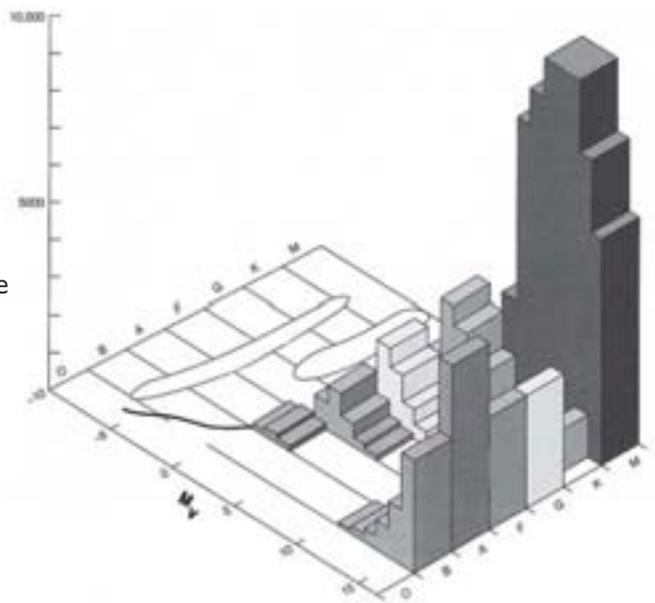
Apresentar as possibilidades de construção e interpretar histogramas.

## OBJETIVOS

Ao final desta aula, o aluno deverá: usar um problema para trabalhar os conceitos de média, desvio padrão e probabilidade a partir do histograma. apontar situações em que o uso de histogramas facilita e auxilia na interpretação dos dados; construir histogramas.

## PRÉ-REQUISITO

O aluno deverá conhecer os tipos de erros que causam flutuações em medidas experimentais.



Vimos na aula passada que qualquer medida está sujeita à ocorrência inevitável dos chamados erros aleatórios. Esses erros decorrem da impossibilidade de que um ser humano repita uma operação exatamente da mesma maneira, mesmo considerando que as repetições sejam feitas dentro do

## INTRODUÇÃO

maior rigor do método. Para este tipo de erro, há tratamentos estatísticos que nos permitem utilizar os dados obtidos, mesmo contendo os erros aleatórios, de modo que nossas medidas sejam confiáveis.

Nesta aula, trataremos de outra questão em que podemos recorrer à estatística para tornar nossa vida mais prática. Vamos imaginar que por alguma razão você precisa de informações sobre a altura das pessoas adultas do sexo feminino presentes em uma cidade. Será que você precisaria medir todas as pessoas de interesse? Existe alguma alternativa estatisticamente confiável para que se possa tratar do problema sem ter que realizar medidas da população completa?



Se quisermos ter a estatística exata, realmente teremos que medir a população completa. Porém, na inviabilidade disso, podemos trabalhar com estimativas obtidas a partir de um número menor de medições. Essas estimativas podem ser boas ou ruins, em termos da representatividade da população, dependendo de um fator crucial: a amostragem obrigatoriamente deve ser aleatória. Isto significa que a escolha dos indivíduos a serem medidos deve ser aleatória e, após cada medida, devem retornar ao conjunto inicial, caso contrário a população original está sendo alterada.

**OUTRA HISTÓRIA****CONCEITUALMENTE, POPULAÇÃO  
E AMOSTRA**

População - Qualquer coleção de indivíduos, finita ou infinita.

Amostra – Uma parte da população, normalmente selecionada com o objetivo de fazer inferências sobre a população.

Amostra representativa – Apresenta características relevantes da população na mesma proporção em que elas ocorrem na própria população.

Amostra aleatórias – Amostra de  $N$  valores ou indivíduos obtida de tal forma que todos os possíveis conjuntos de  $N$  valores na população tenham a mesma chance de ser escolhidos.

Qual a distribuição de alturas das pessoas do sexo feminino?

Solução medir todos, um por um? Não, pois isto é inviável.

Buscamos um método que possamos aplicar, mesmo que nossa população seja virtualmente infinita.

Usar uma abordagem estatística da questão? Sim...

Resultado das medições

Medir uma pessoa ao acaso: 1,6188 m

Escolhida uma segunda pessoa ao acaso: 1,7673 m .

Qual desses valores é a resposta que procuramos?

Nenhum dos dois, em princípio, pois como a altura varia de uma pessoa para outra, de nada adianta saber a altura de duas pessoas ao acaso, pois nos interessa saber a altura média das pessoas. Quanto mais numerosa esta amostra, mais bem sucedida será nossa empreitada. Isso vai ocorrer com qualquer tipo de investigação experimental, em que sempre uma população é o objetivo da coleta de dados, a fim de se chegar a conclusões sobre a população.

