

Sumário

Sumário	1
1 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA, CONJUNTOS DE DADOS	3
1.1 Expectativa de Vida por País	3
1.1.1 Sobre o Conjunto	3
1.1.2 Como utilizar o conjunto (cópia de arquivo CSV, localmente)	4
1.2 Consumo de Cigarro	4
1.2.1 Sobre o Conjunto	4
1.2.2 Como utilizar pelo R	5
1.2.3 Invocação de conjunto, por arquivo local	6
1.3 Inteligência de prodígios	6
1.3.1 Sobre o Conjunto	6
1.3.2 Invocação de conjunto, por arquivo local	7
1.4 SAT - Professores	7
1.4.1 Sobre o Conjunto	7
1.4.2 Como utilizar pelo R	7
1.4.3 Invocação de conjunto, por arquivo local	8
1.5 Toxicidade	8
1.5.1 Sobre o Conjunto	8
1.5.2 Como utilizar pelo R	9
1.5.3 Invocação de conjunto, por arquivo local	9
1.6 Fertilidade Feminina	10
1.6.1 Sobre o Conjunto	10
1.6.2 Como utilizar pelo R	10
1.6.3 Invocação de conjunto, por arquivo local	11
1.7 Atitudes em relação à empresa	11
1.7.1 Sobre o Conjunto	11
1.7.2 Como utilizar pelo R	12
1.7.3 Invocação de conjunto, por arquivo local	12
1.8 Hipótese de ciclos-de-economia salarial	12
1.8.1 Sobre o Conjunto	13
1.8.2 Como utilizar pelo R	13

1.8.3	Invocação de conjunto, por arquivo local	13
1.9	Preço de Computadores 1993-1995	14
1.9.1	Dados sobre o Conjunto	14
1.9.2	Como utilizar pelo R	14
1.9.3	Invocação de conjunto, por arquivo local	15
1.10	Peso ao Nascimento	15
1.10.1	Dados sobre o Conjunto	15
1.10.2	Invocação de conjunto, por arquivo local	16
1.11	Crime	16
1.11.1	Dados sobre o Conjunto	16
1.11.2	Invocação de conjunto, por arquivo local	17
1.12	Performance de Estudantes	18
1.12.1	Dados sobre o Conjunto	18
1.12.2	Invocação de conjunto, por arquivo local	19

1 Análise de Regressão Linear Múltipla, Conjuntos de Dados

1.1 Expectativa de Vida por País

- Modelar a variável expectativa de vida

1.1.1 Sobre o Conjunto

- Dados de todos países, derivados da OMS.
- Dados:
 - Country: país,
 - Year: ano correspondente aos dados,
 - Status: estado da industrialização,
 - Life expectancy: expectativa de vida,
 - Adult Mortality 15 < idade < 60: mortalidade adulta por mil habitantes,
 - Infant deaths: mortalidade infantil por mil habitantes,
 - Alcohol: uso de álcool por litro puro (+15 anos),
 - Percentage expenditure: gasto em saúde em percentagem do PIB per capita,
 - Hepatitis B: percentagem de imunização entre bebês com um ano,
 - Measles: quantidade de casos reportados de Rubeola por mil habitantes,
 - BMI: média de index de massa corpórea da população,
 - under-five deaths: quantidade de mortes antes dos cinco anos por mil habitantes,
 - Polio: quantidade de imunização entre bebês com um ano (percentagem),
 - Total expenditure: percentagem do gasto total do governo com saúde,
 - Diphtheria: percentagem de imunização de tétano (≤ 1 ano),
 - HIV/AIDS: morte a cada mil nascimentos (contando nascimento de 0-4 anos de idade),
 - GDP: PIB em dólares,
 - Population: População,

- thinness 1-19 years: prevalência de pessoas magras (1-19 anos de idade),
- thinness 5-9 years: prevalência de pessoas magras (5-9 anos de idade),
- Income composition of resources: IDH (qualidade de vida) em termos de composição por salário pessoal (provida por salário próprio),
- Schooling: média do número de anos sendo educado (escolaridade).

1.1.2 Como utilizar o conjunto (cópia de arquivo CSV, localmente)

O conjunto apenas está disponível por meio da **cópia da nuvem**, para o computador local. Após baixar o arquivo `Life_Expectancy_Data.csv`, para seu computador, a análise deve ser feita, em relação a um determinado ano, a qual foi indicado ao grupo.

