DISTRIBUIÇÃO DE FREQUÊNCIA

É o tipo de série estatística na qual permanece constante o fato, o local e a época. Os dados são colocados em classes preestabelecidas, registrando a frequência de ocorrência. Uma distribuição de frequência pode ser dividida em dois tipos:

Tipo A ou Tipo I

Usada para variáveis qualitativas ou então quantitativas discretas com poucos valores diferentes.

Ex.: notas do aluno X na disciplina de estatística

nº de acidentes na cidade de Santa Maria;

Quantidade de livros de programação na biblioteca da UNIFRA.

Notas do Aluno "X" na Disciplina de Estatística segundo critérios de avaliação do DE da UFSM – 1990

X_i	f_i
6.3	2
8.4	3
5.3	2
9.5	3
6.5	5
Σ	15

Fonte: Departamento de Estatística (1990)

Onde:

X_i ____

f_i ____.

Tipo B ou Tipo II: Usada para variáveis quantitativas contínuas ou discretas com muitos valores diferentes. Não é possível enumerar todos os valores.

Exemplo de distribuição de frequências para variável contínua (Tipo B)

X = Notas finais de 50 estudantes

22	46	9	40	57	22	22	13	50	42
35	2	15	41	34	52	32	75	69	44
26	42	60	56	30	3	17	79	45	37
0	12	62	50	45	41	59	11	66	39
43	33	70	50	47	20	36	40	67	29

A distribuição de frequência será expressa pela tabela:

Notas	f_i
0 ⊢— 10	4
10 ⊢— 20	5
20 ⊢— 30	6
30 ⊢— 40	8
40 ⊢— 50	12
50 ⊢— 60	7
60 ⊢— 70	5
70 ⊢— 80	3
Total	50

Para explicar a colocação das notas dos alunos, segundo uma distribuição do tipo B, necessitaremos de algumas definições. Assim:

1) Dados Brutos: Aqueles que não foram numericamente organizados, como é o caso das 50 notas dos alunos.

2) Rol: é o arranjo dos dados em ordem de grandeza crescente ou decrescente. Portanto teríamos

0	2	3	9	11	12	13	15	17	20
22	22	22	26	29	30	32	33	34	35
36	37	39	40	40	41	41	42	42	43
44	45	45	46	47	50	50	50	52	56
57	59	60	62	66	67	69	70	75	79

3) Limites da Classe: são os números extremos de cada classe, sendo assim, temos um limite inferior e um superior.

Ex.: Se o 1^0 intervalo for notas de 0 até 10, o 0 será o limite inferior enquanto que o 10 será o limite superior desta classe.

4) Intervalo de Classe: Existem várias maneiras de apresentarmos o intervalo de classes iguais ou diferentes entre si. Porém, sempre que possível, deveremos optar por intervalos iguais, o que facilitará os cálculos posteriores. Mas mesmo com intervalos iguais, as distribuições poderão apresentar-se da seguinte forma:

0 — 10: compreende todos os valores entre 0 e 10, exclusive os extremos.

0 ⊢⊣ 10: compreende todos os valores entre 0 e 10, inclusive os extremos.

0 —⊣ 10: compreende todos os valores entre 0 e 10, inclusive o 10 e exclusive o 0.

0 ⊢— 10: compreende todos os valores entre 0 e 10, inclusive o 0 e exclusive o 10.

Como optamos por este último tipo (0 α 10), poderemos definir como intervalo de classe a diferença entre o limite superior e o limite inferior da classe. Portanto, no exemplo, 10-0=10 é o intervalo ou amplitude da classe.

5) Amplitude Total ou "Range": é a diferença entre o maior e o menor dado. Em nosso caso, a nota maior é 7,9 e a menor é 0, logo nossa amplitude é 79-0=79

OBS.:____

____

6) Ponto Médio das Classes (X_i): é a média aritmética entre o limite superior e o limite inferior da classe. Assim, se a classe for 0 α 10, teremos $\frac{0+10}{2}=5$, que será o ponto médio da classe.

7) Número de Classes (K): quantas classes serão necessárias para representar o fato? Existem vários critérios que podem ser utilizados a fim de possuirmos uma ideia do melhor número de classes, porém tais critérios servirão apenas como indicação e nunca como regra fixa, pois caberá sempre ao pesquisador estabelecer o melhor número, levando-se em conta o intervalo de classe e a facilidade para os posteriores cálculos numéricos. Assim, podemos citar o critério que leva em consideração:

a)

n^0 de elementos	número de classes (k)	
	mínimo	máximo
até 50	5	10
51 a 100	8	16
101 a 200	10	20
201 a 300	12	24
301 a 500	15	30
mais de 500	20	40

b) Fórmula de Sturges: k = número de classes = 1+3,3.log n onde n = número de elementos observados

Usando a tabela, em nosso exemplo de notas, segundo o número de elementos, poderíamos utilizar de 5 a 10 classes, pois possuímos 50 notas. E segundo a fórmula de Sturges, teríamos

K = 1 + 3,3 log n
K = 1 + 3,3 log 50
K = 1 + 3,3(1,6989)
K = 1 + 5,6 = 6,6 ou arredondando: 7 classes.

Existe ainda uma terceira forma que pode ser utilizada para descobrir o número de classes: $k=\sqrt{N}$

Assim, $k=\sqrt{N}=\sqrt{50}\cong 7$

8) Amplitude ou Intervalo de Classes (h):

$$h=\frac{\text{amplitude total}}{\text{número de classes}}$$

Portanto, a distribuição de frequências do tipo B, para as notas dos alunos, ficaria:

Notas	f_i
0 ⊢ 10	4
10 ⊢ 20	5
20 ⊢ 30	6
30 ⊢ 40	8
40 ⊢ 50	12
0 ⊢ 60	7
60 ⊢ 70	5
70 ⊢ 80	3
	50

9) Frequência absoluta (*fi*): É o número de indivíduos por classe. Deve-se cuidar a contagem dos indivíduos nas classes, em função do tipo de intervalo utilizado.

10)Frequência Relativa da Classe (f_{ri}): corresponde ao quociente entre a frequência absoluta da classe e o total de elementos.

No exemplo, a frequência relativa da 7ª classe é: $f_{ri} = \frac{f_i}{N} = \frac{5}{50} = 0{,}1$

11) Frequência Acumulada (F_i): corresponde à soma de frequências de determinada classe com as anteriores. No exemplo, a frequência acumulada crescente da 4ª classe será:

$f_1 + f_2 + f_3 + f_4 = 4 + 5 + 6 + 8 = 23$.

No exemplo, a frequência acumulada decrescente da 4ª classe será:

$f_4 + f_5 + f_6 + f_7 + f_8 = 8 + 12 + 7 + 5 + 3 = 35$.