Como já dissemos, o principal requisito para isto é que toda a amostra tenha sido escolhida aleatoriamente, podendo ser sorteada novamente. Isto garante que a população não se altera durante a amostragem. Não precisamos levar em conta aqui a probabilidade de a mesma pessoa ser sorteada mais de uma vez.



(Fonte: <http://www.eletromil.net.tif>).

A tabela a seguir mostra as alturas de 140 pessoas medidas aleatoriamente.

---

**Altura de pessoas do sexo feminino (em metros).**


---

1,62	1,77	1,68	1,74	1,68	1,69	1,70
1,68	1,69	1,64	1,67	1,71	1,69	1,73
1,74	1,71	1,77	1,71	1,70	1,70	1,67
1,75	1,68	1,66	1,67	1,65	1,69	1,68
1,68	1,71	1,66	1,75	1,73	1,80	1,67
1,78	1,74	1,69	1,69	1,64	1,70	1,71
1,73	1,75	1,65	1,69	1,71	1,73	1,72
1,67	1,67	1,75	1,68	1,67	1,70	1,73
1,73	1,71	1,68	1,74	1,70	1,70	1,74
1,73	1,69	1,70	1,66	1,69	1,64	1,62
1,75	1,71	1,68	1,67	1,73	1,72	1,70
1,70	1,70	1,67	1,71	1,75	1,67	1,75
1,71	1,71	1,66	1,68	1,762	1,66	1,75
1,68	1,63	1,70	1,75	1,70	1,67	1,71
1,76	1,65	1,77	1,67	1,72	1,72	1,71
1,69	1,72	1,72	1,66	1,67	1,73	1,70
1,69	1,68	1,63	1,72	1,68	1,65	1,69
1,76	1,68	1,76	1,72	1,74	1,66	1,69
1,77	1,76	1,75	1,69	1,68	1,73	1,71
1,64	1,78	1,70	1,69	1,69	1,72	1,67

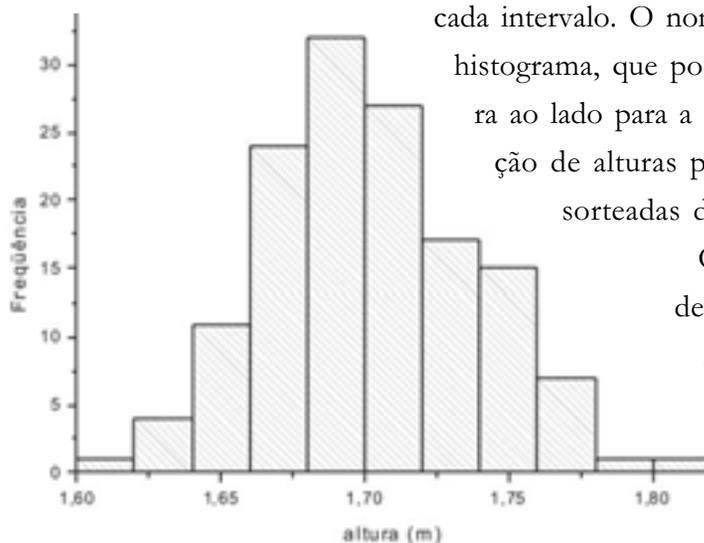
---

Percebemos que todas as alturas se encontram na faixa de 1,60 m até 1,82 m, porém, a ocorrência de valores não se distribui igualmente em toda a faixa possível. Se dividirmos a faixa toda em intervalos de 0,02 m e contarmos quantas pessoas tem altura em cada intervalo, encontraremos a chamada frequência de ocorrências.

**Distribuição de alturas de 140 pessoas escolhidas aleatoriamente:**

Intervalo (m)	Nº de ocorrências	Frequência
1,60 – 1,62	1	0,007
1,62 – 1,64	4	0,029
1,64 – 1,66	11	0,079
1,66 – 1,68	24	0,171
1,68 – 1,70	32	0,229
1,70 – 1,72	27	0,193
1,72 – 1,74	17	0,121
1,74 – 1,76	15	0,107
1,76 – 1,78	7	0,050
1,78 – 1,80	1	0,007
1,80 – 1,82	1	0,007
Total	140	1

Um conjunto de dados como este pode ser analisado mais facilmente por meio de um gráfico. No caso da distribuição de frequências, a representação gráfica tradicional associa a cada intervalo um retângulo, cuja base coincide com a largura do próprio intervalo e cuja altura representa a frequência associada a



cada intervalo. O nome deste tipo de gráfico é histograma, que pode ser observado na figura ao lado para a representação da distribuição de alturas para a amostra de pessoas sorteadas dentro da população.

Observe que o significado deste gráfico fica muito claro, mostrando que há poucas pessoas com as alturas menores e também com as maiores. Há um número maior de pesso-

as com alturas medianas, entre 1,68 m e 1,70 m. Nota-se, também, uma certa simetria na distribuição: a parte que fica à direita da região central é mais ou menos a imagem especular da parte que fica à esquerda.

