

Estatística Descritiva II

Exemplo 1 : Projeto Qualidade de Vida

Dados *CEA15P01*: Projeto “*Avaliação da fadiga em mulheres com câncer de mama durante o tratamento quimioterápico*”

- Estudo realizado pela Escola de Enfermagem da Universidade de São Paulo - EEUSP
- **Objetivo**: Quantificar uma possível associação entre fadiga, qualidade de vida e nível de hemoglobina em mulheres com câncer de mama em tratamento quimioterápico
- **Coleta dos Dados**: junho/2010 a maio/2011
- **Análise Estatística**: realizada pelo Centro de Estatística Aplicada (CEA), IME-USP, em 2015.

- **Amostra:** 30 mulheres com idades acima de 19 anos, com diagnóstico de câncer de mama, tratadas no Ambulatório de Mastologia da Faculdade de Medicina da USP-Ribeirão Preto, no período de junho/2010 a maio/2011.
- **Dados coletados:**
 - Variáveis demográficas e biométricas
 - Fadiga e qualidade de vida: mensuradas por meio do questionário FACIT-F versão 4
 - Nível de hemoglobina (g/dL)



Coletadas nas sessões 1, 5 e 8 de quimioterapia (dados longitudinais).

Variáveis:

- **Idade** (em anos)
- **Etnia (codificação no arquivo de dados):** branca (1), preta (2), parda (3), amarela (4)
- **Estado Civil (codificação no arquivo de dados):** casada (1), solteira (2), separada (3) e viúva (4)
- **Escolaridade (codificação no arquivo de dados):** ensino fundamental incompleto (1), ensino fundamental completo (2), ensino médio incompleto (3), ensino médio completo (4), ensino superior incompleto (5) e ensino superior completo (6)

- **Situação do trabalho (codificação no arquivo de dados):** ativa (1), está de licença saúde (2), desempregada (3) e pensionista (4)
- **Fumante:** fumante (2) ou não fumante (1)
- **Peso:** em Kg
- **IMC - Índice de massa corporal:** em Kg/m^2

Questionário FACIT-F:

- Dividido em 4 domínios: BES, BEFi, BEE, BEF
- Cada domínio tem 6 ou 7 questões
- Cada questão tem pontuação de 0 a 4
- Após o paciente responder, calcula-se o escore de cada um dos domínios

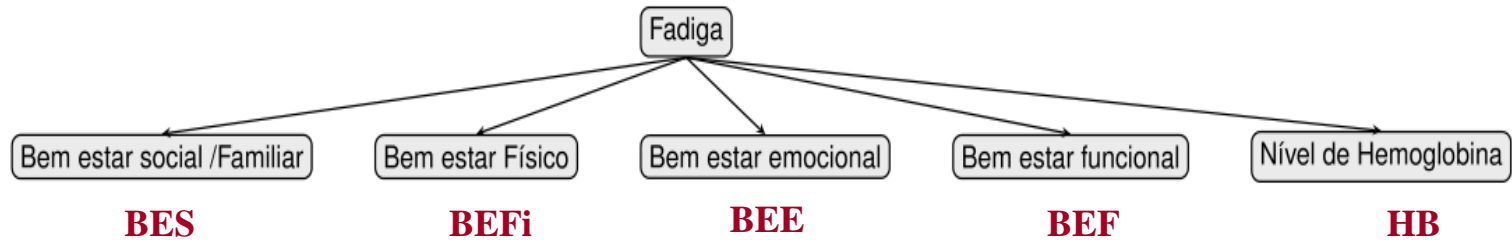


Figura 1: Diagrama da relação entre fadiga com as componentes de qualidade de vida e Nível de hemoglobina.

Exemplo

Pontuação **BEFi**: 0 a 28 - quanto maior o escore, melhor a sensação de bem-estar.

- **Bem estar físico (*BEFi*):** variável relacionada à pré-disposição física, sensação de dores, enjoo, entre outros. A pontuação varia de 0 a 28.
- **Bem estar social/familiar (*BES*):** variável relacionada à interação da paciente com a família e amigos. A pontuação varia de 0 a 28.
- **Bem estar emocional (*BEE*):** variável relacionada a questões psicológicas, como stress, preocupação com o estágio da doença, etc. A pontuação varia de 0 a 24.
- **Bem estar funcional (*BEF*):** variável relacionada à pré-disposição para trabalhar, alegria de viver, etc. A pontuação varia de 0 a 28.
- **FACT-G:** medida relacionada à qualidade de vida, obtida somando-se as pontuações das variáveis *BEFi*, *BES*, *BEE* e *BEF*. A pontuação varia de 0 a 108.
- **Subescala de fadiga (*SF*):** variável relacionada à fadiga, engloba sintomas como fraqueza, cansaço, falta de energia, dificuldade de dormir, etc. A pontuação varia de 0 a 52

⇒ Nas variáveis de qualidade de vida, **quanto maior o escore, melhor o bem estar físico e menor a fadiga.**

Banco de Dados

indivíduo	Idade	Etnia	EstCivil	Escolar	SitTrab	Fumante	Peso	Altura	IMC
1	49	1	1	1	4	1	70	152	30,3
2	47	1	1	1	4	2	75	162	28,62
3	68	3	4	1	4	2	72,7	146	33,8
4	69	1	1	1	4	1	58	149	26,12
5	58	1	2	1	1	2	77,7	156	31,97
6	63	1	1	1	3	1	59	156	24,27
7	56	2	4	1	5	1	79	161	30,5
8	24	1	2	4	2	1	59,4	160	23,2
9	49	1	1	1	1	2	68	160	26,56
10	41	1	4	1	5	1	65	152	28,13

n=30

Variáveis demográficas e biométricas da paciente. Não variam com os tempos de coleta.

indivíduo	HB	BEFi	BES	BEE	BEF	SF	FACT.G	Tempo	Idade
1	13,1	28	20	18	18	49	84	1	49
2	12	23	8	10	11	35	52	1	47
3	12,8	24	12	20	15	49	71	1	68
4	13,5	28	17	23	20	48	88	1	69
5	13,7	24	17	19	18	45	78	1	58
6	12,6	22	15	18	18	44	73	1	63
7	14,4	20	14	20	15	39	69	1	56
8	13,9	28	22	19	22	51	91	1	24
.
.
.
1	11,9	28	23	23	22	47	96	5	49
2	10,1	13	9	20	12	23	54	5	47
3	12,4	22	9	20	10	37	61	5	68
4	10,9	21	19	20	17	32	77	5	69

Tempos de coleta: 1, 5 e 8

Variáveis do Questionário FACIT-F e do nível de Hemoglobina. Avaliadas em cada sessão de quimioterapia: sessões 1, 5 e 8.