Resumindo, teríamos:

Notas	f_i (freq.. absoluta)	X_i (ponto médio)	FAC (frequência acumulada crescente)	FAD (frequência acumulada decrescente)	f_{ri} (frequência relativa de classe)	f_{rAC} (frequência relativa acumulada)	f percentual simples
0 α 10	4	0+10/2=5	0+4=4	50	4/50=0.08	0,08	8%
10 α 20	5	10+20/2=15	4+5=9	46-5=41	5/50=0,1	0,08+0,1=0,18	10%
20 α 30	6	20+30/2=25	9+6=15	41-6=35	6/50=0,12	0,18+0,12=0,3	12%
30 α 40	8	30+40/2=35	15+8=23	35-8=27	8/50=0,16	0,30+0,16=0,46	16%
40 α 50	12	40+50/2=45	23+12=35	27-12=15	12/50=0,24	0,46+0,24=0,70	24%
50 α 60	7	50+60/2=55	35+7=42	15-7=8	7/50=0,14	0,70+0,14=0,84	14%
60 α 70	5	60+70/2=65	42+5=47	8-5=3	5/50=0,1	0,84+0,10=0,94	10%
70 α 80	3	70+80/2=75	47+3=50	3-3=0	3/50=0,05	0,94+0,05=1	6%
Total	50				1		100%

Exercícios:

1) Nos dados abaixo determinar a amplitude total e dispô-la em um rol:

17	34	56	21	67	43	21	36	31	19	40	16	32

2) Dada a seguinte distribuição:

Valores	2	4	6	8	10
fi	10	15	8	3	2

a) N:

b) Frequência relativa simples:

c) Frequência acumulada crescente:

d) Frequência acumulada decrescente:

3) Temos a seguinte distribuição de frequência acumulada crescente:

Valores	2	3	5	10	20	30
FAC	4	16	22	32	38	40

Determinar:
a) Distribuição de frequência simples:
b) Distribuição de frequência decrescente:

4) As notas de 32 estudantes de uma classe estão descritos na tabela abaixo:

6,0	0,0	2,0	6,5	5,0	3,5	4,0	7,0
8,0	7,0	8,5	6,0	4,5	0,0	6,5	6,0
2,0	5,0	5,5	5,0	7,0	1,5	5,0	5,0
4,0	4,5	4,0	1,0	5,5	3,5	2,5	4,5

Com referência a essa tabela, determinar:
a) O Rol:
b) A distribuição de frequência do tipo B (sugestão: iniciar por "0" e intervalo de classe de 1,5)
c)O maior e o menor grau:
d) A amplitude total:
e) Qual a porcentagem dos alunos que tiraram nota menor do que 4?
f) Qual a porcentagem dos alunos que tiraram nota entre 5 e 7 inclusive?
g) Qual o limite superior da 2ª classe?
h) Qual o limite inferior da 4ª classe?
i) Qual o ponto médio da 3ª classe?
j) Colocar a frequência acumulada crescente e decrescente:

5) Os pesos dos 40 alunos de uma classe estão abaixo descritos:

69	57	72	54	93	68	72	58	64	62
65	76	60	49	74	59	66	83	70	45
60	81	71	67	63	64	53	73	81	50
67	68	53	75	65	58	80	60	63	53

Determinar a distribuição do tipo B (sugestão: iniciar com 45):

6)(ENADE) A taxa de evaporação de água em um reservatório depende da condição climática. Em um modelo simplificado, essa taxa E, pode ser escrita por $E = \alpha v(100 - UR)$ em que α é uma constante, v é a velocidade o vento em m/s e UR é a umidade relativa do ar, em porcentagem. Nas figuras I e II abaixo, são apresentados dados climáticos em determinado reservatório de água, em 12 semanas de observação.

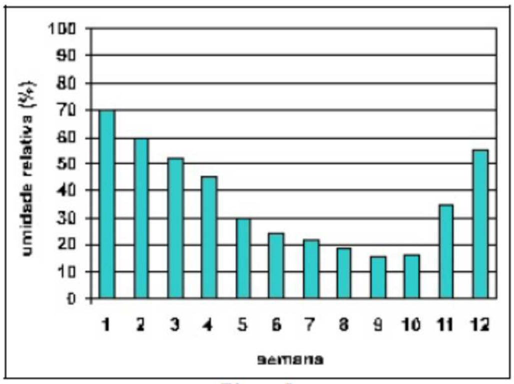


Figura I

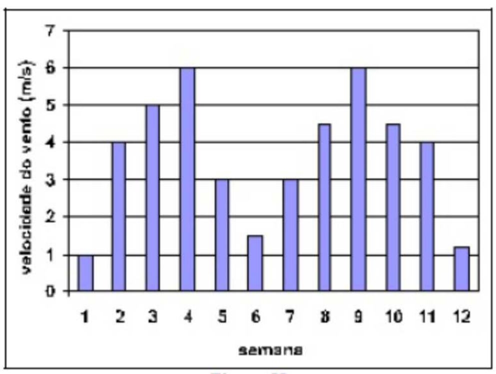


Figura II

As informações acima permitem concluir que a taxa de evaporação de água no reservatório, nas 12 semanas observadas, foi maior na semana:

a) 1 b)4 c)6 d)9 e)12

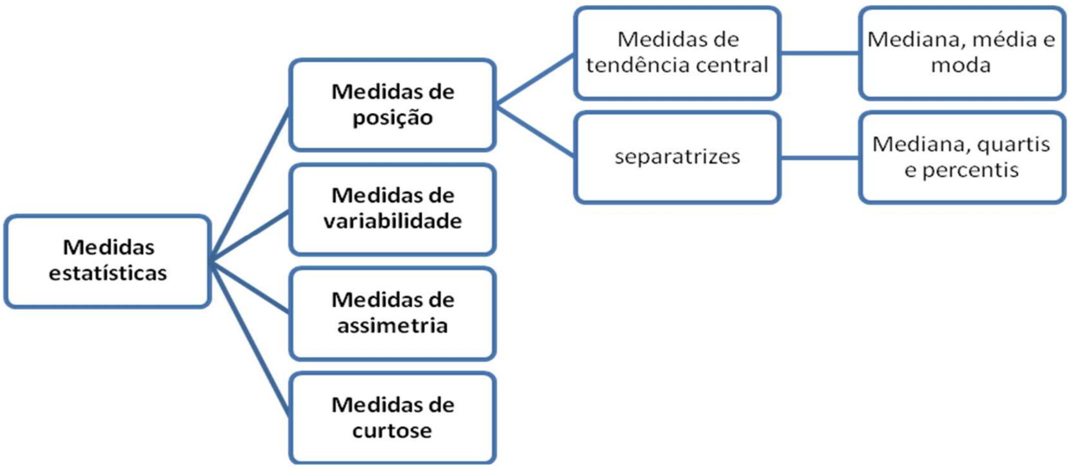
Para estimar a taxa de evaporação de água no reservatório, na 24ª semana, considere que a umidade do ar seja aproximada pelo valor médio dos dados da figura I e que a velocidade do vento seja aproximada por uma função peruódica com período igual a 6 semanas, obtida a partir dos dados da figura II. Qual das opções abaixo melhor estima essa taxa na 24ª semana?

a) 3α b) 80α c) 210α d) 480α e) 1.080α

MEDIDAS DESCRITIVAS:

O estudo que fizemos sobre distribuições de frequência, até agora, permite-nos descrever, de modo geral, os grupos dos valores que uma variável pode assumir. Dessa forma, podemos localizar a maior concentração de valores de uma dada distribuição, isto é, se ela se localiza no início, no meio ou no final, ou, ainda, se há uma distribuição por igual.

Porém, para ressaltar as tendências características de cada distribuição, isoladamente, ou em confronto com outras, necessitamos introduzir conceitos que se expressem através de números, que nos permitam traduzir essas tendências. Esses conceitos são denominados elementos típicos da distribuição e são as:



Dentre os elementos típicos, destacamos, as medidas de posição — estatísticas que representam uma série de dados orientando-nos quanto à posição da distribuição em relação ao eixo horizontal (eixo das abscissas).