Por exemplo, para a análise da expectativa de vida, para o ano de 2015, seria feita invocando os dados da seguinte maneira,

```
life <- read.csv('./data/csv/Life_Expectancy_Data.csv')
life_2015 <- subset(life, Year=="2015")
life_2015[1:6,1:5]
```

	Country	Year	Status	Life.expectancy	Adult.Mortality
1	Afghanistan	2015	Developing	65.0	263
17	Albania	2015	Developing	77.8	74
33	Algeria	2015	Developing	75.6	19
49	Angola	2015	Developing	52.4	335
65	Antigua and Barbuda	2015	Developing	76.4	13
81	Argentina	2015	Developing	76.3	116

Obs: apenas 5 colunas foram selecionadas, por motivos estéticos na formatação.

1.2 Consumo de Cigarro

- Modelar quantidade de venda (sales).

1.2.1 Sobre o Conjunto

- Usado em:
 - Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.

- Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.
 - Baltagi, B.H. and D. Levin (1992) “Cigarette taxation: raising revenues and reducing consumption”, *Structural Changes and Economic Dynamics*, 3(2), pp. 321–335.
 - Baltagi, B.H., J.M. Griffin and W. Xiong (2000) “To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand”, *Review of Economics and Statistics*, 82(1), pp. 117–126.
- Dados: anos 1963-1992
 - State: estados norte americanos, $\{1, 2, 3, \dots, 51\}$.
 - Year: Anos 63-92.
 - Price: preço do maço.
 - Pop: população de fumantes.
 - Pop16: população acima de 16.
 - CPI: índice do preço (ideal) dado pelo consumidor.
 - NDI: Salário líquido per capta.
 - Sales: venda de cigarro por maço.
 - Pimin: preço mínimo de cigarro em estados vizinhos

1.2.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote **Ecdat** (Econometrics Datasets [Conjuntos de Econometria]). Desta forma, precisamos instalá-lo.

```
install.packages("Ecdat", mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote.

```
load(Ecdat)
```

Finalmente, podemos acessar o conjunto.

```
cigarro <- data("Cigar")
cigarro
```

1.2.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `cigarro.csv` da seguinte maneira.

```
data <- read.csv('./data/csv/cigarro.csv')
cigarro<- subset(data, year==90)
head(data,5)
```

	state	year	price	pop	pop16	cpi	ndi	sales	pimin
28	1	90	139.1	4129.2	3148.6	130.7	12806	108.6	132.3
58	3	90	130.2	3598.2	2703.7	130.7	13826	88.9	133.6
88	4	90	141.2	2411.1	1834.9	130.7	12370	113.1	129.1
118	5	90	163.8	29602.1	22490.6	130.7	17384	77.8	130.2
148	7	90	171.2	3242.9	2567.1	130.7	21447	91.5	150.9

1.3 Inteligência de prodígios

- Modele o IQ da criança com as outras variáveis

1.3.1 Sobre o Conjunto

- Referências:
 - Graybill, F.A. & Iyer, H.K., (1994) Regression Analysis: Concepts and Applications, Duxbury, p. 511-6.
- Dados:
 1. IQ da Criança.
 2. IQ Pai.
 3. IQ Mãe.
 4. Período em meses, até primeiras palavras.
 5. Período em meses, até quanto contou até dez.
 6. Tempo passado, pelos pais, lendo livros, semanalmente.
 7. Tempo passado assistindo programas educativos, semanalmente.
 8. Tempo assistindo desenhos (`cartoons`), semanalmente.

1.3.2 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `gifted.csv` da seguinte maneira.

```
data <- read.csv('./data/csv/gifted.csv')
head(data,5)
```

	score	fatheriq	motheriq	speak	count	read	edutv	cartoons
1	159	115	117	18	26	1.9	3.00	2.00
2	164	117	113	20	37	2.5	1.75	3.25
3	154	115	118	20	32	2.2	2.75	2.50
4	157	113	131	12	24	1.7	2.75	2.25
5	156	110	109	17	34	2.2	2.25	2.50

1.4 SAT - Professores

- Modelar variável SAT (nota geral)

1.4.1 Sobre o Conjunto

- Dados:
 - State: estado norte americano
 - Expend: termo médio de aula assistida diária.
 - Ratio: razão de alunos por professor.
 - Salary: salário do professor médio, anual.
 - Frac: percentagem de alunos elegíveis a participar do SAT.
 - Verbal: nota pra proficiência verbal no SAT.
 - Math: nota para proficiência em exatadas no SAT.
 - SAT: nota geral do SAT final.

1.4.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote `mosaicData`. Desta forma, precisamos instalá-lo.

