

# Contents

<b>1</b>	<b>Como acessar o conjunto de dados pelo link</b>	<b>3</b>
<b>2</b>	<b>Análise de Regressão Linear e Logísticas Múltiplas</b>	<b>5</b>
2.1	Hipótese de ciclos-de-economia salarial . . . . .	5
2.1.1	Sobre o Conjunto . . . . .	5
2.1.2	Como utilizar pelo R . . . . .	5
2.1.3	Invocação de conjunto, por arquivo local . . . . .	6
2.2	Inteligência de prodígios . . . . .	7
2.2.1	Sobre o Conjunto . . . . .	7
2.2.2	Invocação de conjunto, por arquivo local . . . . .	7
2.3	Expectativa de Vida por País . . . . .	8
2.3.1	Sobre o Conjunto . . . . .	8
2.3.2	Como utilizar o conjunto (cópia de arquivo CSV, localmente) . . . . .	9
2.4	Peso ao Nascimento . . . . .	10
2.4.1	Dados sobre o Conjunto . . . . .	10
2.4.2	Invocação de conjunto, por arquivo local . . . . .	10
2.5	Crime . . . . .	12
2.5.1	Dados sobre o Conjunto . . . . .	12
2.5.2	Invocação de conjunto, por arquivo local . . . . .	12
2.6	Atitudes em relação à empresa . . . . .	14
2.6.1	Sobre o Conjunto . . . . .	14
2.6.2	Como utilizar pelo R . . . . .	14
2.6.3	Invocação de conjunto, por arquivo local . . . . .	15
2.7	Cancer de Mama . . . . .	16
2.7.1	Sobre o Conjunto . . . . .	16
2.7.2	Invocação de conjunto, por arquivo local . . . . .	16
2.8	Diabetes . . . . .	18
2.8.1	Sobre o Conjunto . . . . .	18
2.8.2	Invocação de conjunto, por arquivo local . . . . .	18
2.9	Arroizal . . . . .	20
2.9.1	Sobre o Conjunto . . . . .	20
2.9.2	Invocação de conjunto, por arquivo local . . . . .	20
2.10	Queimadas em Florestas . . . . .	22
2.10.1	Sobre o Conjunto . . . . .	22
2.10.2	Invocação de conjunto, por arquivo local . . . . .	22

2.11	Método Contraceptivo . . . . .	24
2.11.1	Invocação de conjuto, por arquivo local . . . . .	24
2.12	Cirurgia Torácica . . . . .	26
2.12.1	Sobre o Conjunto . . . . .	26
2.12.2	Invocação de conjuto, por arquivo local . . . . .	26

# 1 Como acessar o conjunto de dados pelo link

Para cada conjunto de dado documentado, há um hyperlink correspondente pelo qual pode-se acessar o conjunto específico.

