



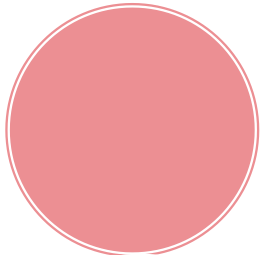
FLACSO
MÉXICO



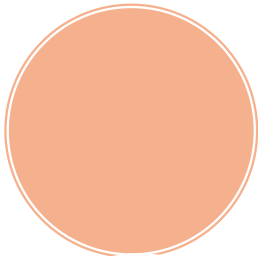
Construção de Bases de Dados e primeiros resultados

Mtra. Martha Estrada
FLACSO-México

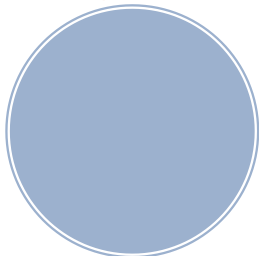
Do que se trata esta apresentação?



Revisar ideias sobre Bases de Dados de forma geral

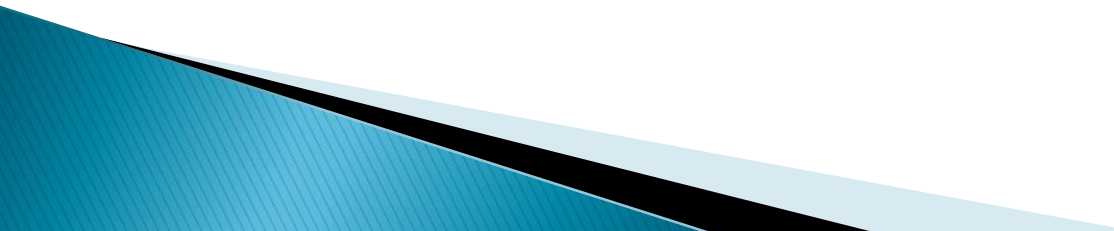


Revisar algunos ejemplos de trabajo en bases pequeñas

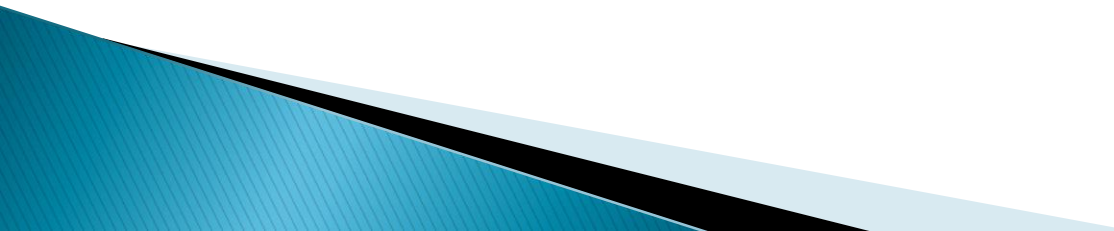


Exercício estatístico

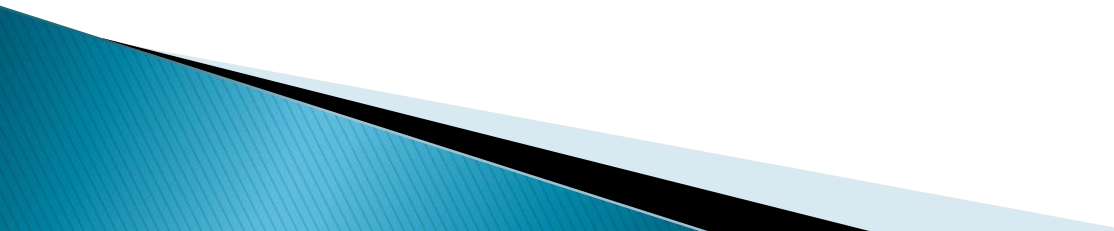
Para que serve uma BD?

- ▶ Conjunto de informação ordenada.
 - ▶ Armazenar a informação.
 - ▶ Recuperação de informação.
 - ▶ Se adaptam às necessidades específicas.
- 

Exemplos de BD

- ▶ Guias telefónicas.
 - ▶ Arquivos pessoais.
 - ▶ Bibliotecas públicas.
 - ▶ Registros de negócios financeiros.
 - ▶ Historia médica.
 - ▶ Coordenadas geográficas.
 - ▶ Nascimentos e óbitos em uma localidade.
 - ▶ Renda (remessas) e gastos de migrantes em território.
- 

A folha do Excel

- ▶ Em cada coluna colocamos uma variável (sexo, idade, renda, etc)
 - ▶ Em cada linha colocamos um sujeito (indivíduos, valores, etc.)
 - ▶ Cada célula se identifica com a letra da coluna e o número da linha. Por exemplo:
 - célula B2
 - célula C5
 - célula D1
 - célula E14
- 