```
install.packages("mosaicData",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(mosaicData)
```

Finalmente, podemos acessar o conjunto,

```
data <- SAT  
head(data)
```

1.4.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto **professores-SAT** da seguinte maneira.

```
data <- read.csv('./data/csv/professores-SAT.csv')  
head(data,5)
```

	state	expend	ratio	salary	frac	verbal	math	sat
1	Alabama	4.405	17.2	31.144	8	491	538	1029
2	Alaska	8.963	17.6	47.951	47	445	489	934
3	Arizona	4.778	19.3	32.175	27	448	496	944
4	Arkansas	4.459	17.1	28.934	6	482	523	1005
5	California	4.992	24.0	41.078	45	417	485	902

1.5 Toxicidade

- Modelar nível de toxicidade aquática de ácidos caboxílicos, baseado nos outros descritores moleculares.

1.5.1 Sobre o Conjunto

- Dados:
 - toxicity

toxicidade aquática, definida como o $\log(\text{IGC50}^{-1})$; tipicamente, a "reposta".

- logKow log Kow, coeficiente de estacionamento
- pKa pKa: constante dissociativa
- ELUMO Energia do menor orbital molecular ocupado
- Ecarb Estado eletrotopológico do grupo carboxílico
- Emet Estado eletrotopológico do grupo metil
- RM Reflectibilidade Molar
- IR Index de Refração
- Ts Tensão superficial
- P Polaridade

1.5.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote `robustbase`. Desta forma, precisamos instalá-lo.

```
install.packages("robustbase",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(robustbase)
```

Finalmente, podemos acessar o conjunto,

```
data <- toxicity
head(data)
```

1.5.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `toxicidade` da seguinte maneira.

```
data <- read.csv('./data/csv/toxicity.csv')
head(data,5)
```

	toxicity	logKow	pKa	ELUMO	Ecarb	Emet	RM	IR	Ts	P
1	-0.15	1.68	1.00	4.81	17.8635	1.4838	31.36	1.425	31.3	12.43
2	-0.33	0.94	0.98	4.68	16.9491	0.0000	22.10	1.408	30.4	8.76

3	-0.34	1.16	0.96	4.86	17.1806	0.2778	26.73	1.418	30.9	10.59
4	0.03	2.75	1.00	4.83	18.4794	3.5836	40.63	1.435	31.8	16.10
5	-0.57	0.79	0.97	4.80	16.8022	1.0232	22.14	1.411	32.5	8.77

1.6 Fertilidade Feminina

- **Modelar MeanAFC** o qual diretamente impacta a habilidade de ter filhos.

1.6.1 Sobre o Conjunto

- Dados:
 - Idade
 - LowAFC: Conta da menor quantidade de folículos antrais.
 - MeanAFC: Média de folículos antrais.
 - FSH: Máxima quantidade de níveis de hormônios estimulantes aos folículos.
 - E2: Nível de fertilidade.
 - MaxE2: Máximo nível de fertilidade.
 - MaxDailyGn: Nível máximo de gonadotrofina.
 - TotalGn: Nível total de gonadotrofina.
 - Oocytes: Quantidade de ovócitos/oócitos.
 - Embryos: Quantidade de embriões.

1.6.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote **Stat2Data**. Desta forma, precisamos instalá-lo.

```
install.packages("Stat2Data",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(Stat2Data)
```

Finalmente, podemos acessar o conjunto,

```
data <- data("Fertility")
head(data)
```

1.6.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto **Fertilidade** da seguinte maneira.