O histograma representa graficamente todas as 140 alturas medidas para nossa amostra. Suas características básicas se referem à localização das observações em uma certa região do eixo horizontal e sua dispersão ao longo desta região. Seria impossível perceber isso olhando para a tabela de dado e este é um bom exemplo do que temos dito desde o começo, que a estatística, apesar de ser considerada uma ferramenta seca e repleta de parâmetros complicados, na realidade fornece elementos de praticidade.

## CONSTRUÇÃO DE UM HISTOGRAMA

Definir o intervalo, depois de olhar os dados.

Cada intervalo é representado por um retângulo.

A base do retângulo coincide com a largura do próprio intervalo.

A área é idêntica ou proporcional a sua frequência.

---

**C**omo conclusões desta aula, podemos destacar que é possível obter estimativas bastante confiáveis sobre uma população bastante numerosa, a partir de uma série de medidas feitas em um número restrito dentro dessa população, ou seja, uma amostra representativa dessa população.

## CONCLUSÃO

O principal requisito nesta tarefa é que a amostragem tenha sido absolutamente aleatória para não alterar a população original. De posse de dados sobre a amostra em questão, tais dados podem ser utilizados para a construção de um histograma, que é um gráfico de blocos que permite observar a frequência com que cada intervalo de valores possíveis ocorre na amostra.

## RESUMO



Ao estudar os tipos de erros, ficamos convencidos da impossibilidade de realizar medidas experimentais que forneçam valores idênticos, mesmo que um suposto valor “real” pudesse ser esperado. Vamos supor que estejamos frente a uma situação em que realmente são esperados valores diferentes e com variação aleatória, como por exemplo as alturas das pessoas presentes em um recinto. Os valores das alturas são tão aleatórios quanto a ocorrência de erros sistemáticos na medida de uma propriedade em que o tal suposto valor “real” seja esperado. Podemos tirar conclusões sobre as medidas das alturas que também serão válidas para o outro caso.

O conjunto de todas as alturas existentes no referido recinto é conhecido como população. Imagine que o recinto seja um estádio esportivo onde cabem milhares de pessoas. Neste caso, uma pergunta crucial é: será que seremos obrigados a medir todos para que as conclusões sejam confiáveis? Felizmente não. Basta medir uma fração desta população, de modo que quanto maior esta fração, mais nossa estimativa de valores se aproxima do valor referente à população. Um requisito é que esta amostra seja representativa, ou seja, que apresente características relevantes da população na mesma proporção em que elas ocorrem na própria população. O que garante isso é que cada observação seja aleatória, isto é, escolhida ao acaso e depois retorne ao conjunto original, para que cada observação saia da mesma população.

Após a realização das medidas da amostra, vemos que os valores em geral flutuam, mas as dificuldades são imensas quando tentamos entendê-los olhando para uma lista. Um conjunto de dados como este pode ser analisado mais facilmente por meio de um gráfico. No caso da distribuição de frequências, a representação gráfica tradicional associa a cada intervalo um retângulo, cuja base coincide com a largura do próprio intervalo e cuja altura representa a frequência associada a cada intervalo: é o chamado histograma.

## QUAL A UTILIDADE DE SE MEDIR A ALTURA DAS PESSOAS?

Segundo a página do IBGE na Internet: [http://www.ibge.gov.br/home/estatistica/indicadores/sipd/segundo\\_apres\\_pof07.shtm](http://www.ibge.gov.br/home/estatistica/indicadores/sipd/segundo_apres_pof07.shtm) (acessada em 07/02/2008), as medidas do peso e altura das pessoas são conhecidas como medidas antropométricas. Estes dados são considerados relevantes para a pesquisa de orçamentos familiares, tendo como objetivo “Investigar os orçamentos familiares combinados com outras informações sobre as condições de vida das famílias brasileiras, com destaque para antropometria e estudo sobre consumo alimentar Efetivo”. Para este estudo, os dados foram coletados em 65.000 domicílios no Brasil, incluindo áreas rurais e urbanas.

---

## PRÓXIMA AULA



Em nossa próxima aula, vamos aprender que a maneira como os valores observados para alguma medida, ao longo de uma amostra, pode ser classificada em tipos de distribuição, dos quais o mais comum é a chamada distribuição normal.

## REFERÊNCIAS

BARROS NETO, B.; SCARMINIO, I. E.; BRUNS, R. E. **Planejamento e otimização de experimentos**. Campinas: Editora da Unicamp, 1995.

BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. **Statistics for experimenters. An introduction to design, data analysis and model building**. New York: Wiley, 1978.

BUSSAB, W. O.; MORETIN, P. A. **Estatística básica**. São Paulo: Atual, 1985.