Nuevo Abrir Guardar Imprimir Importar Copiar Pegar Formato Deshacer Rehacer Auto

Hojas Gráficos

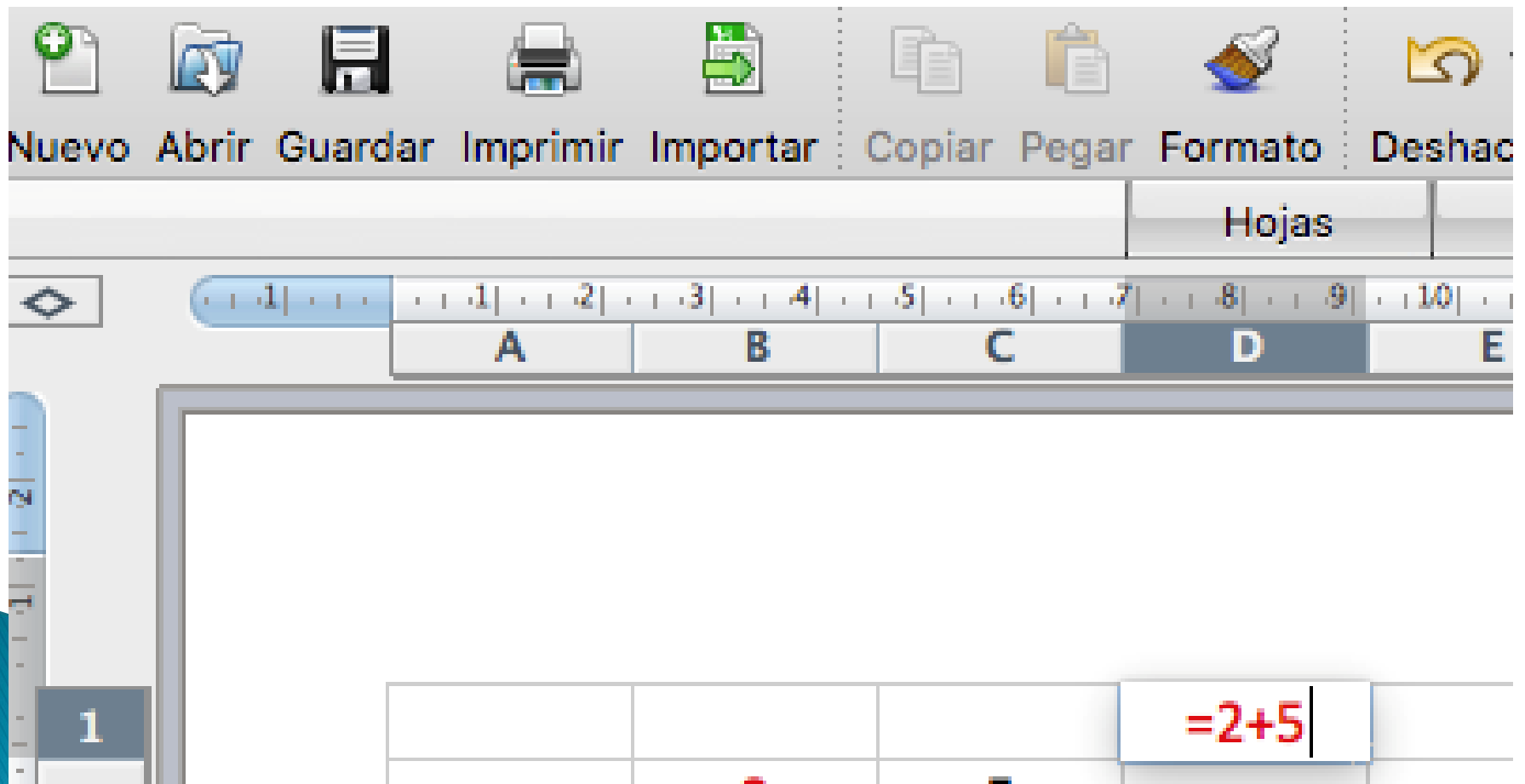
1 2 3 4 5 6 7 8 9 10 11 12 13 14

A B C D E F

1				D1	
2		B2			
3					
4					
5			C5		
6					
7					
8					
9					
10					
11					
12					
13					
14				E14	
15					
16					

Operações matemáticas

- ▶ Dar click com o cursor encima da célula onde vai escrever
- ▶ Sempre iniciar com o sinal de = (igual)
- ▶ Depois escrever a fórmula = 2 + 5 (enter)



Operações matemáticas

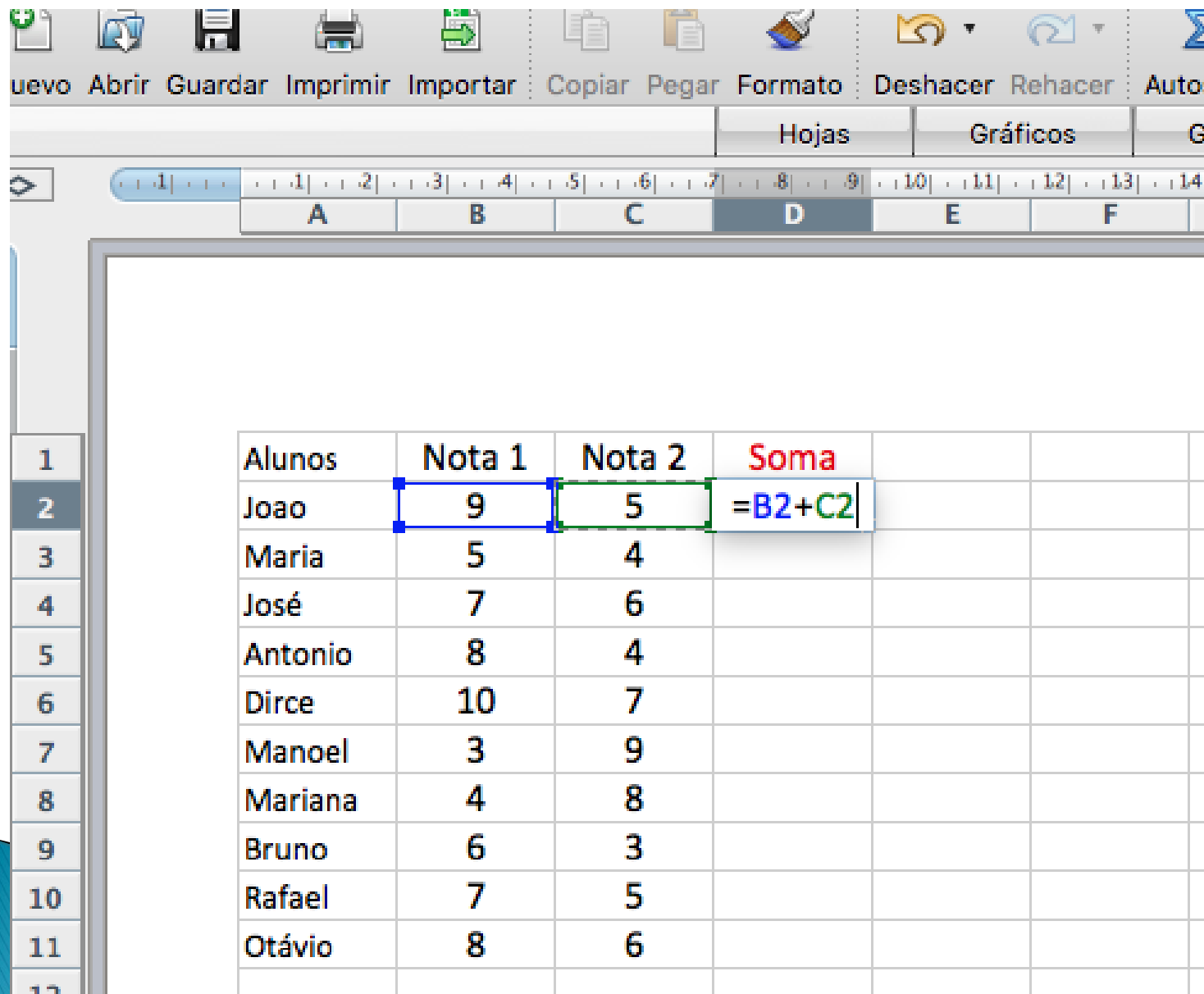
- ▶ Se já tiver os números nas células, dar clique com o cursor na célula onde você quer o resultado e escrever