```
data <- read.csv('./data/csv/Fertility.csv')
head(data,5)
```

	Age	LowAFC	MeanAFC	FSH	E2	MaxE2	MaxDailyGn	TotalGn	Oocytes	Embryos
1	40	40	51.5	5.3	45	1427	300	2700	25	13
2	37	41	41.0	7.1	53	802	225	1800	7	6
3	40	38	41.0	4.9	40	4533	450	4850	27	15
4	40	36	37.5	3.9	26	1804	300	2700	9	4
5	30	36	36.0	4.0	49	2526	150	1500	19	12

1.7 Atitudes em relação à empresa

- Modelar e modelar o rating

1.7.1 Sobre o Conjunto

- Dados: (em percentagem)
 - Complaints: percentagem de resoluções de reclamações.
 - Privileges: percentagem de intolerância de privilégios.
 - Learning: percentagem de oportunidade de aprendizado.
 - Raises: percentagem de aumento correspondente à performance.
 - Critical: percentagem de atitudes e críticas exacerbadas.
 - Advancel: percentagem de evolução percebida da empresa e posto.
 - Rating: avaliação geral positiva da empresa

1.7.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote `datasets`. Desta forma, precisamos instalá-lo.

```
install.packages("datasets",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(datasets)
```

Finalmente, podemos acessar o conjunto,

```
data <- data("attitude")  
head(data)
```

1.7.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `Attitude` da seguinte maneira.

```
data <- read.csv('./data/csv/attitude.csv')  
head(data,5)
```

	rating	complaints	privileges	learning	raises	critical	advance
1	43	51	30	39	61	92	45
2	63	64	51	54	63	73	47
3	71	70	68	69	76	86	48
4	61	63	45	47	54	84	35
5	81	78	56	66	71	83	47

1.8 Hipótese de ciclos-de-economia salarial

- Modelar `sr`, a partir das outras variáveis
- Hipótese formulada por Franco Modigliani 1960-1970, de que essas (outras) variáveis eram explicativas do fenômeno 'sr'.

1.8.1 Sobre o Conjunto

- Dados:
 - Sr: valor agregado à economia particular (razão entre valor total de economias pessoais e salário líquido)
 - Pop15: população sob quinze anos de idade.
 - Pop75: população acima de setenta e cinco anos de idade.
 - dpi: valor de salário líquido per-capita médio.
 - ddpi: taxa de crescimento de dpi.

1.8.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote `datasets`. Desta forma, precisamos instalá-lo.

```
install.packages("datasets",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(datasets)
```

Finalmente, podemos acessar o conjunto,

```
data <- data("LifeCycleSavings")  
head(data)
```

1.8.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `savings.csv` da seguinte maneira.

```
data <- read.csv('./data/csv/savings.csv')  
head(data,5)
```

	sr	pop15	pop75	dpi	ddpi
Australia	11.43	29.35	2.87	2329.68	2.87
Austria	12.07	23.32	4.41	1507.99	3.93
Belgium	13.17	23.80	4.43	2108.47	3.82

Bolivia	5.75	41.89	1.67	189.13	0.22
Brazil	12.88	42.19	0.83	728.47	4.56

1.9 Preço de Computadores 1993-1995

- Modelar preço (price) a partir das variáveis.

1.9.1 Dados sobre o Conjunto

- Dados:
 - Speed: velocidade de rotação em MHz do processador (clock speed)
 - HD: tamanho da memória do disco rígido em MB
 - RAM: tamanho da RAM em MB
 - Screen: tamanho da tela em polegadas
 - CD: tem ou não entrada para CDs.
 - Multi: kit multimídia incluso ou não (caixa de som etc).
 - Premium: manufatura feita por companhia conhecida (IBM etc).
 - Ads: quantidade de anúncio do computador por mês.
 - Trend: quanto tempo está no mercado.

1.9.2 Como utilizar pelo R

O conjunto de dados se encontra sob o pacote Ecdat. Desta forma, precisamos instalá-lo.

```
install.packages("Ecdat",mirror="https://vps.fmvz.usp.br/CRAN/")
```

Após instalação, precisamos invocar o pacote,

```
library(Ecdat)
```

Finalmente, podemos acessar o conjunto,

```
data <- data("Computers")  
head(data)
```

1.9.3 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto `computers.csv` da seguinte maneira.