As medidas de posição mais importantes são as medidas de tendência central, que recebem tal denominação pelo fato de os dados observados tenderem, em geral, a se agrupar em torno dos valores centrais. Dentre as medidas de tendência central, destacamos: a média aritmética; a mediana e a moda.

As outras medidas de posição são as separatrizes, que englobam a própria mediana; os quartis e os percentis.

MEDIDAS DE POSIÇÃO:

Medidas de Tendência Central: Dentre as medidas de tendência central, destacamos: a média aritmética, a mediana e a moda.

1.MÉDIA ARITMÉTICA: ($\bar{x}$) amostral e (μ) populacional

É o quociente da divisão da soma dos valores da variável pelo número deles:

$$\bar{x} = \frac{\sum x_i}{n} \qquad \text{ou} \qquad \mu = \frac{\sum x_i}{N}$$

Onde: $\bar{x}$ é a média aritmética
X_i= os dados das variáveis
n = o número de valores

OBS.: Existem vários tipos de média (aritmética, ponderada, geométrica, harmônica, etc.), mas estudaremos apenas a média aritmética.

Dados não-agrupados

Quando desejamos conhecer a média dos dados não-agrupados, determinamos a média aritmética simples.

Exemplo:

Sabendo-se que o excesso de etanol em uma determinada reação química, conforme as observações realizadas durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 ml, temos uma média de quantos ml durante uma semana:

$$\bar{x} = \frac{10+14+13+15+16+18+12}{7} = \frac{98}{7} = 14 \qquad \text{Logo: } \bar{x} = 14 \text{ ml}$$

Às vezes, a média pode ser um número diferente de todos os da série de dados que ela representa. Esse será o número representativo dessa série de valores, embora não esteja nos dados originais. Neste caso, diz-se que a média não tem existência concreta.

Exemplo: Tabela 1 – Volumes respiratórios forçados em um segundo para 10 adolescentes que respiram uma substância tóxica.

Indivíduo	*FEV(litros)*
1	2,30
2	2,15
3	3,50
4	2,60
5	2,75
6	2,82
7	4,05
8	2,25
9	2,68
10	3,00
$\sum =$	

R: 2,81 litros

Dados agrupados simples

Consideremos a distribuição relativa a 34 famílias de 4 filhos, tomando para variável o número de filhos do sexo masculino:

Tabela 2 – Número de filhos do sexo masculino de 34 famílias moradoras na zona rural de Santa Maria/2009

Número de meninos	f_i
0	2
1	6
2	10
3	12
4	4
	$\sum$ **= 34**

Neste caso, como as frequências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a média aritmética ponderada, dada pela fórmula:

$$\bar{x} = \frac{\sum x_i . f_i}{\sum f_i}$$

Um modo prático de obtenção da média ponderada é abrir, na tabela, uma coluna correspondente aos produtos $x_i.f_i$.

Nº de meninos	f_i	$x_i . f_i$
0	2	
1	6	
2	10	
3	12	
4	4	
	$\sum$ = 34	**$\sum$ =**

OBS: O valor médio obtido acima de 2,3 meninos sugere, neste caso, que o maior número de famílias tem 2 meninos e 2 meninas, sendo, porém, a tendência geral uma leve superioridade numérica em relação ao número de meninos.

Com intervalos de classes

Neste caso, convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio, e determinamos a sua média aritmética ponderada por meio da fórmula:

$$\bar{x} = \frac{\sum X_i . f_i}{\sum f_i}$$ onde X_i é o ponto médio das classes

Tabela 3 – Altura de 40 alunos da escola X – Santa Maria - 2007

i	*Estaturas (cm)*	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\sum$ = 40

Pela mesma razão do caso anterior, vamos, inicialmente, abrir uma coluna para os pontos médios e outra para os produtos $x_i . f_i$

i	**Estaturas (cm)**	**f_i**	**x_i(ponto médio)**	**$x_i . f_i$**
1	150 ⊢ 154	4		
2	154 ⊢ 158	9		
3	158 ⊢ 162	11		
4	162 ⊢ 166	8		
5	166 ⊢ 170	5		
6	170 ⊢ 174	3		
		$\sum$ = 40		**$\sum$=**

Tabela 3 – Frequência absoluta de níveis de dióxido de carbono para homens de Santa Maria com idades entre 25 e 34 anos.

Nível de colesterol (mg/100ml)	f_i	x_i	$x_i \cdot f_i$
80 ⊢ 120	13		
120 ⊢ 160	150		
160 ⊢ 200	442		
200 ⊢ 240	299		
240 ⊢ 280	115		
280 ⊢ 320	34		
320 ⊢ 360	9		
360 ⊢ 400	5		
	$\sum$=1067		$\sum$=

Vantagens e desvantagens da média aritmética
Por ser muito influenciada por valores extremos da série, não representa bem as distribuições em que estes valores ocorrem com frequência acentuada, como, por exemplo, a série cujos elementos são os seguintes: 18, 20, 22, 24 e 850 (onde a média aritmética é igual a 186,8, resultado que foi muito influenciado pelo elemento 850).

1) Apesar de a média aritmética situar-se entre o menor e o maior resultado da distribuição de frequências, ela não tem, necessariamente, a existência real. Podemos obter, por exemplo, uma média do tamanho de família de 4,5 pessoas, que é um valor inexistente.

2) Pode ser calculada para distribuições com classes, mas os seus resultados não são considerados reais.

3) Pode ser calculada diretamente usando qualquer calculadora eletrônica.

4) Depende de todos os valores da distribuição.

5) Evidencia bastante estabilidade de amostra para amostra, ou seja, se pesquisarmos numerosas amostras extraídas de uma mesma população, os valores das médias obtidas variam pouco (pouca variabilidade com amostras da mesma população)

3. Moda(M_0)

Denominamos moda o valor que ocorre com maior frequência em uma série de valores.

Dados não-agrupados

Quando lidamos com valores não-agrupados, a moda é facilmente reconhecida: basta procurar o valor que mais se repete.

A série de dados: *7, 8, 9, 10, 10, 10, 11, 12, 13, 15* tem moda igual a 10.

Podemos, entretanto, encontrar séries nas quais não exista valor modal, isto é, nas quais nenhum valor apareça mais vezes que outros.

É o caso da série: *3, 5, 8, 10, 12, 13,* que não apresenta moda (amodal).

Em outros casos, ao contrário, pode haver dois ou mais valores de concentração. Dizemos, então, que a série tem dois ou mais valores modais.

Na série: *2, 3, 4, 4, 4, 5, 6, 7, 7, 7, 8, 9* temos duas modas: 4 e 7 (bimodal).

A moda é utilizada:

- *quando desejamos obter uma medida rápida e aproximada de posição;*
- *quando a medida de posição deve ser o valor mais típico da distribuição.*

Dados agrupados COM intervalo de classe: A classe que apresenta a maior frequência é chamada classe modal. Pela definição, podemos afirmar que a moda, neste caso, é o valor dominante que está compreendido entre os limites da classe modal. O método mais simples para o cálculo consiste em tomar o ponto médio da classe modal e chamarmos esse valor de MODA BRUTA.

Para calcular a moda utilizamos a fórmula de Czuber:

$$M_o = L_{\inf} + \frac{\Delta_1}{\Delta_1 + \Delta_2}.h$$

Onde:

- L_{inf} = limite inferior da classe modal
- Δ_1= diferença entre a frequência da classe modal e a imediatamente anterior
- Δ_2= diferença entre a frequência da classe modal e a imediatamente posterior.
- h= amplitude da classe.