$$= b2 + c2$$

The screenshot shows a spreadsheet application interface. The menu bar includes options: Nuevo, Abrir, Guardar, Imprimir, Importar, Copiar, Pegar, Formato, and Des. The sheet name is 'Hojas'. The grid shows columns A, B, C, and D, and rows 1, 2, 3, and 4. Cell B2 contains the number '2' and cell C2 contains the number '5'. Cell D2 is selected and contains the formula '=B2+C2'. The result '7' is displayed above cell D2. The formula bar shows the formula '=B2+C2'.

	A	B	C	D
1				7
2		2	5	=B2+C2
3				
4				

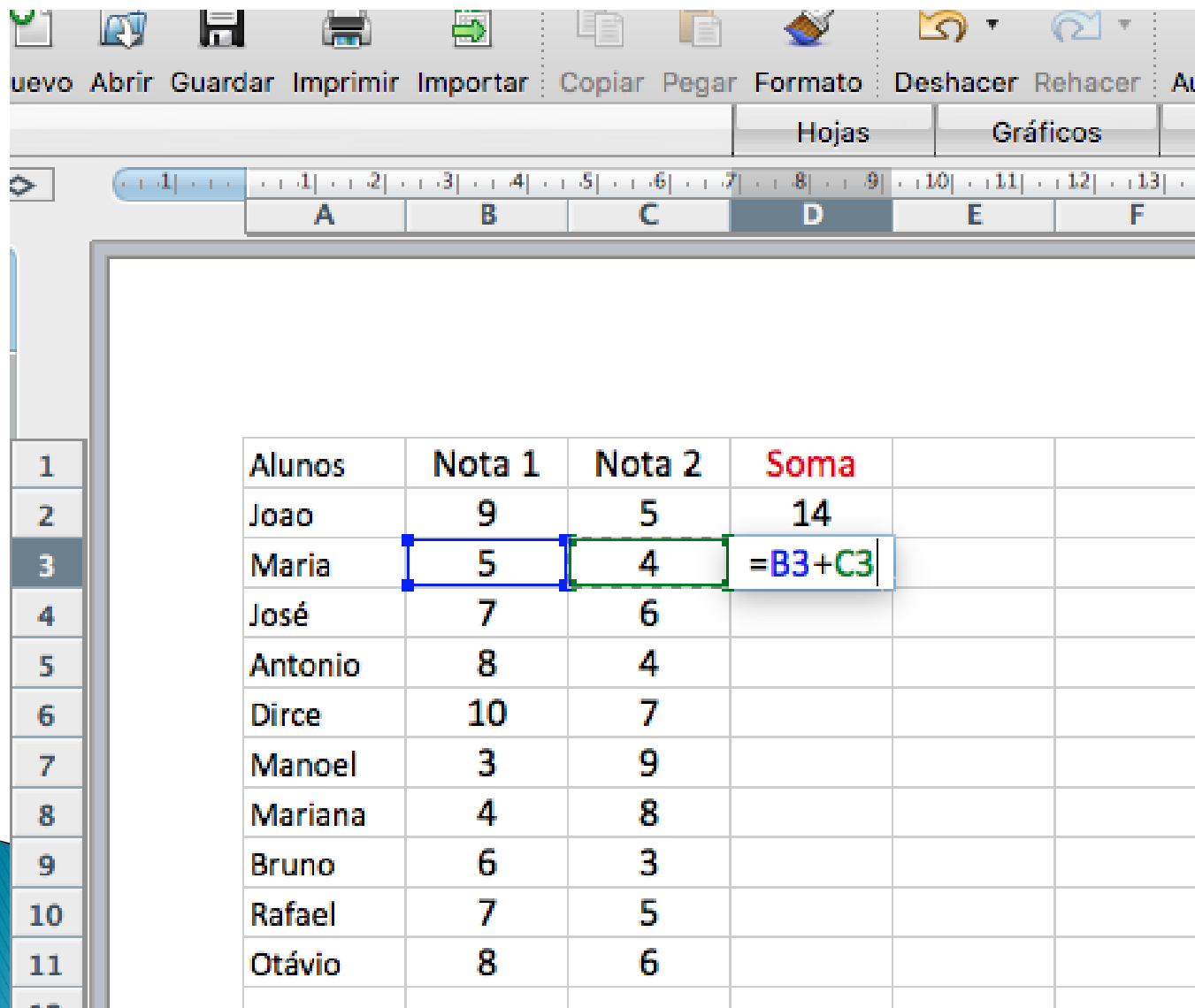
COPIEM A TABELA ABAIXO



The image shows a screenshot of a spreadsheet application. The menu bar includes options like 'Arquivo', 'Editar', 'Formato', 'Ferramentas', 'Ajuda', 'Inserir', 'Dados', 'Exibir', 'Referências', 'Formulas', 'Tabelas', 'Gráficos', 'Linha de Tabela', 'Coluna de Tabela', 'Células', 'Células em Branco', 'Células com Erros', 'Células com Fórmulas', 'Células com Comentários', 'Células com