APOIO COMPUTACIONAL

Software sugerido: *R*



- Vantagem: *software* livre
- *Download*: <http://www.r-project.org/>
 - Escolher opção *Download R*
 - Seguir os passos de instalação

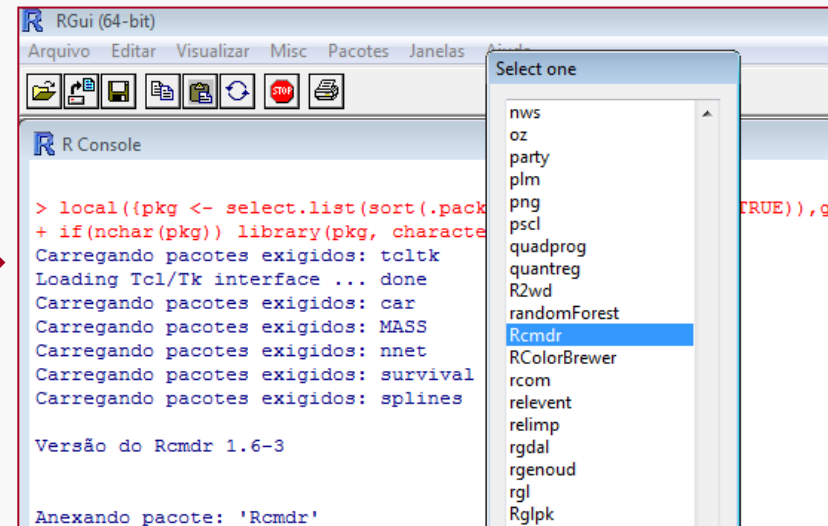
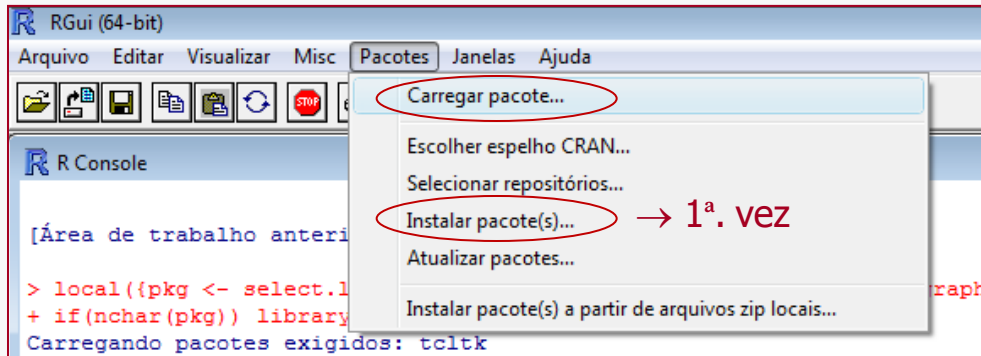
Biblioteca *Rcmdr*



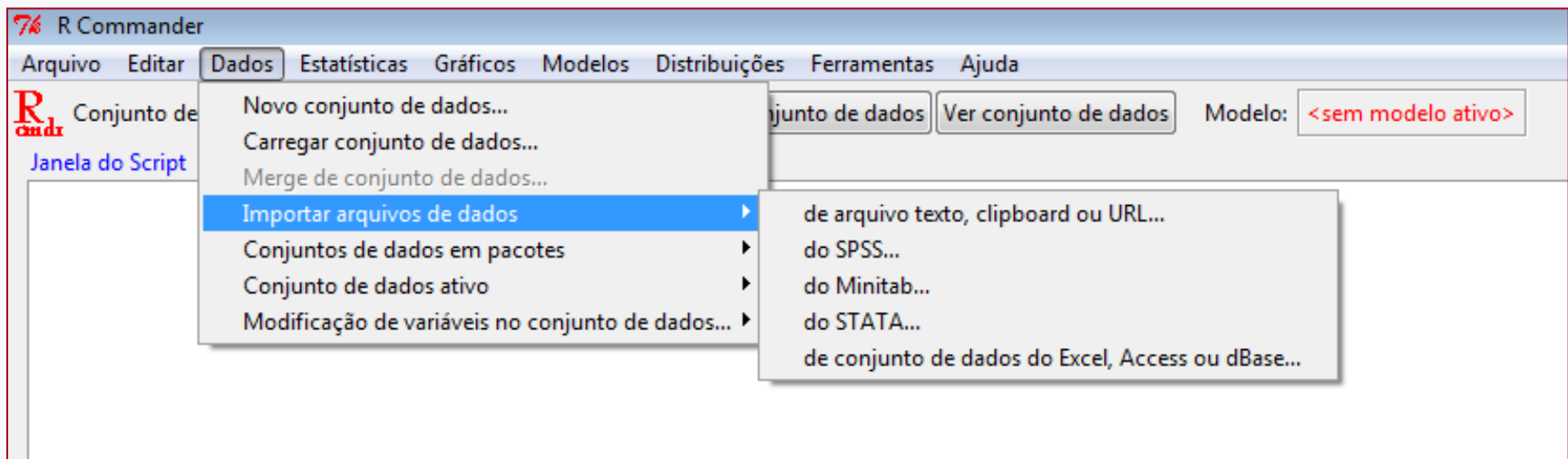
- Ambiente baseado em menus
- Deve ser instalada após instalação do *R*
- Instruções de instalação no material de apoio

Carregando dados no R : Arquivo QV

(1) Carregar *Rcmdr*:



(2) Importar dados (Excel):



Projeto Qualidade de Vida: Dados demográficos e biométricos

Visualizar / Editar dados

	indivíduo	Idade	Etnia	EstCivil	Escolaridade	SitTrab	Fumante	Peso	Altura	IMC
1	1	49	1	1	1	4	1	70.0	152	30.30
2	2	47	1	1	1	4	2	75.0	162	28.62
3	3	68	3	4	1	4	2	72.7	146	33.80
4	4	69	1	1	1	4	1	58.0	149	26.12
5	5	58	1	2	1	1	2	77.7	156	31.97
6	6	63	1	1	1	3	1	59.0	156	24.27
7	7	56	2	4	1	5	1	79.0	161	30.50
8	8	24	1	2	4	2	1	59.4	160	23.20
9	9	49	1	1	1	1	2	68.0	160	26.56
10	10	41	1	4	1	5	1	65.0	152	28.13
11	11	49	1	1	1	2	2	92.2	150	40.97
12	12	47	3	1	2	2	2	86.0	164	32.00
13	13	45	1	3	2	4	2	97.2	156	40.00
14	14	50	1	1	4	1	1	69.0	162	26.33
15	15	47	1	3	1	1	2	75.0	148	34.24
16	16	36	1	1	1	1	1	87.5	158	35.14
17	17	50	1	1	6	1	1	84.0	172	28.47
18	18	52	1	1	1	4	1	71.1	154	30.00
19	19	45	1	3	5	2	1	111.0	175	36.27
20	20	49	1	2	3	4	2	63.4	161	24.40
21	21	57	1	1	1	4	1	70.0	165	25.73
22	22	25	1	2	4	2	1	87.2	158	35.00
23	23	49	3	2	5	2	1	72.0	165	26.47
24	24	68	1	3	1	3	2	68.0	156	27.98
25	25	57	1	1	1	1	1	92.4	150	41.00
26	26	35	3	1	1	4	1	97.0	160	37.89
27	27	57	1	3	1	1	1	73.5	152	31.81
28	28	62	1	1	1	4	1	85.6	159	33.90
29	29	49	1	1	1	2	2	72.5	154	30.50
30	30	43	1	2	4	2	1	60.9	160	23.78