Exemplo: Altura de 40 alunos:

Estaturas	fi
150 α 154	4
154 α 158	9
158 α 162	11
162 α 166	8
166 α 170	5
170 α 174	3
	Σ=40

4.Mediana

A mediana é outra medida de posição, definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Em outras palavras, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

Dada uma série de valores, como, por exemplo: 5, 13, 10, 2, 18, 15, 6, 16, 9, de acordo com a definição de mediana, o primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores:

2, 5, 6, 9, 10, 13, 15, 16, 18

Em seguida, tomamos aquele valor central que apresenta o mesmo número de elementos à direita e à esquerda. Em nosso exemplo, esse valor é o 10, já que, nessa série, há quatro elementos acima dele e quatro abaixo.

Temos, então: Md = 10 Se, porém, a série dada tiver um número par de termos, a mediana será, por definição, qualquer dos números compreendidos entre os dois valores centrais da série. Convencionou-se utilizar o ponto médio. Assim, a série de valores: *2, 6, 7, 10, 12, 13, 18, 21* tem para mediana a média aritmética entre 10 e 12. Md = $\frac{10+12}{2}=11$

Verificamos que, estando ordenados os valores de uma série e sendo n o número de elementos da série, o valor mediano será:

• o termo de ordem $\frac{n+1}{2}$, se n for ímpar; • a média aritmética dos termos de ordem $\frac{n}{2}$ e $\frac{n}{2}+1$, se n for par.

A mediana é utilizada:

- *quando desejamos obter o ponto que divide a distribuição em partes iguais;*
- *quando há valores extremos que afetam de uma maneira acentuada a média.*

Dados agrupados sem intervalo de classe:

Neste caso, é o bastante identificar a frequência acumulada imediatamente superior à metade da soma das frequências. A mediana será aquele valor da variável que corresponde a tal frequência acumulada.

Nº de meninos	fi	FAC
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	Σ=34	

$\frac{\sum f_i}{2}=\frac{34}{2}=17$. A menor frequência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável, este será o valor da mediana.

No caso de existir uma frequência acumulada, tal que $FAC=\frac{\sum f_i}{2}$, a mediana será dada por $M_d=\frac{x_i+x_{i+1}}{2}$. Isto é, a mediana será a média aritmética entre o valor da variável correspondente a essa frequência acumulada e o seguinte.

Exemplo:

Xi	fi	FAC
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8
	Σ=8	

Dados agrupados COM intervalo de classe:

Neste caso, o problema consiste em determinar o ponto do intervalo em que está compreendida a mediana. Para tanto, temos inicialmente que determinar a classe na qual se acham a mediana- classe mediana. Tal classe será, evidentemente, aquela correspondente a frequência acumulada imediatamente superior a $\frac{\sum f_i}{2}$. Considerando a tabela abaixo:

Altura de 40 alunos da escola X

Estaturas	fi	FAC
150 α 154	4	
154 α 158	9	
158 α 162	11	
162 α 166	8	
166 α 170	5	
170 α 174	3	
	Σ=40	

$$M_d = L_{\inf} + \frac{\left[\frac{\sum f_i}{2} - F(ant)\right].h^*}{f^*}$$

Onde:

- L_{inf}= limite inferior da classe
- F(ant) = frequência acumulada da classe anterior à classe mediana;
- f*= frequência simples da classe mediana
- h*= amplitude do intervalo da classe mediana

OBS: No cálculo da média, todos os valores da amostra são levados em conta, ao passo que no caso da mediana isto não acontece. Por esta razão, valores muito grandes ou muito pequenos, comparados aos demais valores da amostra, causam grandes variações na média, o que em geral não ocorre com a mediana. Por isso, dizemos que a mediana é robusta, isto é, ela é resistente a valores atípicos.

EXERCÍCIOS

01. A tabela abaixo lista as durações das terapias para dez pessoas inscritas em um estudo que investiga os efeitos da exposição a determinada substância química. Determine a média desses valores. (Resposta: 8,6 anos)

Tabela 1 – Durações das terapias para 10 pessoas expostas a uma determinada substância química

Indivíduo	***Duração***
1	12
2	11
3	12
4	6
5	11
6	11
7	8
8	5
9	5
10	5
$\sum =$	

02. Na sequência temos a massa (peso) em gramas, de ratos da raça Wistar com 30 dias de idade. (Fonte: Vieira, S., 1980). Calcule a média aritmética. R: 67

50	62	70	86	66	55
60	77	82	64	58	74

03. Os tempos de reação de um indivíduo a determinados estímulos foram medidos por um psicologista como sendo 0,53; 0,46; 0,50; 0,49; 0,52; 0,53; 0,44 e 0,55 segundos, respectivamente. Determinar: os tempos médio, modal e mediano de reação do indivíduo a esses estímulos. R: 0,50 / 0,53 / 0,51

04. Calcule a média dos números de dentes perdidos ou danificados em uma amostra de 50 pessoas tratadas em determinada clínica dentária (Fonte: Callegari- Jacques, S. 2003). R: 3,2 dentes

Tabela 3 – Número de dentes perdidos/danificados de pessoas tratadas em uma clínica dentária – Santa Maria/RS

Número de dentes (X)	Número de pessoas (Fi)	X . Fi
0	9	
1	5	
2	6	
3	7	
4	9	
5	5	
6	4	
7	3	
8	2	
Σ	50	

05. Calcule o número médio de dentes cariados, para cada sexo, a partir dos dados apresentados na tabela a seguir:

Tabela 4 – Número de dentes cariados das pessoas tratadas em uma clínica dentária – Santa Maria/RS

Nº de dentes cariados	Sexo	
	Masculino	Feminino
0	16	13
1	2	5
2	3	3
3	2	2
4	2	2

Fonte: MOREIRA et alli (1985)

R: Masculino: 0,88   Feminino: 1

06. Em uma classe de 50 alunos, as notas obtidas formaram a seguinte distribuição.

Notas	2	3	4	5	6	7	8	9	10
Número de alunos	1	3	6	10	13	8	5	3	1

Calcule a nota média, a nota mediana e a nota modal. R: 5,92 / 6 / 6

07. Quinze indivíduos foram sujeitos à recolha de urina em dois momentos, antes da toma de um diurético e após a toma desse diurético tendo-se obtido os seguintes valores em litros/dia:

Ind	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sem	1,2	1,2	1,2	1,2	1,1	1,3	1,8	1,2	1,1	1,4	1,1	1,3	1,1	1,2	1,3
Com	1,4	1,3	1,5	1,4	1,3	1,6	2,1	1,4	1,3	1,5	1,2	1,4	1,2	1,2	1,3

a) Determine as medidas de localização central da urina sem diurético.

R: Média: 1,25   Mediana: 1,2   Moda: 1,2

b) Determine as medidas de tendência central da urina com diurético.