Formatação', 'Células com Links', 'Células com Mídias', 'Células com Objetos', 'Células com Tabelas', 'Células com Textos', 'Células com Tabelas', 'Células com Textos', 'Células com Tabelas', 'Células com Textos'. The spreadsheet has columns labeled A through F and rows numbered 1 through 12. The data is as follows:

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma		
2	Joao	9	5	=B2+C2		
3	Maria	5	4			
4	José	7	6			
5	Antonio	8	4			
6	Dirce	10	7			
7	Manoel	3	9			
8	Mariana	4	8			
9	Bruno	6	3			
10	Rafael	7	5			
11	Otávio	8	6			
12						

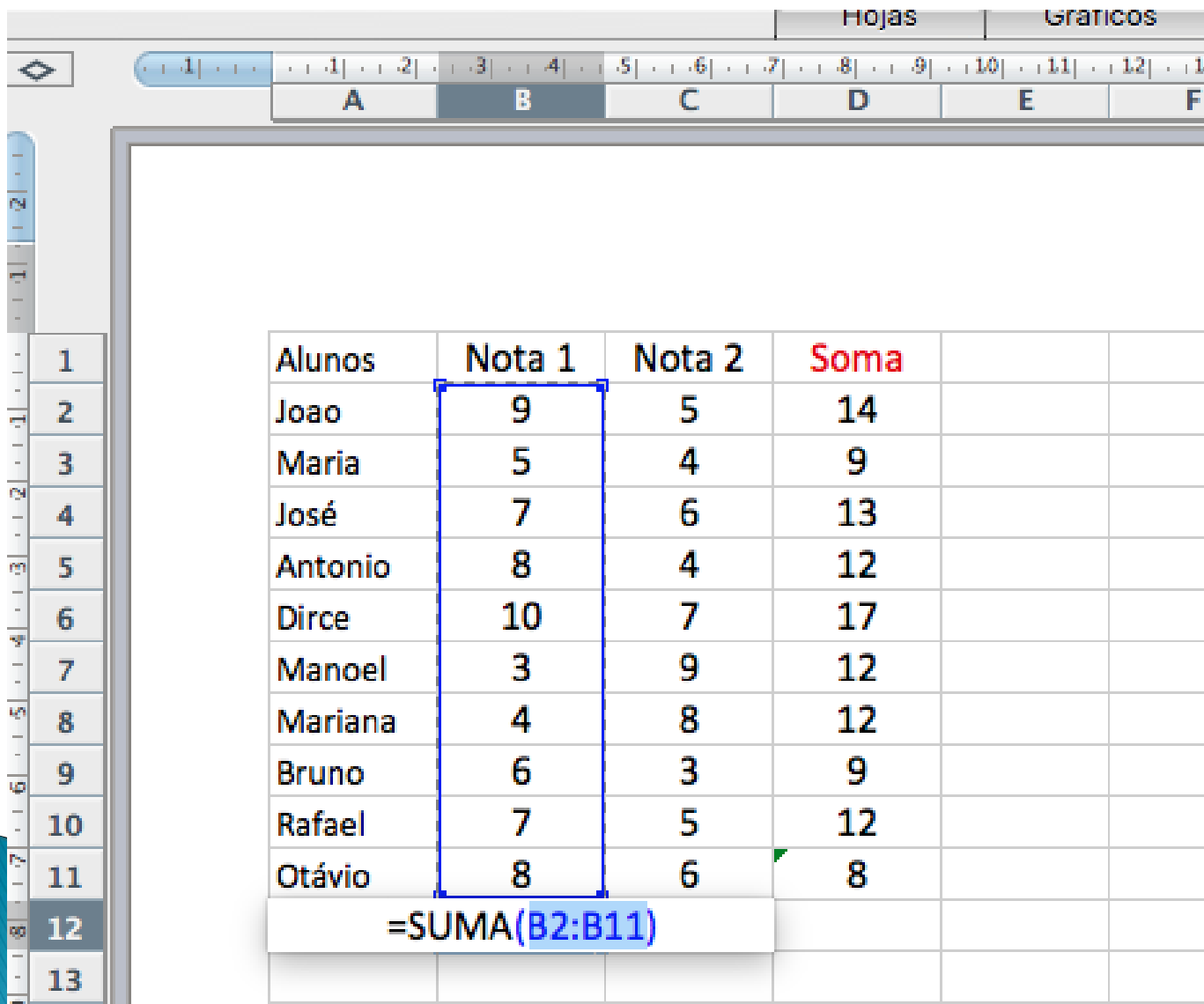
SOME OS NÚMEROS EM CADA LINHA



The image shows a spreadsheet application interface. The menu bar includes options like 'Nuevo', 'Abrir', 'Guardar', 'Imprimir', 'Importar', 'Copiar', 'Pegar', 'Formato', 'Deshacer', and 'Rehacer'. Below the menu bar, there are tabs for 'Hojas' and 'Gráficos'. The spreadsheet grid shows columns A through F and rows 1 through 11. The data is as follows:

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma		
2	Joao	9	5	14		
3	Maria	5	4	=B3+C3		
4	José	7	6			
5	Antonio	8	4			
6	Dirce	10	7			
7	Manoel	3	9			
8	Mariana	4	8			
9	Bruno	6	3			
10	Rafael	7	5			
11	Otávio	8	6			

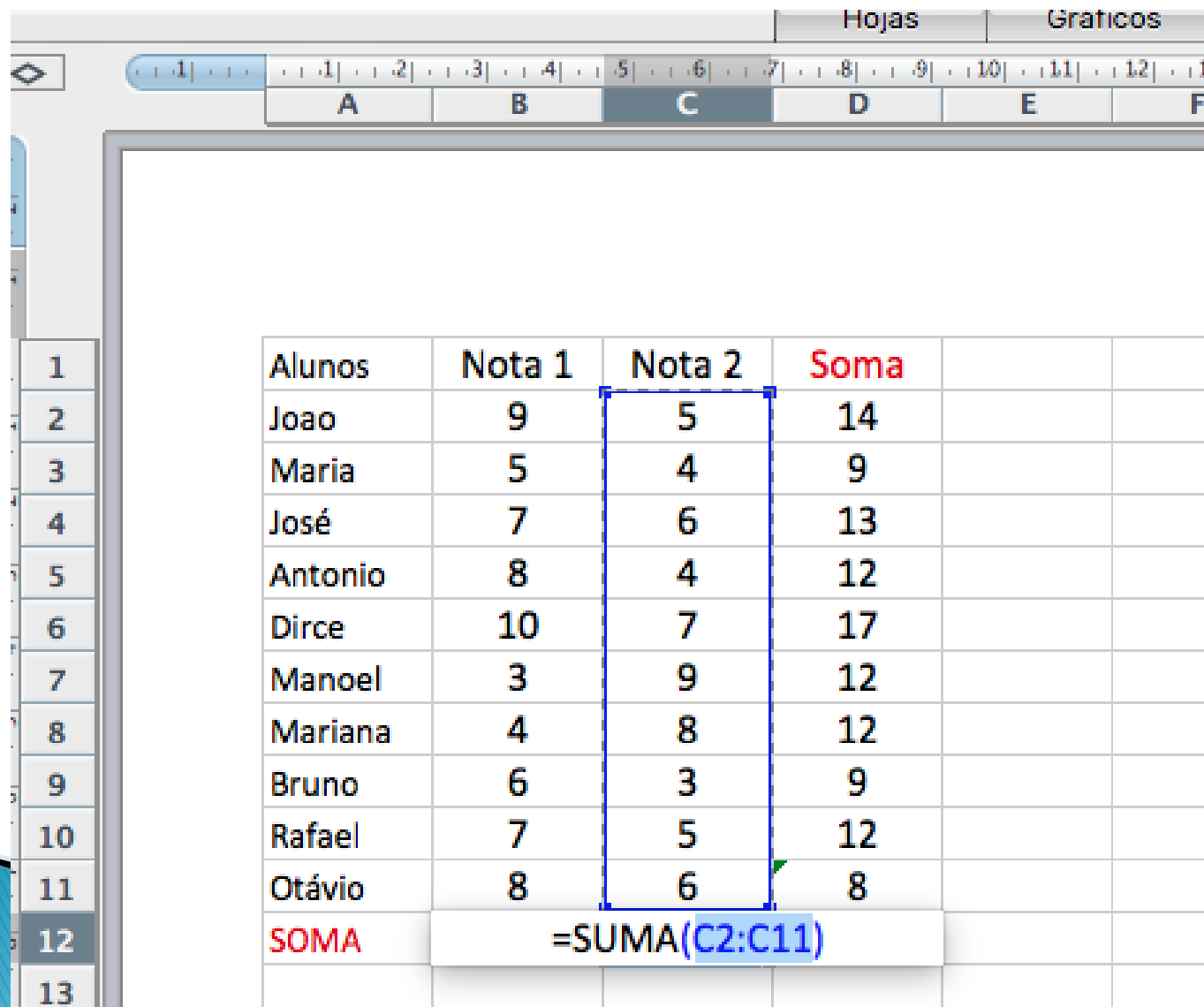
SOME OS NÚMEROS EM CADA COLUNA



The image shows a spreadsheet interface with a table of student scores. The columns are labeled 'Alunos', 'Nota 1', 'Nota 2', and 'Soma'. The rows list students: Joao, Maria, José, Antonio, Dirce, Manoel, Mariana, Bruno, Rafael, and Otávio. A blue selection box highlights the range B2:B11. A formula bar at the bottom shows the formula `=SUMA(B2:B11)`.

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma		
2	Joao	9	5	14		
3	Maria	5	4	9		
4	José	7	6	13		
5	Antonio	8	4	12		
6	Dirce	10	7	17		
7	Manoel	3	9	12		
8	Mariana	4	8	12		
9	Bruno	6	3	9		
10	Rafael	7	5	12		
11	Otávio	8	6	8		
12		<code>=SUMA(B2:B11)</code>				
13						

SOME OS NÚMEROS EM CADA COLUNA



The image shows a spreadsheet interface with a table of student grades. The columns are labeled 'Alunos', 'Nota 1', 'Nota 2', and 'Soma'. The rows list students: Joao, Maria, José, Antonio, Dirce, Manoel, Mariana, Bruno, Rafael, and Otávio. A blue selection box highlights the 'Nota 2' column from row 2 to row 11. Below the table, a formula bar shows the formula `=SUMA(C2:C11)` entered in cell C12, which is labeled 'SOMA' in red. The spreadsheet interface includes a ribbon with 'Hojas' and 'Gráficos' tabs, a sheet tab labeled '1', and column headers A through F.