Figure 1: Exemplo conjunto gifted.csv

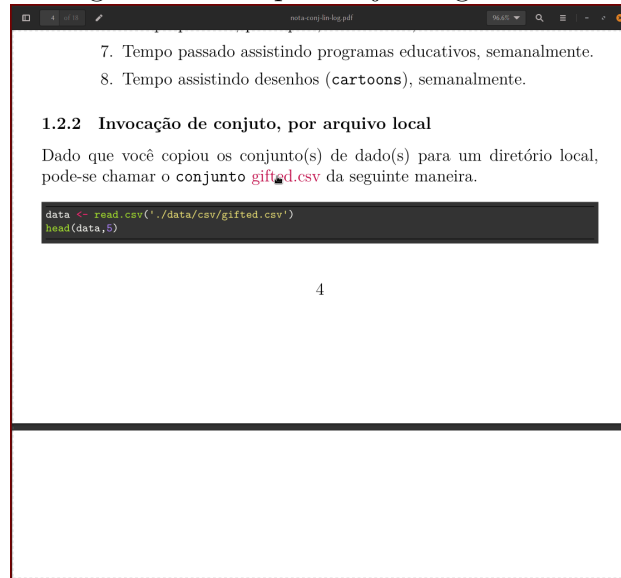


Figure 2: Exemplo direcionamento do hyperlink gifted.csv

	A	B	C	D	E	F	G	H
	idfamily	idfamily	sexes	count	read	write	distance	
1	100	110	111	10	20	1.0	2	0
2	104	117	113	20	37	2.4	1.75	0.05
3	104	110	110	20	30	2.2	2.75	0.0
4	107	119	101	10	24	1.7	2.75	0.00
5	100	110	100	17	34	2.2	0.25	0.0
6	100	110	100	10	20	1.0	1.00	0.00
7	100	114	114	10	24	1.4	2	0
8	101	117	120	10	30	2.2	2.25	0.0
9	100	111	100	20	24	2.1	-	0
10	102	102	100	10	27	2.1	0.25	0.70
11	104	111	111	10	30	2.2	1.75	0.00
12	100	112	100	20	30	2.2	2	0.70
13	107	110	100	20	30	2.2	0.75	0
14	100	100	114	20	31	1.7	0.5	0.00
15	100	114	100	17	27	1.4	1.5	0.70
16	100	111	110	10	20	2	1.75	0.00
17	100	111	110	21	30	2.2	1.75	0.00
18	104	110	111	10	30	2.2	2	0
19	100	100	111	10	30	2.2	2.75	1.70
20	101	110	100	21	28	2	2	0.70
21	100	114	100	10	30	2.4	2	0.00
22	101	110	102	20	30	2.1	2	0.0
23	100	110	114	10	30	2.2	2.00	0.70
24	100	112	101	20	30	2.2	2	0.70
25	100	110	104	10	30	2.2	2	0.0
26	101	117	110	20	31	2.2	2	0
27	100	110	100	17	30	2.4	1.25	0.70
28	100	114	110	20	30	2.4	2.00	0.70
29	100	110	101	10	24	1.4	1.5	0.0
30	100	111	117	10	28	2.1	1.75	0.00
31	100	112	117	10	20	2	2.75	0.00
32	107	110	111	10	31	2.2	1.75	0.0
33	111	101	101	10	20	2.1	2	0.00
34	100	110	111	20	30	2.1	1.00	0.0
35	104	111	101	10	30	2.2	-	0.0
36	100	114	101	10	30	2.2	1.50	0.00

Ou, acesse os links com todos os dados, e faça o download de todos eles:

- Regressão Linear
- Regressão Logística

E, seguida, siga a secção do documento referente ao conjunto do seu grupo.

## 2 Análise de Regressão Linear e Logísticas Múltiplas

### 2.1 Hipótese de ciclos-de-economia salarial

- Nome no conjunto de dados: `savings.csv`
- Variável resposta: `sr`.
- Hipótese formulada por Franco Modigliani 1960-1970, de que essas (outras) variáveis eram explicativas do fenômeno 'sr'.

#### 2.1.1 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`.

#### 2.1.2 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)
```

### 2.1.3 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

## 2.2 Inteligência de prodígios

- Nome no conjunto de dados: `gifted.csv`
- Variável resposta: `score`.

### 2.2.1 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.

### 2.2.2 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

## 2.3 Expectativa de Vida por País

- Nome no conjunto de dados: `Life_Expectancy_Data.csv`
- Variável resposta: `Life.expectancy`.

### 2.3.1 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 Rubéola 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 ( $\leq 1$  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).

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

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

Por exemplo, para a análise da expectativa de vida, para o ano de 2015, seria feita invocamos 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.

## 2.4 Peso ao Nascimento

- Nome no conjunto de dados: **Birth Weight**
- Variável resposta: **Birthweight..**

### 2.4.1 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)
  - fedyrs: 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)

### 2.4.2 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

## 2.5 Crime

- Nome no conjunto de dados: **Crimes**
- Variável resposta: **CrimeRate..**

### 2.5.1 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.

### 2.5.2 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
5	1	660	162

## 2.6 Atitudes em relação à empresa

- Nome no conjunto de dados: *Atitude*
- Variável resposta: *rating*.

### 2.6.1 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

### 2.6.2 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)
```

### 2.6.3 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 **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

## 2.7 Cancer de Mama

- Nome no conjunto de dados: `wdbc.data`
- Variável resposta: `Diagnosis`.

### 2.7.1 Sobre o Conjunto

- Dados:
  1. Número Identificador (ID).
  2. Diagnóstico (M=malígnico, B=benigno).
  3. Dez variáveis contínuas reais, com três medidas, médias, erro e pior medida (maiores valores)
    - (a) Raio da mancha
    - (b) Textura
    - (c) Perímetro
    - (d) Área
    - (e) Suavidade
    - (f) Compactividade (Quão compacto)
    - (g) Concavidade
    - (h) Pontos de Concavidade
    - (i) Simetria
    - (j) Dimensão Fractal (formato das bordas)

### 2.7.2 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 `wdbc.data` da seguinte maneira.

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

	ID	Diagnosis	Radius.mean	Texture.mean	Perimeter.mean	Area.mean
1	842302	M	17.99	10.38	122.80	1001.0
2	842517	M	20.57	17.77	132.90	1326.0
3	84300903	M	19.69	21.25	130.00	1203.0