```
data <- read.csv('./data/csv/computers.csv')
head(data,5)
```

	price	speed	hd	ram	screen	cd	multi	premium	ads	trend
1	1499	25	80	4	14	no	no	yes	94	1
2	1795	33	85	2	14	no	no	yes	94	1
3	1595	25	170	4	15	no	no	yes	94	1
4	1849	25	170	8	14	no	no	no	94	1
5	3295	33	340	16	14	no	no	yes	94	1

1.10 Peso ao Nascimento

- Modelar peso ao nascimento (libras) (Birthweight) a partir das variáveis.

1.10.1 Dados sobre o Conjunto

- Dados:
 - lowbwt: baixo peso ao nascer.
 - ID: Identificador
 - mage35: mãe mais velha do que 35 anos.
 - fnocig: número de cigarros fumado pelo pai diariamente.
 - fheight: peso do pai (kg)
 - fedysr: quantidade de anos de educação do pai
 - fage: idade do pai
 - mppwt: peso da mãe pré gravidez.
 - mheight: tamanho da mãe (cm).
 - mnocig: número de cigarros fumados diariamente pela mãe.
 - Motherage: idade da mãe.
 - Gestation Smoker: fumante durante gestação
 - headcircumfer: circunferência da cabeça
 - length: tamanho (cm)

1.10.2 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto **Birth Weight** da seguinte maneira.

```
data <- read.csv('./data/csv/Birthweight_reduced_kg_R.csv')
head(data,5)
```

	X...ID	Length	Birthweight	Headcirc	Gestation	smoker	mage	mnocig	mheight	mppwt
1	1360	56	4.55	34	44	0	20	0	162	57
2	1016	53	4.32	36	40	0	19	0	171	62
3	462	58	4.10	39	41	0	35	0	172	58
4	1187	53	4.07	38	44	0	20	0	174	68
5	553	54	3.94	37	42	0	24	0	175	66

	fage	fedyrs	fnocig	fheight	lowbwt	mage35
1	23	10	35	179	0	0
2	19	12	0	183	0	0
3	31	16	25	185	0	1
4	26	14	25	189	0	0
5	30	12	0	184	0	0

1.11 Crime

- Modelar taxa de crime (**CrimeRate**) a partir das variáveis.

1.11.1 Dados sobre o Conjunto

- Dados:
 - CrimeRate: Taxa de criminalidade (ofensas por milhão de habitantes).
 - Youth: Jovens entre 18-24 anos a cada mil habitantes.
 - Southern: Estado do sul?
 - Education: Tempo de educação (anos de estudo até 25)
 - ExpenditureYear0: Dinheiro para segurança pública per capta.
 - LabourForce: Quantidade de jovens trabalhadores por 1000 habitantes.
 - Males: Homens (por cada 1000 mulheres).
 - MoreMales: Mais homens que mulheres?.

- StateSize: Tamanho do estado em milhares.
- YouthUnemployment: Desemprego de jovens por mil.
- BelowWage: Número de famílias abaixo de meio salário por mil.
- Wage: Salário médio semanal.
- MatureUnemployment: Desemprego de sêniores (35-39) por mil.

1.11.2 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto **Crimes** da seguinte maneira.