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma		
2	Joao	9	5	14		
3	Maria	5	4	9		
4	José	7	6	13		
5	Antonio	8	4	12		
6	Dirce	10	7	17		
7	Manoel	3	9	12		
8	Mariana	4	8	12		
9	Bruno	6	3	9		
10	Rafael	7	5	12		
11	Otávio	8	6	8		
12	SOMA		=SUMA(C2:C11)			
13						

CALCULE A MEDIA DA COLUNA

The image shows a spreadsheet application window with a grid of data. The columns are labeled A through E, and the rows are numbered 1 through 14. The data is as follows:

Alunos	Nota 1	Nota 2	Soma	
Joao	9	5	14	
Maria	5	4	9	
José	7	6	13	
Antonio	8	4	12	
Dirce	10	7	17	
Manoel	3	9	12	
Mariana	4	8	12	
Bruno	6	3	9	
Rafael	7	5	12	
Otávio	8	6	8	
SOMA	67	57		
MEDIA	=B12/10			

The spreadsheet interface includes a ribbon with 'Hojas' and 'Gráficos' tabs, a column header row (A-E), and a row header column (1-14). The 'SOMA' row (row 12) is highlighted in red, and the 'MEDIA' cell (row 13, column B) is highlighted in green and contains the formula '=B12/10'.

CALCULE A MEDIDA DA COLUNA

The image shows a spreadsheet interface with a grid of data. The columns are labeled A through F, and the rows are numbered 1 through 15. The data is as follows:

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma		
2	Joao	9	5	14		
3	Maria	5	4	9		
4	José	7	6	13		
5	Antonio	8	4	12		
6	Dirce	10	7	17		
7	Manoel	3	9	12		
8	Mariana	4	8	12		
9	Bruno	6	3	9		
10	Rafael	7	5	12		
11	Otávio	8	6	8		
12	SOMA	67	57			
13	MEDIA	6.7	=C12/10			
14						
15						

CALCULE A MEDIA DA LINHA

The image shows a spreadsheet interface with a table of student data. The columns are labeled 'Alunos', 'Nota 1', 'Nota 2', 'Soma', and 'MEDIA'. The rows list students: Joao, Maria, José, Antonio, Dirce, Manoel, Mariana, Bruno, Rafael, and Otávio. The 'Soma' column contains the sum of the two notes for each student. The 'MEDIA' column contains the average of the two notes, calculated using the formula $=D2/2$ for Joao. The 'SOMA' and 'MEDIA' rows at the bottom of the table show the totals for each column.

	A	B	C	D	E	F
1	Alunos	Nota 1	Nota 2	Soma	MEDIA	
2	Joao	9	5	14	$=D2/2$	
3	Maria	5	4	9		
4	José	7	6	13		
5	Antonio	8	4	12		
6	Dirce	10	7	17		
7	Manoel	3	9	12		
8	Mariana	4	8	12		
9	Bruno	6	3	9		
10	Rafael	7	5	12		
11	Otávio	8	6	8		
12	SOMA	67	57			
13	MEDIA	6.7	5.7			
14						