Variáveis qualitativas



Etnia
Fumante
Escolaridade

} Nominal

- Ordinal

Variáveis quantitativas



Idade, Peso
IMC
BEFi
BEE

} Contínua

} Discreta

Variáveis Quantitativas

Medidas de posição

Média (\bar{x})

Mediana (md)

Quartis ($Q1, Q3$)

Máximo (max)

Mínimo (min)

Medidas de dispersão

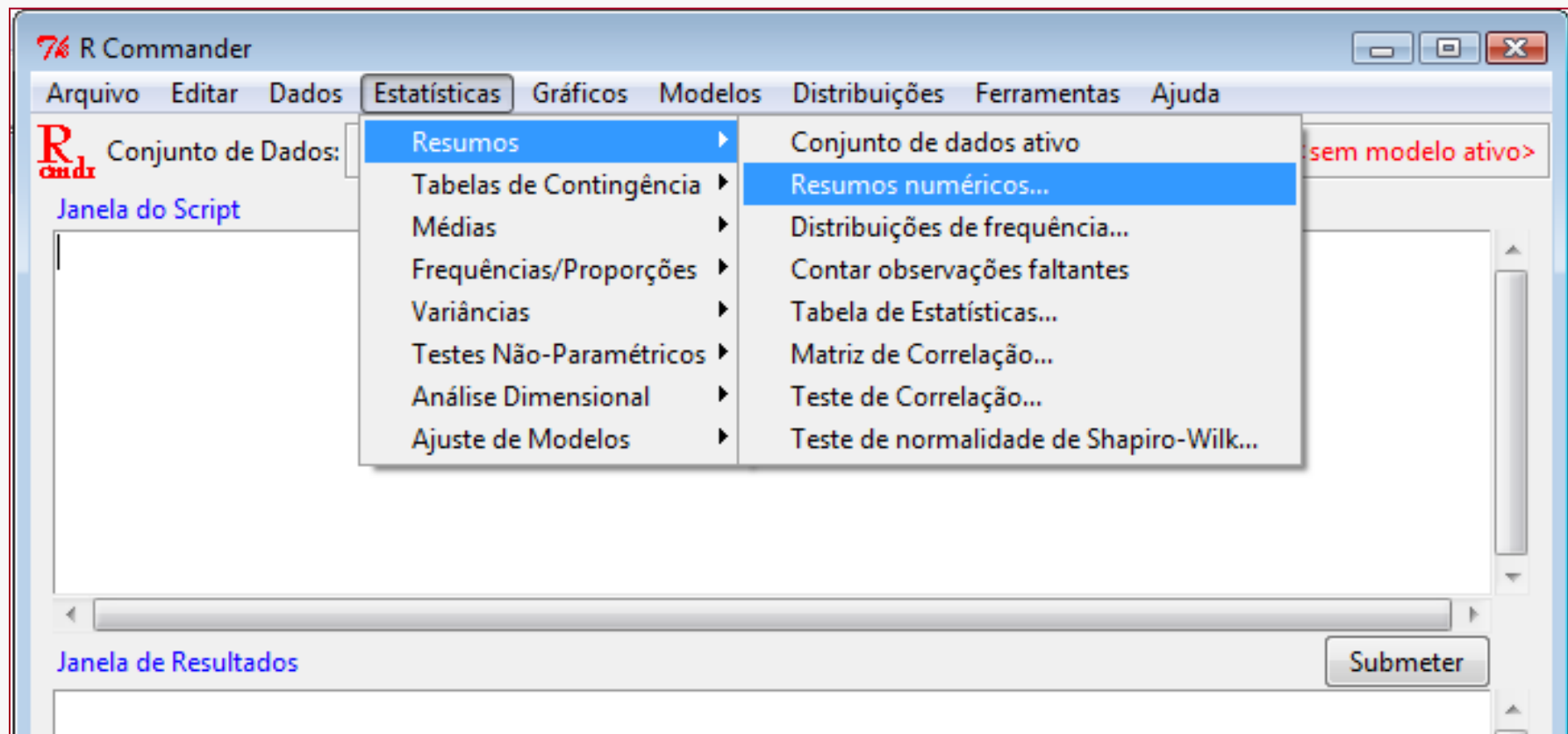
Variância (s^2)

Desvio padrão (s)

Distância interquartil ($Q3 - Q1$)

Coeficiente de variação (CV)

Rcmdr



Estatísticas → Resumos → Resumos Numéricos

Saída Rcmdr:

Variável: Idade

mean	sd	cv	0%	25%	50%	75%	100%	n
49.8667	10.9976	0.22054	24	45.5	49	57	69	30

Variável: IMC

mean	sd	cv	0%	25%	50%	75%	100%	n
30.845	5.1336	0.16643	23.2	26.4925	30.4	34.155	41	30



Projeto Qualidade de Vida - Estatísticas Descritivas

	Idade (anos)	Peso (kg)	Altura (cm)	IMC (kg/cm^2)
Mín	24,00	58,00	146,00	23,20
1o quartil	45,50	68,25	152,50	26,49
Mediana	49,00	73,10	158,00	30,40
Média	49,87	76,64	157,80	30,84
3o quartil	57,00	85,90	161,00	34,16
Máx	69,00	111,00	175,00	41,00
Desvio Padrão	11,00	12,97	6,68	5,13
Coeficiente de variação	0,22	0,17	0,04	0,17

Estadísticas Descriptivas

Variável: Idade

mean	sd	cv	0%	25%	50%	75%	100%	n
49.8667	10.9976	0.22054	24	45.5	49	57	69	30

Variável: IMC

mean	sd	cv	0%	25%	50%	75%	100%	n
30.845	5.1336	0.16643	23.2	26.4925	30.4	34.155	41	30

- 50% das pacientes tem idade menor ou igual a 49 anos. A idade mínima é de 24 anos e a idade máxima é de 69 anos;
- 25% das pacientes tem *IMC* inferior a 26,49 kg/m^2 ;
- *IMC* de 75% das pacientes foi igual ou menor a 34,16 kg/m^2 ;
- A média da idade das pacientes foi de 49,86 *a.* e desvio padrão de 11 *a.*
- O *IMC* médio das pacientes foi de 30,85 e desvio padrão de 5,13 kg/m^2 .
- As pacientes em tratamento de câncer de mama apresentaram maior variabilidade quanto à idade, quando comparada ao *IMC* ($cv=22\%$ e $cv=17\%$, respectivamente).