R: Média: 1,41   Mediana: 1,4   Moda: 1,4

08. Foram analisadas amostras de sangue de 100 adultos tendo-se obtido os seguintes resultados para valores de haptoglobina (g/l)

1,51	1,51	1,51	1,51	1,52	1,52	1,52	1,52	1,52	1,52
1,52	1,53	1,53	1,53	1,54	1,54	1,54	1,54	1,54	1,54
1,55	1,55	1,55	1,55	1,56	1,56	1,56	1,56	1,57	1,57
1,58	1,59	1,59	1,59	1,59	1,59	1,59	1,59	1,59	1,59
1,60	1,60	1,61	1,61	1,61	1,61	1,62	1,62	1,62	1,63
1,63	1,63	1,63	1,63	1,63	1,63	1,64	1,64	1,64	1,64
1,65	1,65	1,65	1,66	1,67	1,67	1,68	1,68	1,68	1,69
1,69	1,70	1,70	1,70	1,71	1,72	1,73	1,73	1,74	1,75
1,76	1,77	1,78	1,78	1,79	1,80	1,81	1,81	1,81	1,82
1,83	1,83	1,83	1,83	1,83	1,83	1,83	1,84	1,84	1,84

a) Calcule a média, moda e mediana. R: Média: 1,65   Mediana: 1,63   Moda: 1,59

b) Agrupe em uma tabela e recalcule as medidas acima.

09. Num estudo sobre obesidade registrou-se o índice de massa corporal (IMC) em 50 doentes conforme segue abaixo:

25,6	26,2	26,9	27,5	28,1	28,7	29,4	32,0	34,8
25,8	26,3	26,9	27,6	28,2	28,8	29,4	32,0	34,9
25,9	26,4	27,0	27,7	28,3	29,0	29,9	32,0	
26,1	26,6	27,0	27,9	28,6	29,0	30,6	32,0	
26,1	26,7	27,3	28,0	28,7	29,2	31,5	34,1	
26,1	26,8	27,4	28,1	28,7	29,4	31,6	34,7	

a) Organize uma tabela com dados agrupados em classes.

b) Determine a média, moda e mediana dos dados acima. R: Média: 28,75 Mediana: 28,15 Moda: 32

10. Durante um derramamento de óleo no mar, recolheu-se um certo número de mortos, em 40 localidades distintos, obtendo-se os seguintes dados.

Mortos(número)	0	1	2	3	4	5	6	7
Pontos da baia	7	11	10	7	1	2	1	1

a) Calcule as medidas de posição central. Média: 1,98   Mediana: 2   Moda: 1

b) Calcule a porcentagem de localidades com pelo menos dois mortos. R: 55%

c) Calcule a porcentagem de localidades com no máximo 2 mortos. R: 70%

d) Calcule a porcentagem de localidades com no mínimo 3 mortos. R: 30%

11. A tabela mostra a composição por idade e sexo de um grupo de trabalhadores, do polo petroquímico, numa determinada cidade.

Idade(anos)	***Homem***	***Mulher***	***Total***
14 ⊢ 19	2	2	4
19 ⊢ 24	10	5	15
24 ⊢ 29	33	9	42
29 ⊢ 34	45	12	57
34 ⊢ 39	39	8	47
39 ⊢ 44	21	4	25
Total			

Pede-se:

Qual é a média de idade dos trabalhadores do sexo masculino e feminino . R: M: 35,9 a F: 33,7 a

12. O gerente do programa de assistência química de um determinado município brasileiro, desejando iniciar o planejamento de medicamentos para o programa de saúde mental, levantou os seguintes dados: a população do município, segundo dados do DATASUS relativos a 2003, é de 97.948 habitantes; há 1,5%

de prevalência das doenças mentais na área; o município conta com 4 unidades de saúde responsáveis pelo acompanhamento dos doentes de saúde mental, sendo que o número total de pacientes cadastrados no programa até o momento é de 1300 pessoas; 20% dos pacientes cadastrados utilizam o medicamento fenobarbital 100mg, 2 vezes ao dia; 40% utilizam o medicamento fenitoína 100mg, 3 vezes ao dia, associado à carbamazepina 200mg, 2 vezes ao dia; 20% utilizam o medicamento clonazepam 0,5 mg, 3 vezes ao dia; o restante utiliza o medicamento ácido valpróico 500mg, 2 vezes ao dia (SINAES, 2004). Numa proposta de programação dos medicamentos para um período de 12 meses, quais as quantidades de comprimidos dos medicamentos deste programa a serem adquiridas?

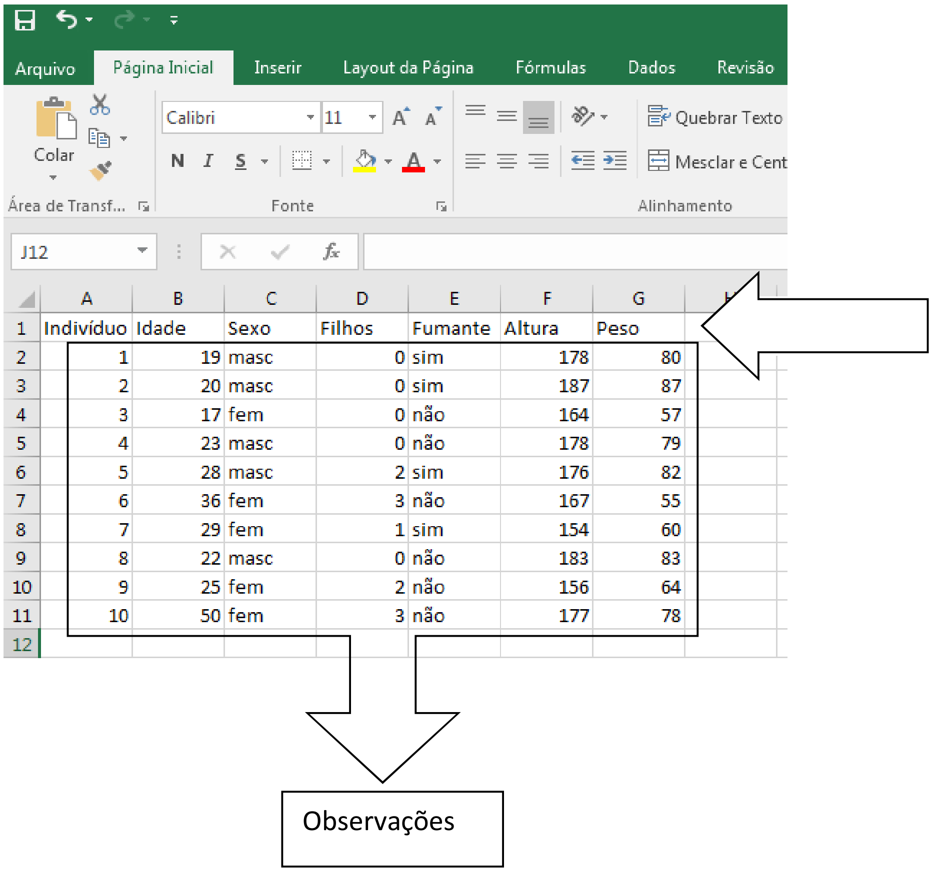
	fenobarbital	fenitoína	carbamazepina	clonazepam	ácido valpróico
a	15600	46800	31200	23400	15600
b	26000	46800	31200	23400	15600
c	26000	52000	26000	26000	26000
d	187200	561600	187800	280800	374400
e	187200	561600	374400	280800	187200

Resposta: letra e

ORGANIZAÇÃO DOS DADOS NO EXCEL

O Excel é uma planilha eletrônica, ou seja, uma tabela onde os dados das pesquisas são organizados a fim de serem melhor entendidos e analisados.

Normalmente colocamos as variáveis nas colunas e os indivíduos (observações), nas linhas.