4	84348301 M	11.42	20.38	77.58	386.1
5	84358402 M	20.29	14.34	135.10	1297.0
	Smoothness.mean	Compactness.mean	Concavity.mean	Concavity.poits.mean	...
1	0.11840	0.27760	0.3001	0.14710	...
2	0.08474	0.07864	0.0869	0.07017	...
3	0.10960	0.15990	0.1974	0.12790	...
4	0.14250	0.28390	0.2414	0.10520	...
5	0.10030	0.13280	0.1980	0.10430	...
	Radius.worst	Texture.worst	Perimeter.worst	Area.worst	Smoothness.worst
1	25.38	17.33	184.60	2019.0	0.1622
2	24.99	23.41	158.80	1956.0	0.1238
3	23.57	25.53	152.50	1709.0	0.1444
4	14.91	26.50	98.87	567.7	0.2098
5	22.54	16.67	152.20	1575.0	0.1374
	Compactness.worst	Concavity.worst	Concavity.poits.worst	Symmetry.worst	
1	0.6656	0.7119	0.2654	0.4601	
2	0.1866	0.2416	0.1860	0.2750	
3	0.4245	0.4504	0.2430	0.3613	
4	0.8663	0.6869	0.2575	0.6638	
5	0.2050	0.4000	0.1625	0.2364	
	Fractal.dim.worst				
1	0.11890				
2	0.08902				
3	0.08758				
4	0.17300				
5	0.07678				

## 2.8 Diabetes

- Nome no conjunto de dados: `diabetes.csv`
- Variável resposta: `Outcome`.

### 2.8.1 Sobre o Conjunto

- Dados:
  1. Pregnancies: número de gravidezes.
  2. Glucose: concentração de glucose durante teste.
  3. BloodPressure: pressão sanguínea.
  4. SkinThickness: espessura da pele.
  5. Insulin: nível de insulina
  6. BMI: index de massa corporal.
  7. DiabetesPedigreeFunction: valor da função modelando hereditariedade.
  8. Age: idade.
  9. Outcome: tem (1) ou não (0) diabetes.

### 2.8.2 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 `diabetes.csv` da seguinte maneira.

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

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
1	6	148	72	35	0	33.6
2	1	85	66	29	0	26.6
3	8	183	64	0	0	23.3
4	1	89	66	23	94	28.1
5	0	137	40	35	168	43.1

	DiabetesPedigreeFunction	Age	Outcome
1	0.627	50	1

2 0.351  
3 0.672  
4 0.167  
5 2.288

31 0  
32 1  
21 0  
33 1

## 2.9 Arroizal

- Nome no conjunto de dados: `rice.csv`
- Variável resposta: `Class`.

### 2.9.1 Sobre o Conjunto

- Dados:
  1. Area: área.
  2. Perimeter: perímetro.
  3. Major.Axis: eixo majoritário.
  4. Minor.Axis: eixo minoritário.
  5. Eccentricity: excentricidade.
  6. Convex.Area: area convexa.
  7. Extent: extensão.
  8. Class: Cammeo ou Osmancik (qual arroizal?)

### 2.9.2 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 `rice.csv` da seguinte maneira.

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

	Area	Perimeter	Major.Axis	Minor.Axis	Eccentricity	Convex.Area	Extent
1	15231	525.579	229.7499	85.09379	0.9288820	15617	0.5728955
2	14656	494.311	206.0201	91.73097	0.8954050	15072	0.6154363
3	14634	501.122	214.1068	87.76829	0.9121181	14954	0.6932588
4	13176	458.343	193.3374	87.44839	0.8918609	13368	0.6406690
5	14688	507.167	211.7434	89.31245	0.9066909	15262	0.6460239
Class							
1	Cammeo						
2	Cammeo						
3	Cammeo						

- 4 Cameo
- 5 Cameo

## 2.10 Queimadas em Florestas

- Nome no conjunto de dados: `forest.csv`
- Variável resposta: `Classes`.

### 2.10.1 Sobre o Conjunto

- Dados:
  1. day: dia.
  2. month: mês.
  3. year: ano.
  4. Temperature: temperatura.
  5. RH: umidade relativa.
  6. Ws: velocidade do vento.
  7. Rain: total do dia em mm.
  8. FFMC: índice de quão inflamável estava a matéria orgânica (Fine fuel moisture code).
  9. DMC: índice de quão inflamável estava materiais em decomposição (Duff moisture code).
  10. DC: índice de quão árido/seco.
  11. ISI: índice de dispersão inicial.
  12. BUI: índice de velocidade de dispersão.
  13. FWI: índice de clima propício à queimada.
  14. Classes: houve ou não queimada.