```
data <- read.csv('./data/csv/Crime_R.csv')
head(data,5)
```

	X...CrimeRate	Youth	Southern	Education	ExpenditureYear0	LabourForce	Males
1	45.5	135	0	12.4	69	540	965
2	52.3	140	0	10.9	55	535	1045
3	56.6	157	1	11.2	47	512	962
4	60.3	139	1	11.9	46	480	968
5	64.2	126	0	12.2	106	599	989

	MoreMales	StateSize	YouthUnemployment	...	ExpenditureYear10	LabourForce10
1	0	6	80	...	71	564
2	1	6	135	...	54	540
3	0	22	97	...	44	529
4	0	19	135	...	41	497
5	0	40	78	...	97	602

	Males10	MoreMales10	StateSize10	YouthUnemploy10	MatureUnemploy10
1	974	0	6	82	20
2	1039	1	7	138	39
3	959	0	24	98	33
4	983	0	20	131	50
5	989	0	42	79	24

	HighYouthUnemploy10	Wage10	BelowWage10
1	1	632	142
2	1	521	210
3	0	359	256
4	0	510	235

1.12 Performance de Estudantes

- Modelar nota final (G3) a partir das variáveis.

1.12.1 Dados sobre o Conjunto

- Dados:
 - school - Escola em que estuda (binary: 'GP' - Gabriel Pereira ou 'MS' - Mousinho da Silveira)
 - sex - Sexo (binary: 'F' - mulher ou 'M' - homem)
 - age - Idade (numeric: from 15 to 22)
 - address - Estadia/tipo da região (binary: 'U' - urbana or 'R' - rural)
 - famsize - Tamanho familiar (binary: 'LE3' - menos que 3 or 'GT3' - mais que 3)
 - Pstatus - Estado de co-habitação dos pais (binary: 'T' - junto or 'A' - aparte)
 - Medu - Educação da mãe (numeric: 0 - nenhuma, 1 - primária (4th ano), 2 - 5th ao 9th ano, 3 - secundária, ou 4 - superior)
 - Fedu - Educação do pai (numeric: 0 - nenhuma, 1 - primária (4th ano), 2 - 5th ao 9th ano, 3 - secundária, ou 4 - superior)
 - Mjob - Trabalho da mãe (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_{home}' or 'other')
 - Fjob - Trabalho do pai (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_{home}' or 'other')
 - reason - Razão de escolha da escola (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
 - guardian - Guardião legal (nominal: 'mother', 'father' or 'other')
 - traveltime - Tempo de viagem, escola à casa (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
 - studytime - Tempo de estudo semanal (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
 - failures - Número de reprovações (numeric: n if $1 \leq n < 3$, else 4)
 - schoolsup - Suporte educacional extra (binary: yes or no)

- famsup - Suporte educacional familiar (binary: yes or no)
- paid - Classes extras pagas (Math or Portuguese) (binary: yes or no)
- activities - Atividades extra-curriculares (binary: yes or no)
- nursery - Atendeu a classes de enfermaria (binary: yes or no)
- higher - Quer seguir educação superior (binary: yes or no)
- internet - Acesso a internet (binary: yes or no)
- romantic - Em relação romântica (binary: yes or no)
- famrel - Qualidade de relação familiar (numeric: from 1 - very bad to 5 - excellent)
- freetime - Tempo livre fora da escola (numeric: from 1 - very low to 5 - very high)
- goout - Tempo gasto saindo com amigos (numeric: from 1 - very low to 5 - very high)
- Dalc - Consumo diário de álcool (numeric: from 1 - very low to 5 - very high)
- Walc - Consumo em finais de semana de álcool (numeric: from 1 - very low to 5 - very high)
- health - Estado de saúde atual (numeric: from 1 - very bad to 5 - very good)
- absences - Número de faltas (numeric: from 0 to 93)
- G1 - Notas primeiro período (numeric: from 0 to 20) [relacionadas com português ou matemática]
- G2 - Nota segundo período (numeric: from 0 to 20) [relacionadas com português ou matemática]
- G3 - Nota final (numeric: from 0 to 20, output target) [relacionadas com português ou matemática]

1.12.2 Invocação de conjunto, por arquivo local

Dado que você copiou os conjunto(s) de dado(s) para um diretório local, pode-se chamar o conjunto de métricas e notas de **Português** e **Matemática** da seguinte maneira,

```
data_mat=read.table("./data/csv/student-mat.csv",sep=";",header=TRUE)
data_por=read.table("./data/csv/student-por.csv",sep=";",header=TRUE)

data=merge(data_mat,data_por,by=c("school","sex","age","address","famsize","Pstatus","Medu",
"Fedu","Mjob","Fjob","reason","nursery","internet"))
print(nrow(data)) # 382 dados

head(data,5)
```

[1] 382

```
  school sex age address famsize Pstatus Medu Fedu Mjob    Fjob    ...
1 GP      F  15  R      GT3      T      1    1   at_home other    ...
2 GP      F  15  R      GT3      T      1    1   other   other    ...
3 GP      F  15  R      GT3      T      2    2   at_home other    ...
4 GP      F  15  R      GT3      T      2    4   services health   ...
5 GP      F  15  R      GT3      T      3    3   services services ...

  famrel.y freetime.y goout.y Dalc.y Walc.y health.y absences.y G1.y G2.y G3.y
1 3          1          2          1          1          1          4          13  13  13
2 3          3          4          2          4          5          2          13  11  11
3 4          3          1          1          1          2          8          14  13  12
4 4          3          2          1          1          5          2          10  11  10
5 4          2          1          2          3          3          2          13  13  13
```