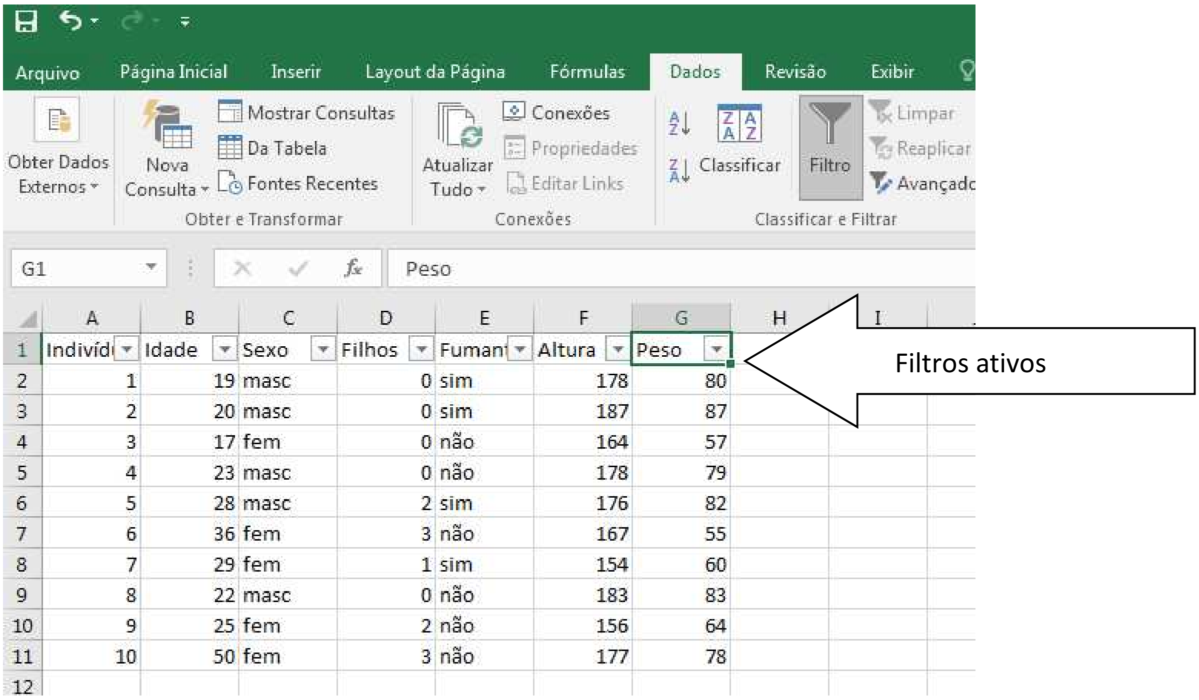


	A	B	C	D	E	F	G
1	Indivíduo	Idade	Sexo	Filhos	Fumante	Altura	Peso
2	1	19	masc	0	sim	178	80
3	2	20	masc	0	sim	187	87
4	3	17	fem	0	não	164	57
5	4	23	masc	0	não	178	79
6	5	28	masc	2	sim	176	82
7	6	36	fem	3	não	167	55
8	7	29	fem	1	sim	154	60
9	8	22	masc	0	não	183	83
10	9	25	fem	2	não	156	64
11	10	50	fem	3	não	177	78
12							

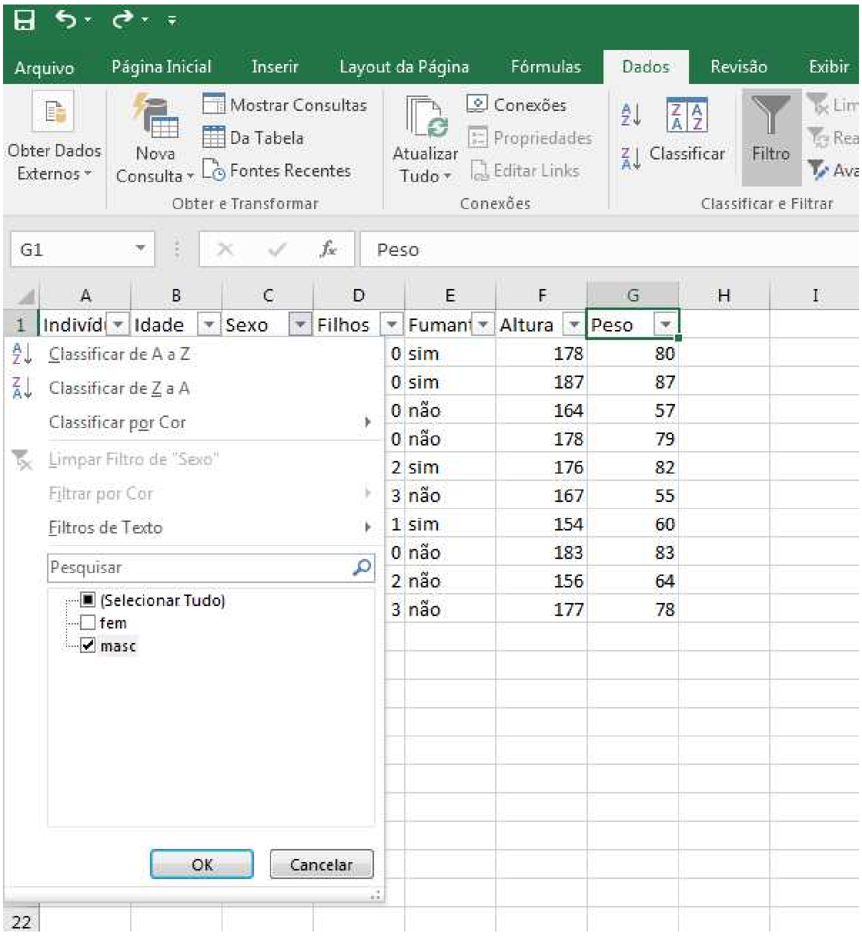
Ferramentas que podem ajudar na organização dos dados:

1) Filtro: permite que se apresente somente os dados desejados de uma ou mais variáveis. Para acessar a ferramenta Filtro, clique sobre qualquer uma das variáveis e em seguida selecione:

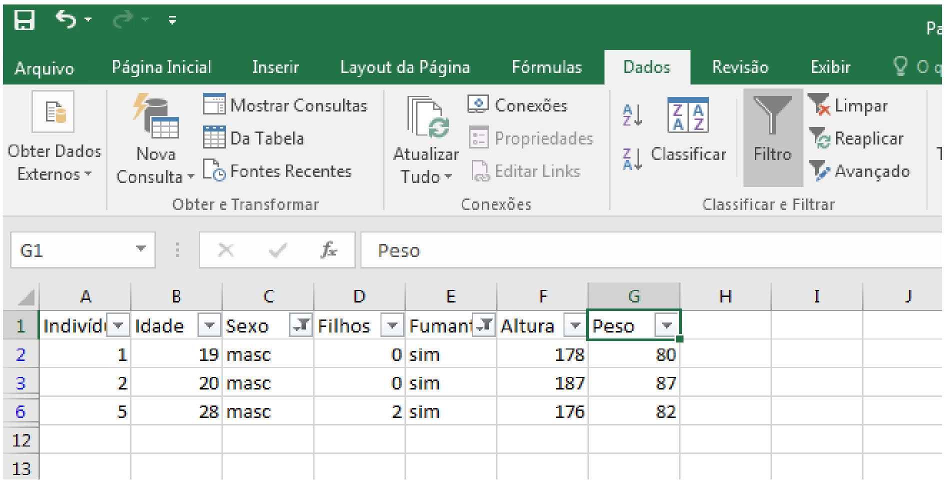
DADOS>Filtro



É possível que queiramos saber quantos homens (masc) são fumantes (sim). Para isto marcamos a opção "masc" no filtro da variável "sexo" e "sim no filtro da variável "Fumante"



Obtendo



Observe que somente os indivíduos 1,2 e 5 são do sexo masculino e fumantes.