### 2.10.2 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 `forest.csv` da seguinte maneira.

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

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
1	01	06	2012	29	57	18	0	65.7	3.4	7.6	1.3	3.4	0.5	not fire
2	02	06	2012	29	61	13	1.3	64.4	4.1	7.6	1	3.9	0.4	not fire
3	03	06	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire
4	04	06	2012	25	89	13	2.5	28.6	1.3	6.9	0	1.7	0	not fire
5	05	06	2012	27	77	16	0	64.8	3	14.2	1.2	3.9	0.5	not fire

## 2.11 Método Contraceptivo

- Nome no conjunto de dados: `cmc.data`
- Variável resposta: `Contraceptive.method`.

### 1. Sobre o Conjunto

- Dados:
  - (a) `Wife.age`: idade da esposa.
  - (b) `Wife.edu`: educação esposa.
  - (c) `Husband.edu`: educação marido.
  - (d) `Children.num`: número de filhos.
  - (e) `Wife.religion`: religião esposa.
  - (f) `Wife.NotWorking.stat`: status da esposa, não trabalhando (1) ou trabalhando (0).
  - (g) `Husband.occupation`: ocupação do marido.
  - (h) `Standard.living`: qualidade de vida.
  - (i) `Media.exposure`: exposição à mídia.
  - (j) `Contraceptive.method`: método contraceptivo.

### 2.11.1 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 `cmc.data` da seguinte maneira.

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

	<code>Wife.age</code>	<code>Wife.edu</code>	<code>Husband.edu</code>	<code>Children.num</code>	<code>Wife.religion</code>	<code>Wife.NotWorking.stat</code>
1	24	2	3	3	1	1
2	45	1	3	10	1	1
3	43	2	3	7	1	1
4	42	3	2	9	1	1
5	36	3	3	8	1	1

	<code>Husband.occupation</code>	<code>Standard.living</code>	<code>Media.exposure</code>	<code>Contraceptive.method</code>
1	2	3	0	1



2 3	4	0	1
3 3	4	0	1
4 3	3	0	1
5 3	2	0	1

## 2.12 Cirurgia Torácica

- Nome no conjunto de dados: **ThoraicSurgery.csv**
- Variável resposta: **Risk1Yr**.

### 2.12.1 Sobre o Conjunto

- Dados:
  1. DNG: combinações específicas de códigos genéticos ativos,
  2. PRE4: capacidade vital forçada (FVC),
  3. PRE5: volume expirado,
  4. PRE6: estatus da performance (escala de Zubrod),
  5. PRE7: dor anterior à cirurgia,
  6. PRE8: hematopoese anterior à cirurgia,
  7. PRE9: dispneia anterior à cirurgia,
  8. PRE10: tosse anterior à cirurgia,
  9. PRE11: fraqueza anterior à cirurgia,
  10. PRE14: tamanho inicial do tumor,
  11. PRE17: presença de diabetes tipo 2,
  12. PRE19: infarto do miocárdio seis meses antes da cirurgia,
  13. PRE25: presença de doenças arterial periférica,
  14. PRE30: fumante,
  15. PRE32: asma,
  16. AGE: idade quando fez cirurgia,
  17. Risk1Yr: período de menos de um ano de sobrevivência após cirurgia (T=morte, F=continuou vivo),

### 2.12.2 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 **ThoraicSurgery.csv** da seguinte maneira.

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

	DNG	PRE4	PRE5	PRE6	PRE7	PRE8	PRE9	PRE10	PRE11	PRE14	PRE17	PRE19	PRE25
1	DGN2	2.88	2.16	PRZ1	FALSE	FALSE	FALSE	TRUE	TRUE	OC14	FALSE	FALSE	FALSE
2	DGN3	3.40	1.88	PRZ0	FALSE	FALSE	FALSE	FALSE	FALSE	OC12	FALSE	FALSE	FALSE
3	DGN3	2.76	2.08	PRZ1	FALSE	FALSE	FALSE	TRUE	FALSE	OC11	FALSE	FALSE	FALSE
4	DGN3	3.68	3.04	PRZ0	FALSE	FALSE	FALSE	FALSE	FALSE	OC11	FALSE	FALSE	FALSE
5	DGN3	2.44	0.96	PRZ2	FALSE	TRUE	FALSE	TRUE	TRUE	OC11	FALSE	FALSE	FALSE
	PRE30	PRE32	AGE	Risk1Yr									
1	TRUE	FALSE	60	FALSE									
2	TRUE	FALSE	51	FALSE									
3	TRUE	FALSE	59	FALSE									
4	FALSE	FALSE	54	FALSE									
5	TRUE	FALSE	73	TRUE									