Rcmdr: Como converter uma variável numérica em fator

The screenshot shows the R Commander interface with the 'Dados' menu open. The 'Modificação de variáveis no conjunto de dados...' option is selected, leading to a submenu where 'Converter variável numérica para fator ...' is highlighted. Below this, the 'Converter Variáveis Numéricas p/ Fator' dialog box is open. In this dialog, 'Fumante' is selected in the 'Variáveis (selecione uma ou mais)' list. Under 'Níveis dos fatores', the radio button for 'Defina nomes dos níveis' is selected. The 'Novo nome de variável ou prefixo para múltiplas variáveis:' field contains 'HabFumar'. At the bottom of this dialog, the 'OK' button is highlighted. A secondary dialog box, 'Nomes dos níveis para HabFu...', is also open, showing the mapping of numeric values to factor levels: '1' maps to 'Não Fumante' and '2' maps to 'Fumante'. The 'OK' button in this secondary dialog is also highlighted.

R Commander

Arquivo Editar **Dados** Estatísticas Gráficos Modelos Distribuições Ferramentas Ajuda

Novo conjunto de dados...
Carregar conjunto de dados...
Merge de conjunto de dados...

Importar arquivos de dados
Conjuntos de dados em pacotes
Conjunto de dados ativo
Modificação de variáveis no conjunto de dados...

Editar conjunto de dados Ver conjunto de dados

Recodificar variáveis...
Computar nova variável...
Adicionar número da observação nos dados
Padronizar variáveis...
Converter variável numérica para fator ...
Agrupar em classes uma variável numérica (para criar fator)...
Reordenar níveis dos fatores
Drop unused factor levels...
Definir contrastes p/ um fator...
Renomear variáveis...
Apagar variáveis de um conjunto de dados...

Converter Variáveis Numéricas p/ Fator

Variáveis (selecione uma ou mais) Níveis dos fatores

Etnia
Fumante
Idade
IMC
individuo
Peso

Defina nomes dos níveis
Use números

Novo nome de variável ou prefixo para múltiplas variáveis: **HabFumar**

Ajuda OK Cancelar

Nomes dos níveis para HabFu...

Valor numérico	Nome do nível
1	Não Fumante
2	Fumante

OK Cancelar

**Rcmdr: Estatísticas → Resumos → Resumos Numéricos
→ Resumo por grupo**

Variável: Idade

HabFumar	mean	sd	0%	25%	50%	75%	100%	data:n
Não Fumante	48.42	12.23	24	42	50	57.0	69	19
Fumante	52.36	8.41	45	47	49	53.5	68	11

Variável: IMC

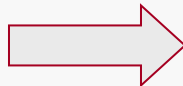
HabFumar	mean	sd	0%	25%	50%	75%	100%	data:n
Não Fumante	30.23	5.14	23.2	26.23	30.00	34.45	41.00	19
Fumante	31.91	5.19	24.4	28.30	31.97	34.02	40.97	11

- A média da idade das pacientes não fumantes é menor que a de fumantes. Contudo, 75% das pacientes do grupo não fumante tem idades inferiores a 57 anos e no grupo de fumantes este valor é de 53 anos. A variabilidade da idade no grupo de não fumantes é maior.
- O hábito de fumar influencia pouco na distribuição do IMC.

Os dados também podem ser resumidos por meio de tabelas de distribuição de frequências.

Distribuição de frequências de uma variável é uma lista dos valores individuais ou dos intervalos de valores que a variável pode assumir, com as respectivas frequências de ocorrência.

**Variável
IDADE**



**Não há perda
de informação**



Distribuição de frequências, var. cont.

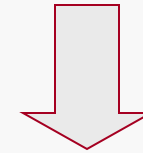
Idade	Freq. Abs.	Porcent. %
24	1	3,33
25	1	3,33
35	1	3,33
36	1	3,33
41	1	3,33
43	1	3,33
45	2	6,67
47	3	10,00
49	6	20,00
50	2	6,67
52	1	3,33
56	1	3,33
57	3	10,00
58	1	3,33
62	1	3,33
63	1	3,33
68	2	6,67
69	1	3,33
n=30		100,00%

Alternativa: construir intervalos de classe

Projeto *QV* – **Variável Idade**

Classes de Idade	<i>f</i>
(24; 33]	2
(33; 42]	3
(42; 51]	14
(51; 60]	6
(60; 69]	5
Total	30

Informações mais
resumidas



Perda de informação

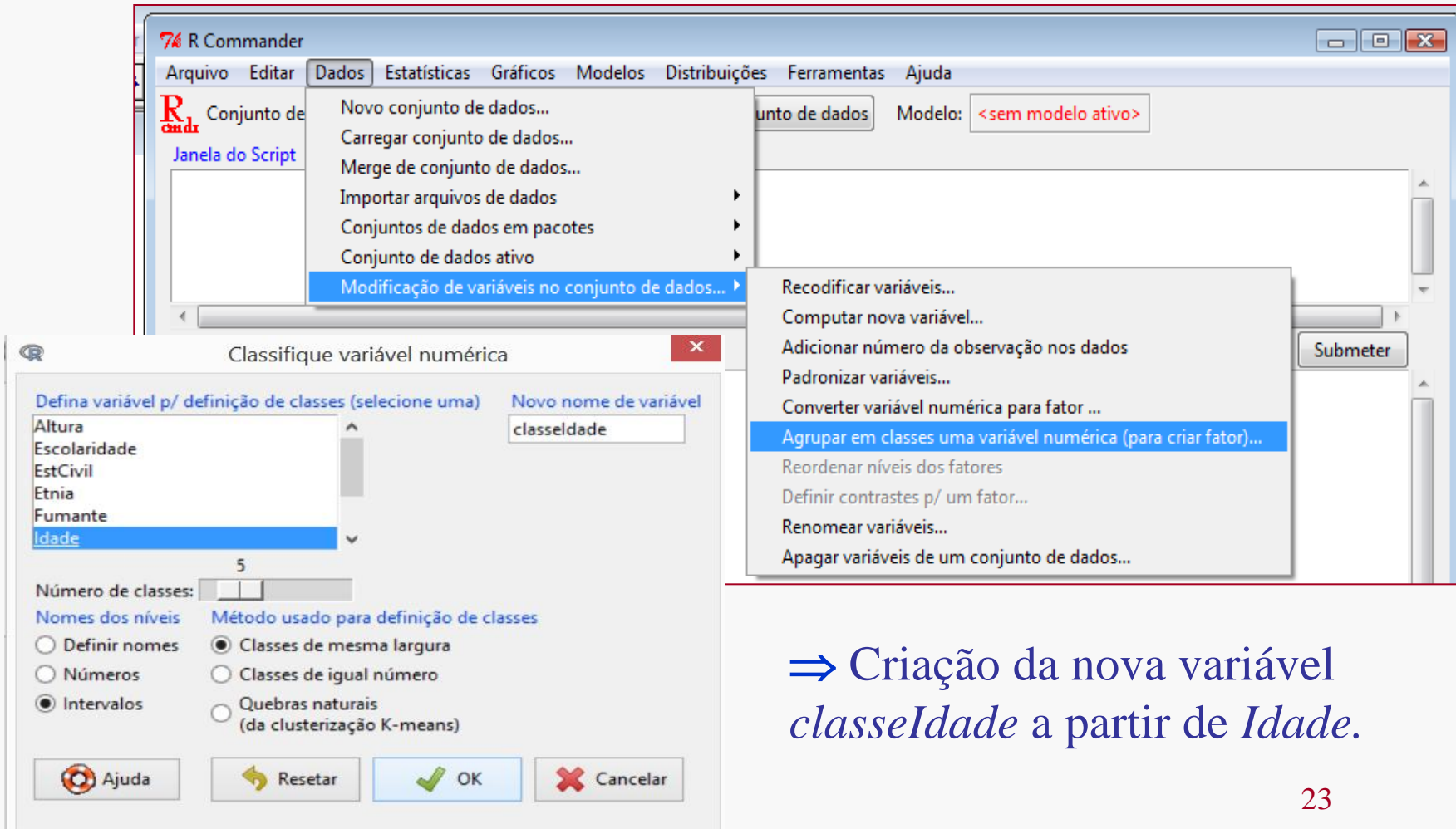
Distribuição das pacientes do projeto *QV* de acordo com o *IMC*