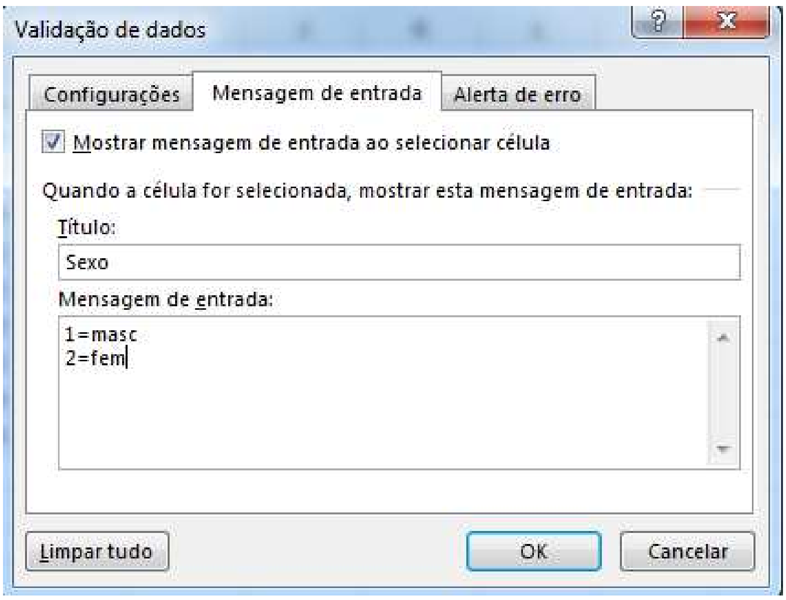
Para" limpar" o filtro e visualizar novamente a planilha original clique sobre o filtro ativo e selecione "Limpar Filtro de".

2) Etiqueta: É comum atribuirmos números aos dados nominais para facilitar o estudo. Podemos atribuir o número "1" para "sim" e o número "2" para "não" aos dados da variável "Fumante", por exemplo. Neste caso, é conveniente colocarmos uma etiqueta que não nos deixe esquecer dos valores atribuídos aos dados.

Primeiro clique sobre o nome da variável que se quer criar uma etiqueta. Em seguida selecione:

DADOS> Validação de Dados>Mensagem de Entrada

Escolha um Título (opcional) e defina os valores que serão atribuídos a cada observação.



Nosso padrão de tabela no Excel:

- Todas as células centralizadas
- O Título e o Rodapé devem englobar exatamente o número de colunas que constituem a Tabela e é preciso "mesclar e centralizar" e ainda "quebrar o texto automaticamente".
- Só são permitidas as seguintes linhas horizontais na Tabela:

  - Separando o Título do Cabeçalho.
  - Separando o Cabeçalho do Corpo.
  - Separando o Corpo do Rodapé.

- A expressão "Tabela 1 - " deve estar em negrito.
- A palavra "Fonte:" deve estar em negrito e itálico.
- A palavra "Total" em negrito é opcional.

Tabela 1- Refeições diárias de alguns alunos do Centro

Refeições	Frequência	F. Relativa
2	4	7%
3	15	25%
4	10	17%
5	21	35%
6	8	13%
7	2	3%
Total	60	

Fonte: Alunos do primeiro semestre de Engenharia Química

Como construir gráficos no Excel:

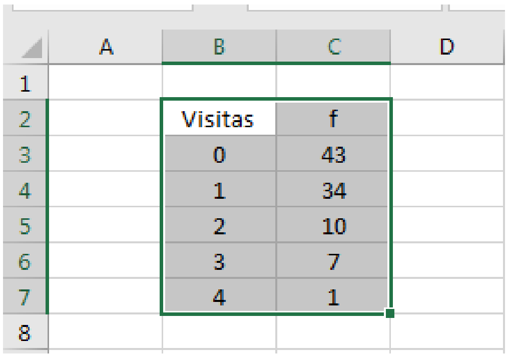
Gráficos de colunas: são comumente utilizados para representarem distribuições de frequências de agrupamentos simples. Cada coluna representa uma observação enquanto que a altura desta coluna representa a frequência desta observação.

Exemplo:

Número de visitas anuais a uma Metalúrgica de uma amostra de funcionários

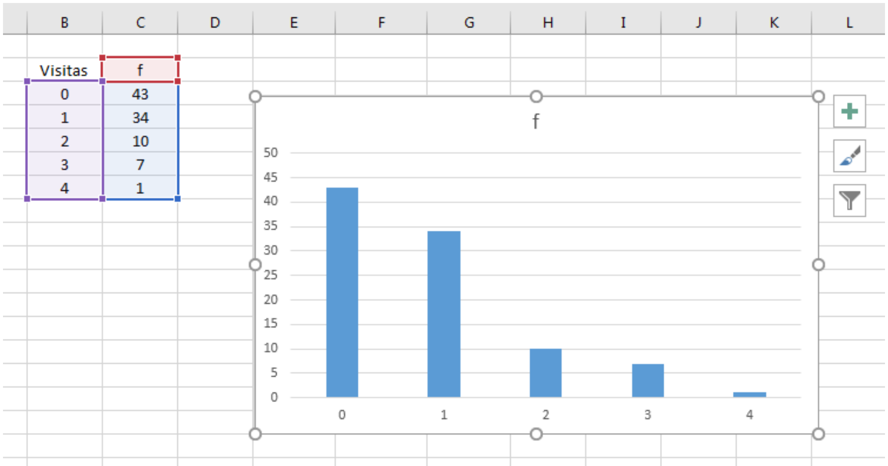
Visitas (Xi)	f_i
0	43
1	34
2	10
3	7
4	1

Passe os dados para o Excel tal como aparecem na tabela e marque-os. Não esqueça de marcar o cabeçalho também.



Visitas	f
0	43
1	34
2	10
3	7
4	1

Em seguida selecione a aba “INSERIR”, encontre “Gráficos Recomendados” e pressione sobre “Colunas Agrupadas”. O resultado será o seguinte gráfico de colunas:



Histogramas: composto por retângulos justapostos onde a base de cada um deles corresponde ao intervalo de classe e a sua altura à respectiva frequência.