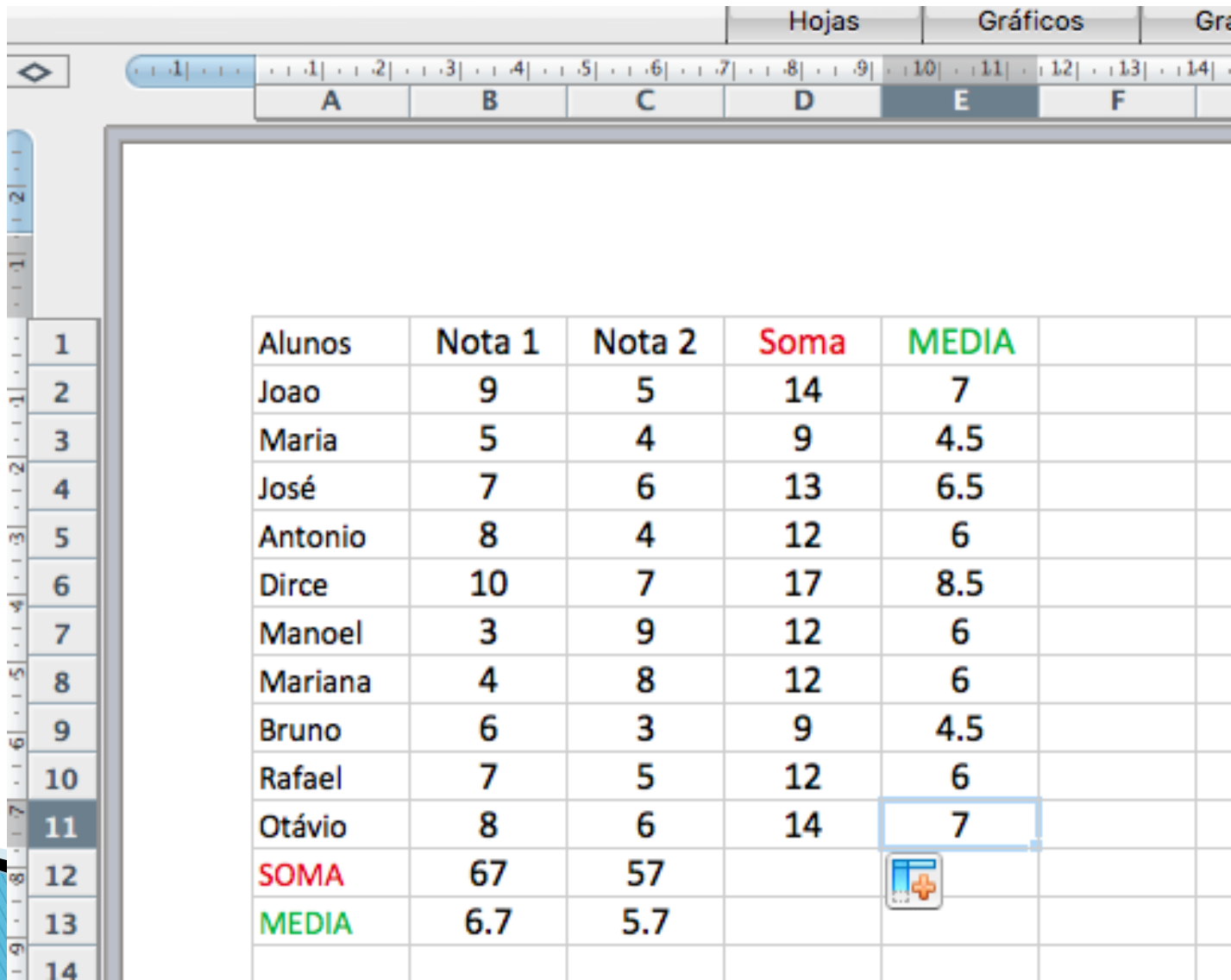
CALCULE A MÉDIA DE TODAS AS LINHAS

Alunos	Nota 1	Nota 2	Soma	MEDIA
Joao	9	5	14	7
Maria	5	4	9	4.5
José	7	6	13	6.5
Antonio	8	4	12	6
Dirce	10	7	17	8.5
Manoel	3	9	12	6
Mariana	4	8	12	6
Bruno	6	3	9	4.5
Rafael	7	5	12	6
Otávio	8	6	14	7
SOMA	67	57		
MEDIA	6.7	5.7		

Que nota se repete mais? MODA

Nota 1: 8 e 7

Nota 2: 5 e 6



The image shows a spreadsheet with columns labeled 'Alunos', 'Nota 1', 'Nota 2', 'Soma', and 'MEDIA'. The data is as follows:

Alunos	Nota 1	Nota 2	Soma	MEDIA
Joao	9	5	14	7
Maria	5	4	9	4.5
José	7	6	13	6.5
Antonio	8	4	12	6
Dirce	10	7	17	8.5
Manoel	3	9	12	6
Mariana	4	8	12	6
Bruno	6	3	9	4.5
Rafael	7	5	12	6
Otávio	8	6	14	7
SOMA	67	57		
MEDIA	6.7	5.7		

The cell containing the value '7' in the 'MEDIA' column for Otávio is highlighted with a blue border. A small icon of a calculator with a plus sign is visible in the bottom right corner of the spreadsheet area.

Exemplo Bases de Datos

BASE CIEES PARA ABRIR EN SPSS (JSR) - Microsoft Excel (Error de activación de productos)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	NUM	Control	Comité	COM	Institución	Dependencia	Entidad Federativa	EDO	Campus	Programa	Nivel educativo	NIV_EDU	Modalidad educativa	MOD_EDUC	Tipo de evaluación	T_EVAL	Subsistema CIEES	Fecha dictame
1	1	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20
2	2	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20
3	3	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20
4	4	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20
5	5	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20
6	6	14-7-20-451	CCSyA	6	Universidad Veracruzana	Facultad de Derecho	Veracruz	30	Xalapa	Derecho	Licenciatura	1	No escolarizada	4	Diagnóstica	6	20.Universidades Públicas Estatales	28/07/20

Exemplo Bases de Datos

Editor de Datos (Navegación) - [g_given_ok]

Archivo Edición Ver Datos Herramientas



unhhid_01[1] 23

	unhhid_01	ent_tip_01	mur_b_01	g8a_01	g8b_01	g7_01															
1	23	11	1	5	450	5															
2	46	11	1	2	550	1															
3	59	11	1	5	15000	1															
4	62	11	1	3	15000	3															
5	87	11	1	10	20000	2															
6	89	11	1	9	14000	3															
7	91	11	1	6	3000	3															
8	129	11	1	15	30000	3															
9	133	11	1	3	5000	1															
10	138	11	1	10	20000	2															
11	139	11	1	2	600	1															
12	142	11	1	10	15000	2															
13	145	11	1	5	35000	1															
14	146	11	1	10	8000	2															
15	147	11	1	10	30000	2															
16	164	11	1	4	33000	1															
17	165	11	1	5	72000	1															
18	168	11	1	6	5000	2															
19	173	11	1	10	150000	2															
20	176	11	1	20	41000	4															
21	179	11	1	6	600	3															
22	182	11	1	5	18000	1															
23	201	11	1	2	1500	1															
24	209	11	1	8	21500	3															
25	215	11	1	4	3000	1															
26	220	11	1	4	5450	2															

Variables

✓ Filtrar variables aquí

<input checked="" type="checkbox"/> Nombre	Etiqueta
<input checked="" type="checkbox"/> unhhid_01	unique household...
<input checked="" type="checkbox"/> ent_tip_01	type of individual ...
<input checked="" type="checkbox"/> mur_b_01	more urban
<input checked="" type="checkbox"/> g8a_01	(sum) g8a_01
<input checked="" type="checkbox"/> g8b_01	(sum) g8b_01
<input checked="" type="checkbox"/> g7_01	(count) g7_01

Variables Copia temporal

Propiedades

Variables	
Nombre	unhhid_01
Etiqueta	unique household
Typo	float
Formato	%9.0g
Etiqueta de valor	
Notas	
Datos	
Nombre de archivo	g_given_ok.dta
Etiqueta	
Notas	
Variables	6
Observaciones	1,398
Tamaño	40.96K