Classes de IMC	f
(23,2; 26,7]	9
(26,7; 30,3]	6
(30,3; 33,9]	6
(33,9; 37,5]	5
(37,5; 41]	4
Total	30

Variável: Idade \Rightarrow quantitativa \Rightarrow

Construir
intervalos
de classe

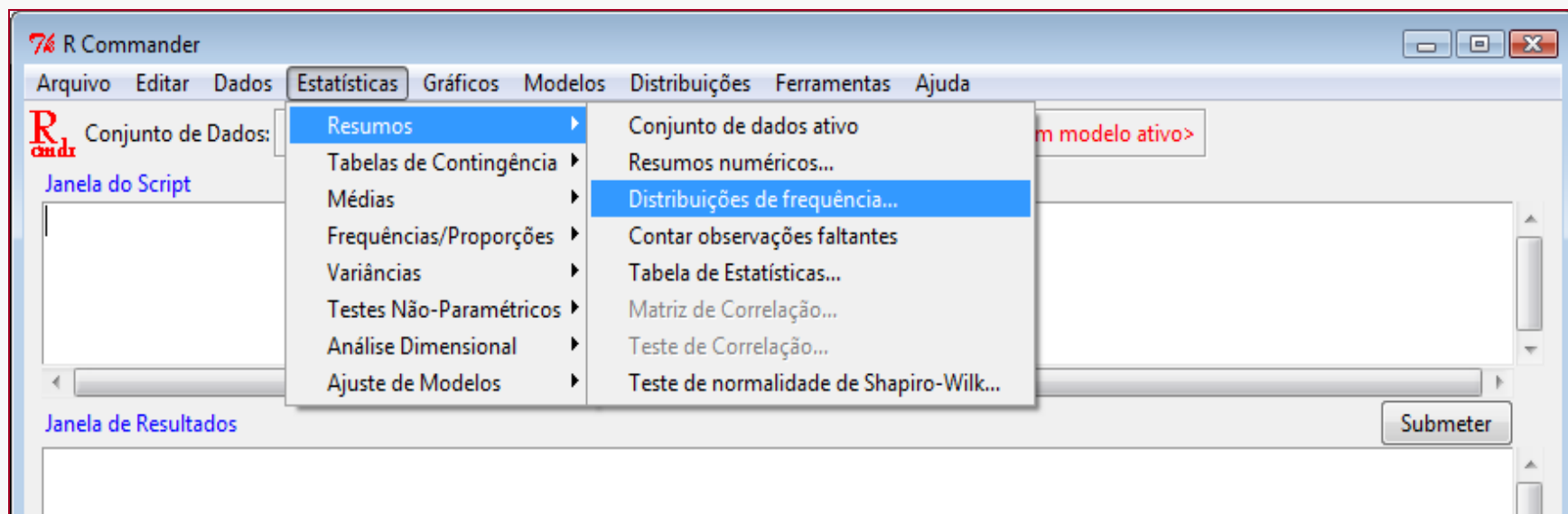
Rcmdr: (1) criar nova variável



\Rightarrow Criação da nova variável
classIdade a partir de *Idade*.

Rcmdr:

(2) obter a distribuição de frequências da nova variável



Distribuição de frequências para *classeIdade*

Classes de Idade	f	fr (%)
(24,33]	2	6,67
(33,42]	3	10,00
(42,51]	14	46,67
(51,60]	6	20,00
(60,69]	5	16,67
Total	30	100,00