Exemplo: Construa o histograma da distribuição de frequência abaixo:

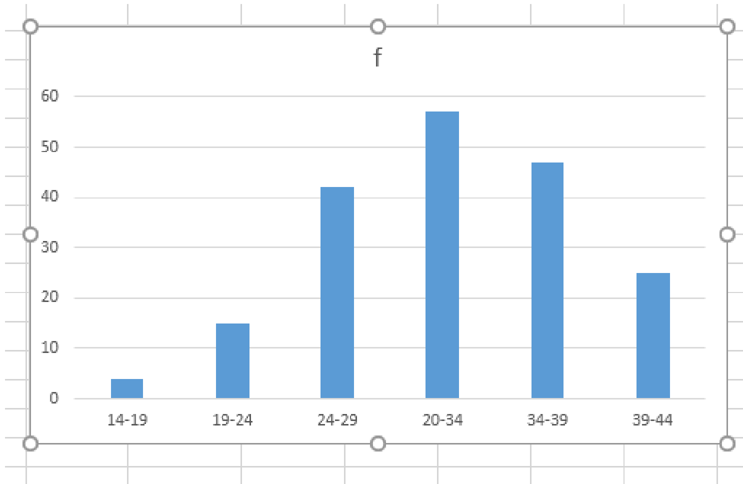
Idade de um grupo de trabalhadores do polo petroquímico

Idade (anos)	f
14↦ 19	4
19↦ 24	15
24↦ 29	42
29↦ 34	57
34↦ 39	47
39↦ 44	25

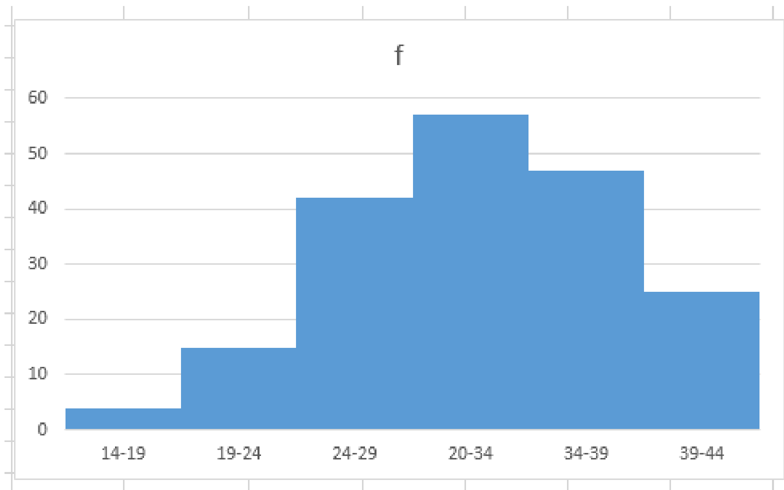
No Excel: copie a tabela para a planilha do Excel

Idade (anos)	f
14-19	4
19-24	15
24-29	42
20-34	57
34-39	47
39-44	25

Selecione toda a tabela e clique sobre inserir> colunas, você vai obter o seguinte gráfico



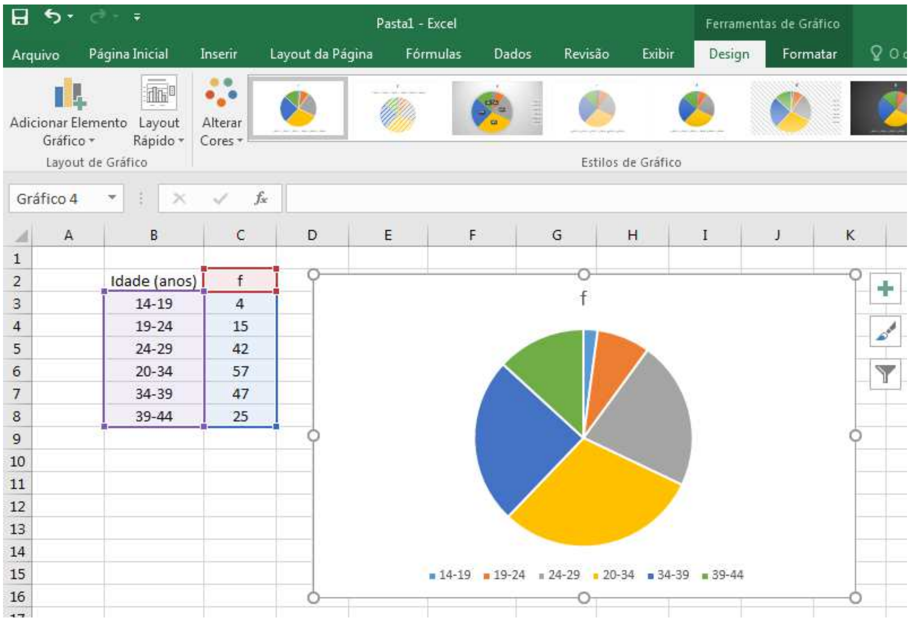
Agora, basta transformar este gráfico de colunas num histograma, para isto, clique com o botão direito sobre qualquer uma das colunas e selecione “Formatar Séries de Dados”. Na janela que se abre, defina “Largura do Espaçamento” como 0% e clique sobre “fechar”. Você obterá o histograma abaixo.



A construção de histogramas tem caráter preliminar em qualquer estudo e é importante indicador da distribuição de dados. Podem indicar se uma distribuição aproximam-se de uma função normal como pode indicar mistura de populações quando se apresentam bimodais.

Gráfico de Pizza: Um gráfico de setores apresenta uma circunferência onde as “fatias” tem tamanhos proporcionais às frequências da distribuição considerada. Para o exemplo das idades visto acima, um gráfico de setores ficaria da seguinte forma:

Selecionar tabela>inserir>pizza , vamos obter o seguinte gráfico:



Exercícios no Excel:

1) A tabela abaixo lista o tempo de duração das visitas em uma fábrica de produtos químicos de 10 funcionários visitantes. Determine a média desses valores.

Duração do tempo de visita em uma fábrica de produtos químicos para 10 funcionários visitantes.

Indivíduo	**Duração**
1	12
2	11
3	12
4	6
5	11
6	11
7	8
8	5
9	5
10	5
TOTAL	

Resp.:8,6

2) Na sequência temos a massa (peso) em gramas, de uma determinada substância. (Fonte: "Adaptado de" Vieira,S.,1980).Calcule a média aritmética.

50	62	70	86	66	55	60	77	82	64	58	74

Resp.: 67

3) O tempo de reação de uma determinada mistura de química a determinados estímulos foram medidos como sendo 0,53;0,46;0,50;0,49;0,52;0,53;0,44 e 0,55 segundos. Determinar o tempo médio, modal e mediano de reação a esses estímulos.

Resp: 0,50;0,53;0,51

4) Calcule a média do número de pontuação obtida em provas de avaliação sob três métodos de instrução (Fonte: "Adaptada de" Callegari-Jaques S. 2003) em uma amostra de 50 pessoas.

Pontuação (x)	Número de pessoas (fi)	x.fi
0	9	
1	5	
2	6	
3	7	
4	9	
5	5	
6	4	
7	3	
8	2	
TOTAL	50	

Resp.: 3,2

5) Durante um derramamento de óleo no mar, recolheu-se um certo número de mortos, em 40 localidades distintos, obtendo-se os seguintes dados.

Mortos(número)	0	1	2	3	4	5	6	7
Pontos da baia	7	11	10	7	1	2	1	1

a) Calcule as medidas de posição central. R. Média: 1,98 Mediana: 2 Moda: 1
b) Calcule a porcentagem de localidades com pelo menos dois mortos. R: 55%
c) Calcule a porcentagem de localidades com no máximo 2 mortos. R: 70%
d) Calcule a porcentagem de localidades com no mínimo 3 mortos. R: 30%

6) . A tabela mostra a composição por idade e sexo de um grupo de trabalhadores, do polo petroquímico, numa determinada cidade.

Idade(anos)	*Homem*	*Mulher*	*Total*
14 ⊢ 19	2	2	4
19 ⊢ 24	10	5	15
24 ⊢ 29	33	9	42
29 ⊢ 34	45	12	57
34 ⊢ 39	39	8	47
39 ⊢ 44	21	4	25
Total			

Pede-se:

Qual é a média de idade dos trabalhadores do sexo masculino e feminino . R: M: 35,9 a F: 33,7 a