```
summary(data)
```

school	sex	age	address
Length:382	Length:382	Min. :15.00	Length:382
Class :character	Class :character	1st Qu.:16.00	Class :character
Mode :character	Mode :character	Median :17.00	Mode :character
		Mean :16.59	
		3rd Qu.:17.00	
		Max. :22.00	
famsize	Pstatus	Medu	Fedu
Length:382	Length:382	Min. :0.000	Min. :0.000
Class :character	Class :character	1st Qu.:2.000	1st Qu.:2.000
Mode :character	Mode :character	Median :3.000	Median :3.000
		Mean :2.806	Mean :2.565
		3rd Qu.:4.000	3rd Qu.:4.000
		Max. :4.000	Max. :4.000
Mjob	Fjob	reason	nursery
Length:382	Length:382	Length:382	Length:382
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

internet	guardian.x	traveltime.x	studytime.x
Length:382	Length:382	Min. :1.000	Min. :1.000
Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000
Mode :character	Mode :character	Median :1.000	Median :2.000
		Mean :1.442	Mean :2.034
		3rd Qu.:2.000	3rd Qu.:2.000
		Max. :4.000	Max. :4.000

failures.x	schoolsup.x	famsup.x	paid.x
Min. :0.0000	Length:382	Length:382	Length:382
1st Qu.:0.0000	Class :character	Class :character	Class :character
Median :0.0000	Mode :character	Mode :character	Mode :character
Mean :0.2906			
3rd Qu.:0.0000			
Max. :3.0000			

activities.x	higher.x	romantic.x	famrel.x
Length:382	Length:382	Length:382	Min. :1.00
Class :character	Class :character	Class :character	1st Qu.:4.00
Mode :character	Mode :character	Mode :character	Median :4.00
			Mean :3.94
			3rd Qu.:5.00
			Max. :5.00

freetime.x	goout.x	Dalc.x	Walc.x	health.x
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000
1st Qu.:3.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.00	1st Qu.:3.000
Median :3.000	Median :3.000	Median :1.000	Median :2.00	Median :4.000
Mean :3.223	Mean :3.113	Mean :1.474	Mean :2.28	Mean :3.579
3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:2.000	3rd Qu.:3.00	3rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000

absences.x	G1.x	G2.x	G3.x
Min. : 0.000	Min. : 3.00	Min. : 0.00	Min. : 0.00
1st Qu.: 0.000	1st Qu.: 8.00	1st Qu.: 8.25	1st Qu.: 8.00
Median : 3.000	Median :10.50	Median :11.00	Median :11.00
Mean : 5.319	Mean :10.86	Mean :10.71	Mean :10.39
3rd Qu.: 8.000	3rd Qu.:13.00	3rd Qu.:13.00	3rd Qu.:14.00
Max. :75.000	Max. :19.00	Max. :19.00	Max. :20.00

guardian.y	traveltime.y	studytime.y	failures.y
------------	--------------	-------------	------------

Length:382	Min. :1.000	Min. :1.000	Min. :0.0000
Class :character	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:0.0000
Mode :character	Median :1.000	Median :2.000	Median :0.0000
	Mean :1.445	Mean :2.039	Mean :0.1414
	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:0.0000
	Max. :4.000	Max. :4.000	Max. :3.0000

schoolsup.y	famsup.y	paid.y	activities.y
Length:382	Length:382	Length:382	Length:382
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

higher.y	romantic.y	famrel.y	freetime.y
Length:382	Length:382	Min. :1.000	Min. :1.00
Class :character	Class :character	1st Qu.:4.000	1st Qu.:3.00
Mode :character	Mode :character	Median :4.000	Median :3.00
		Mean :3.942	Mean :3.23
		3rd Qu.:5.000	3rd Qu.:4.00
		Max. :5.000	Max. :5.00

goout.y	Dalc.y	Walc.y	health.y
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:3.000
Median :3.000	Median :1.000	Median :2.000	Median :4.000
Mean :3.118	Mean :1.476	Mean :2.291	Mean :3.576
3rd Qu.:4.000	3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000

absences.y	G1.y	G2.y	G3.y
Min. : 0.000	Min. : 0.00	Min. : 5.00	Min. : 0.00
1st Qu.: 0.000	1st Qu.:10.00	1st Qu.:11.00	1st Qu.:11.00
Median : 2.000	Median :12.00	Median :12.00	Median :13.00
Mean : 3.673	Mean :12.11	Mean :12.24	Mean :12.52
3rd Qu.: 6.000	3rd Qu.:14.00	3rd Qu.:14.00	3rd Qu.:14.00
Max. :32.000	Max. :19.00	Max. :19.00	Max. :19.00