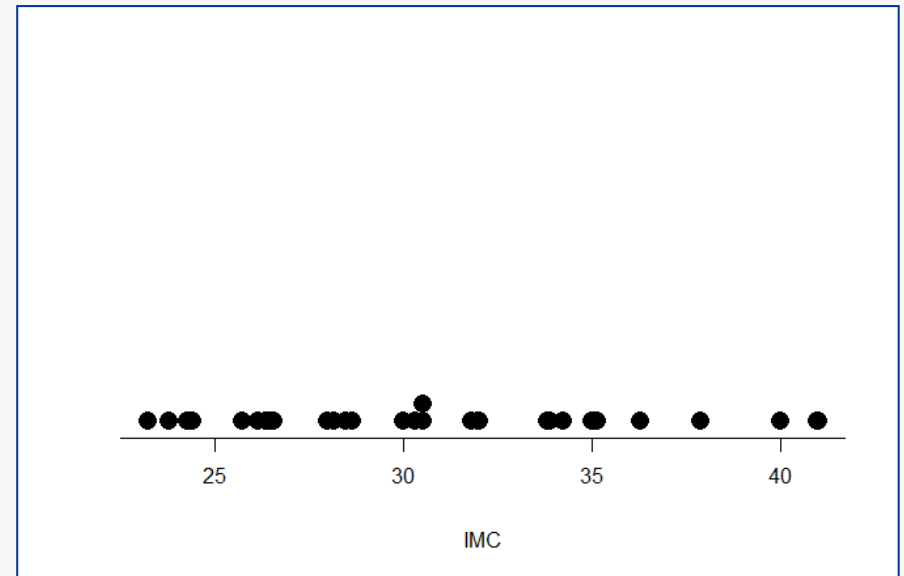
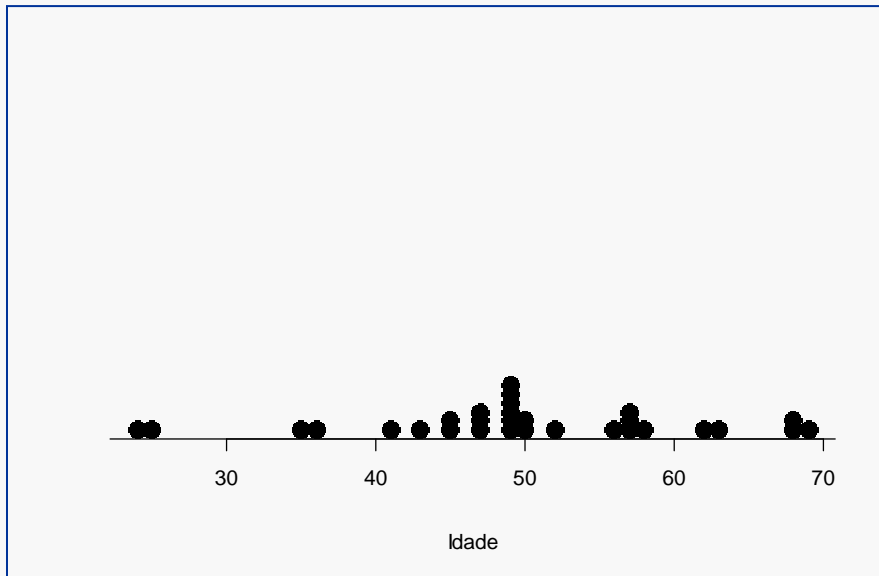
Análise Descritiva - Variáveis Quantitativas

Representação Gráfica

- “Strip Chart” ou “Dotplot” ou “Gráfico de pontos”
- “Boxplot” (desenho esquemático)
- Histograma

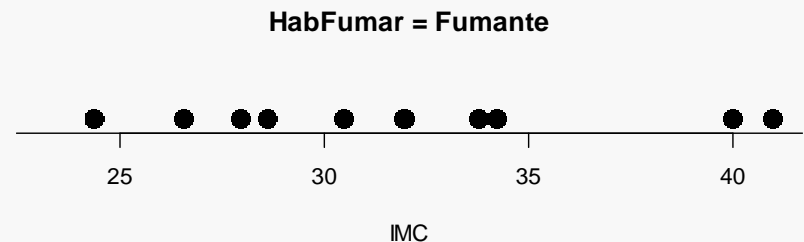
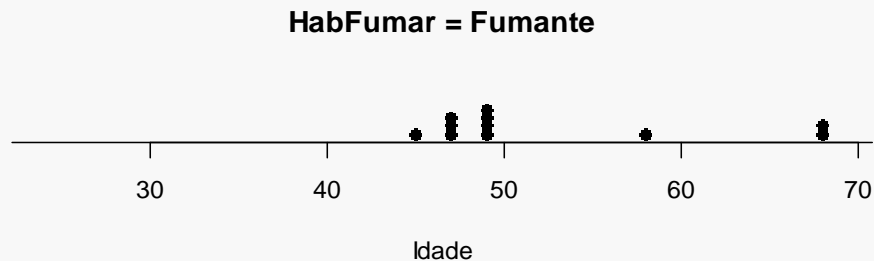
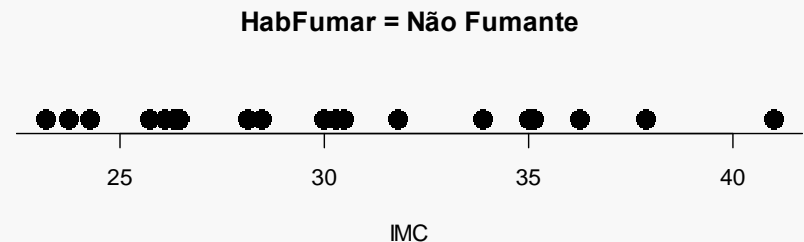
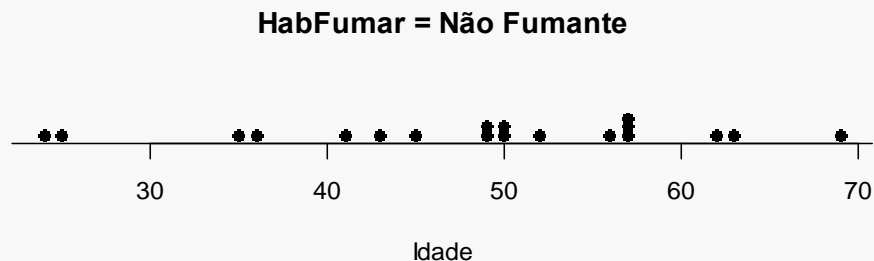
STRIP CHART ou DOT PLOT

Dados do projeto Qualidade de Vida de pacientes em tratamento de câncer – Variáveis Idade e *IMC*



STRIP CHART ou DOT PLOT

Dados do projeto Qualidade de Vida de pacientes em tratamento de câncer de mama – Variáveis Idade e IMC de acordo com o hábito de fumar.