Activo

Vars: 6 Orden: Dataset

Obs: 1,398

Filtro: Apagado

Modo: Navegación CAP NUM

Escribe aquí para buscar

03:42 p. m.
18/08/2020

Exemplo Bases de Datos

The screenshot displays the RStudio interface. The main window shows a data table with the following columns: V29, V30, V31, V32, V33, V34, and V35. The data is as follows:

V29	V30	V31	V32	V33	V34	V35
12	12	Masculino	26	Más de 25 años	Soltero(a)	No
20	15	Masculino	19	17-19 años	Soltero(a)	No
49	19	Masculino	21	20-22 años	Soltero(a)	No
19	65	Masculino	21	20-22 años	Soltero(a)	No
0	0	0	0	0	0	0
30	25	Masculino	20	20-22 años	Soltero(a)	No
10	10	Masculino	18	17-19 años	Soltero(a)	Si
44	27	Masculino	19	17-19 años	Soltero(a)	No
58	12	Masculino	19	17-19 años	Soltero(a)	No
21	30	Masculino	19	17-19 años	Soltero(a)	Si
18	18	Masculino	19	17-19 años	Soltero(a)	No
45	45	Masculino	21	20-22 años	Unión libre	Si
0	0	0	0	0	0	0
25	15	Masculino	19	17-19 años	Soltero(a)	No
0	0	0	0	0	0	0
28	29	Masculino	19	17-19 años	Soltero(a)	No
30	25	Masculino	23	23-25 años	Soltero(a)	Si

The right-hand panel shows the Environment tab with the following data:

- base_tra... 2223 obs. of 100 vari...
- insercio... 2224 obs. of 308 vari...

The bottom of the screen shows the Windows taskbar with the search bar and various application icons. The system tray shows the time as 03:12 a. m. on 19/08/2020.

Variáveis



- Nível de escolaridade
- ° C (temperatura)
- Número de compromissos em um escritório
- Presença de alunos
- Avaliações finais
- Km² (distância)
- Estado civil de uma pessoa: Solteira, casada, divorciada, viúva.
- Nota não numérica de uma avaliação: Excelente, aprovado, aceito, reprovado.
- Cor dos olhos de um indivíduo: preto, marrom, verde, azul, etc.
- Profissão: Engenheiro, arquiteto, médico, advogado, etc.

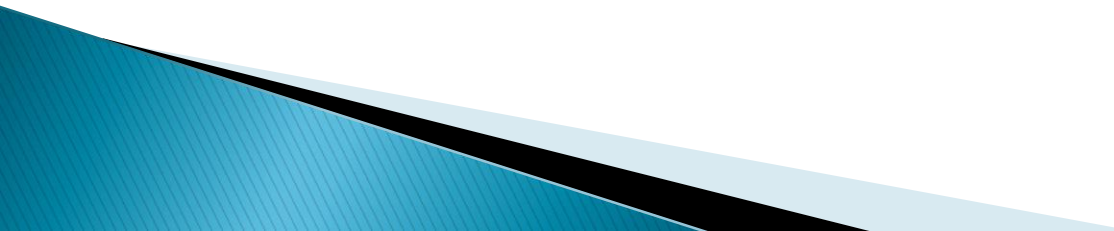
Instrumentos de captação

- ▶ Questionário de perguntas introdutórias
- ▶ Questionário básico / Questionário Ampliado
- ▶ Caderneta de exercícios cognoscitivos
- ▶ Questionário para informante substituto

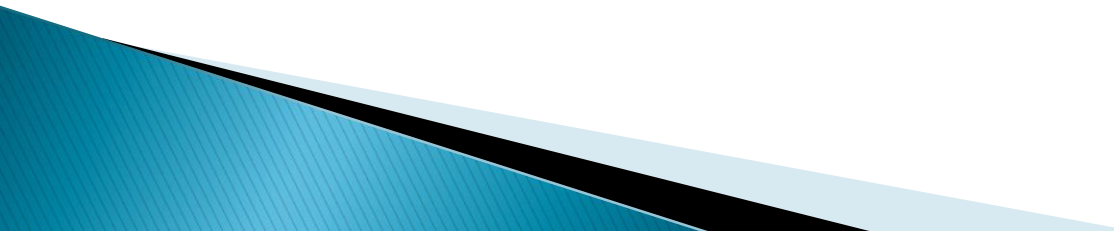
* Revisar sempre os manuais de metodologia y glossários.

Variáveis importantes (exemplos)

- ▶ CUNICAH / UNHHID Clave Única do Domicílio
 - ▶ CODENT01 Identificador de pessoa
 - ▶ SUBHOG_01 Identificador del domicilio

 - ▶ FACTORI_03 Fator de expansão individual
 - ▶ FACTORA_03 Fator de expansão de Biomarcadores
 - ▶ FACTORH_03 Fator de expansão de domicilio
- 

Medições

- ▶ As medidas de tendência central (ou de centralização) são medidas que tendem a localizar em que pontos se encontra a parte central de um conjunto ordenado de dados de uma variável quantitativa.
 - ▶ As medidas de dispersão ou medidas de variabilidade mostram a variabilidade de um conjunto de dados, indicando a maior ou menor concentração de dados em relação com as medias de centralização.
- 

Exercício no excel

- ▶ A família Pérez del Olmo é composta por 6 pessoas, observou-se que durante uma caminhada percorreram as seguintes distâncias: 1,18, 1,19, 1,20, 1,25, 1,26 e 1,29 quilômetros em uma hora.
- ▶ De um grupo de alunos do primeiro semestre foi retirada uma amostra de 10 deles, determinando-se os seguintes pesos em quilogramas: 62,5, 58,4, 62,5, 70,0, 65,7, 65,5, 68,0, 69,5, 69,5, 65,0.

Muchas gracias

Muito obrigada