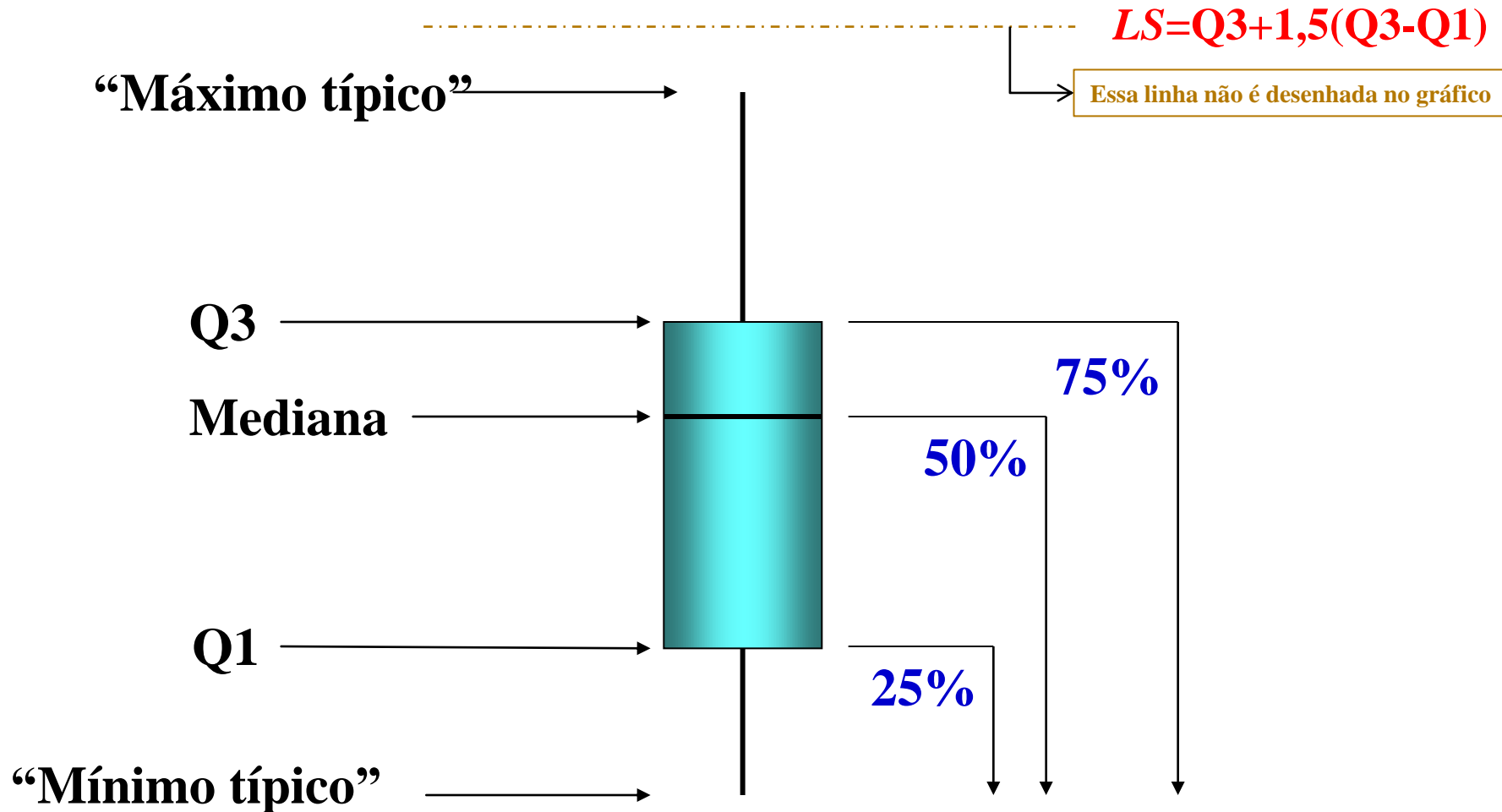


A distribuição da idade é diferente para os dois grupos de pacientes (fumantes e não fumantes), já o IMC é similar nos dois grupos.

BOXPLOT

Representação dos dados por meio de um retângulo construído com os **quartis** e fornece várias informações, incluindo a existência de **valores discrepantes**.

Boxplot - Construção



“Máximo típico” é o maior valor menor que LS ;

“Mínimo típico” é o menor valor maior que LI .

$$LI = Q1 - 1,5(Q3 - Q1)$$

Essa linha não é desenhada no gráfico

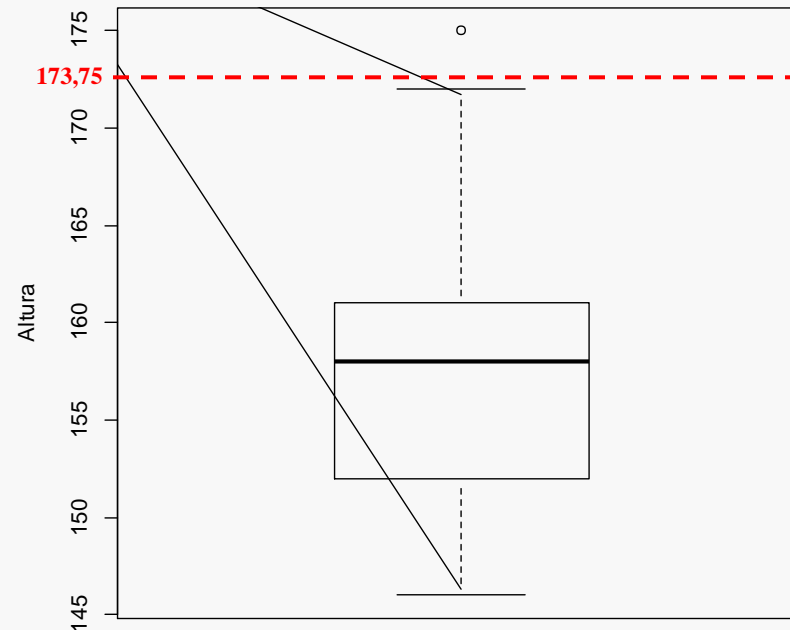
Projeto Qualidade de vida – Variável Altura das pacientes

Dados ordenados ($n=30$)

146	148	149	150	150	152
152	153	154	154	156	156
156	156	158	158	159	160
160	160	160	161	161	161
162	164	165	165	172	175

Observação discrepante

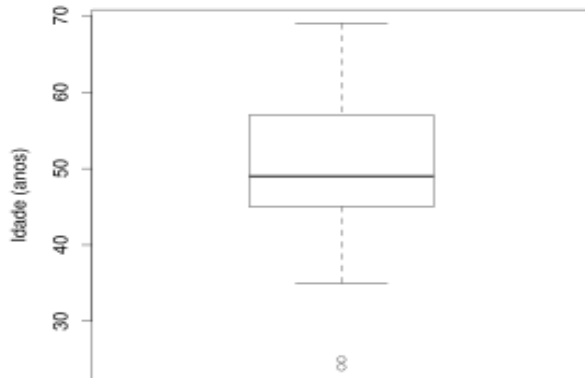
Min=146 $Q1=152,5$ $Q2=158$ $Q3=161$ Max=175



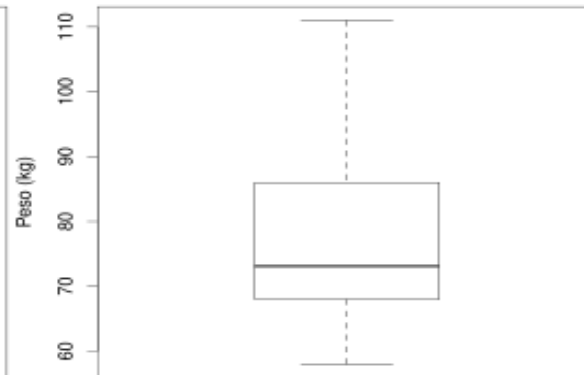
$$LI = Q1 - 1,5(Q3 - Q1) = 139,75$$

$$LS = Q3 + 1,5(Q3 - Q1) = 173,75$$

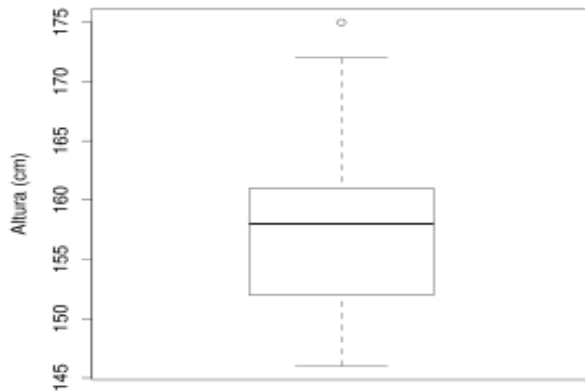
Boxplot – Projeto Qualidade de Vida



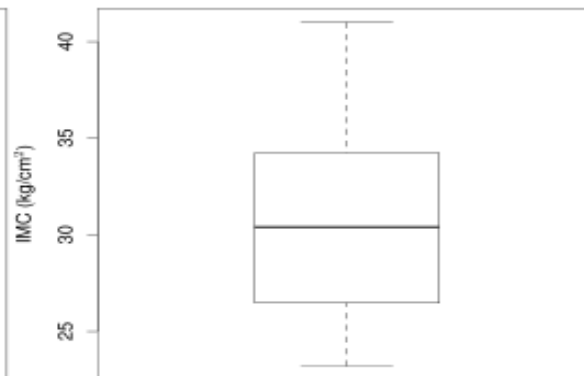
(a) Idade



(b) Peso



(c) Altura



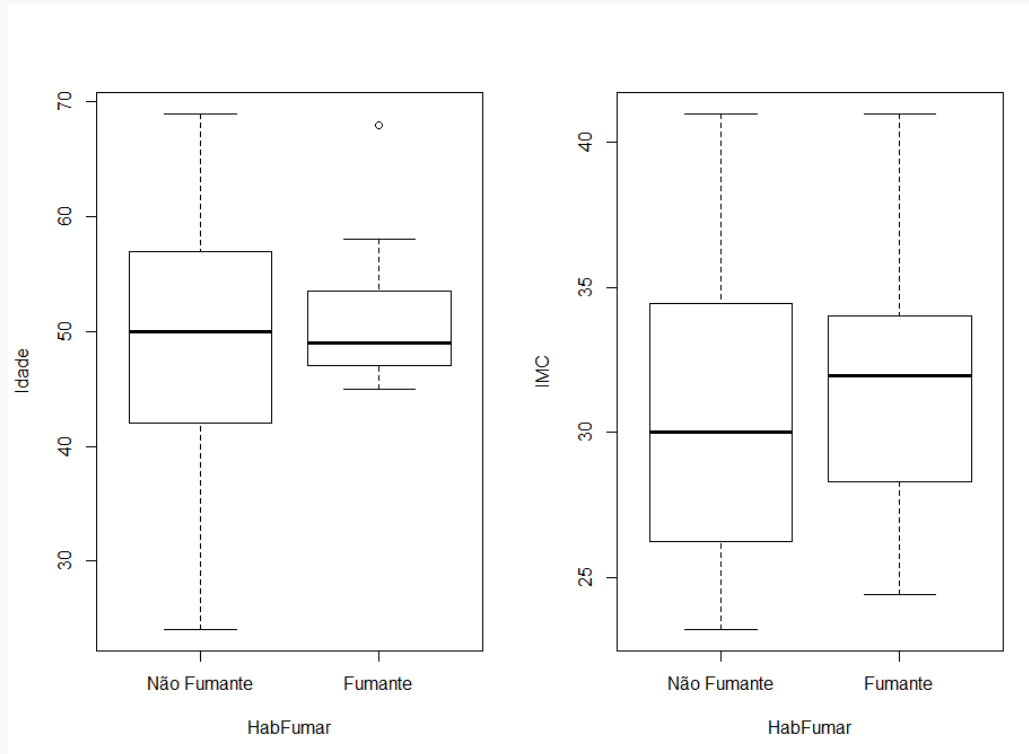
(d) IMC

- A distribuição do IMC das pacientes é mais simétrica quando comparada às demais variáveis.

- Idade e Altura apresentam valores discrepantes.

Boxplot – Projeto Qualidade de Vida

Variáveis **Idade** e **IMC** de acordo com o **hábito de fumar**.



Confirmando as demais análises dos dados, a distribuição da idade é bastante diferente para os dois grupos de pacientes (fumantes e não fumantes).

HISTOGRAMA

Agrupar os dados em intervalos de classes
(distribuição de frequências)

Bases iguais

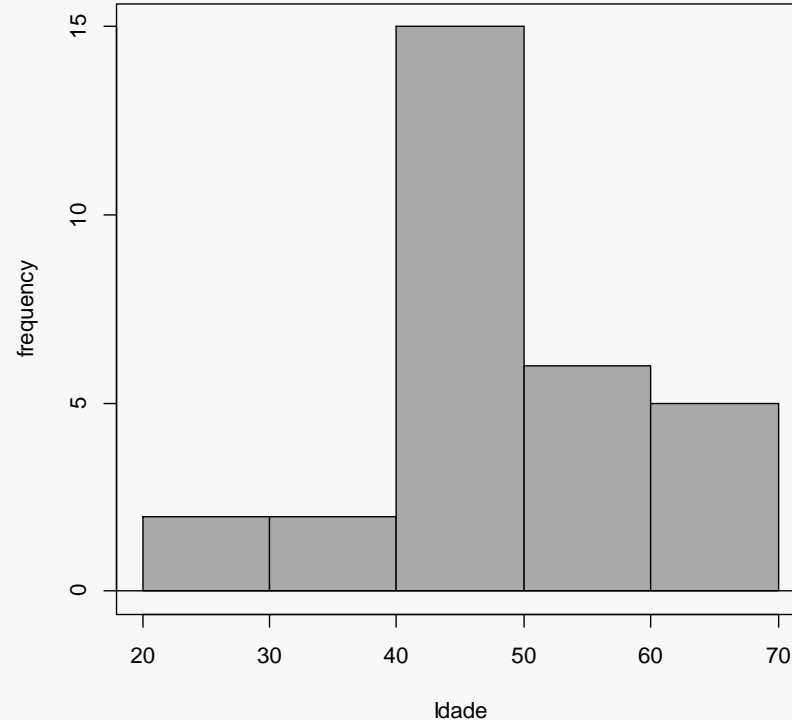
Construir um retângulo para cada classe, com base igual à amplitude do intervalo de classe e *altura proporcional à frequência da classe (f)*.

Bases diferentes

Construir um retângulo para cada classe, com base igual à amplitude do intervalo de classe e *área do retângulo igual a frequência relativa da classe (fr)*.

Projeto Qualidade de vida – Histograma da Idade

Classes de Idade	f
(20; 30]	2
(30; 40]	2
(40; 50]	15
(50; 60]	6
(60; 70]	5
Total	30



No *Rcmdr*: Gráficos → Histograma → Options
(número de classes=5) ...

Projeto Qualidade de vida – Histograma do *IMC*

Classes de <i>IMC</i>	<i>f</i>
-----------------------	----------

(20; 25]	4
----------	---

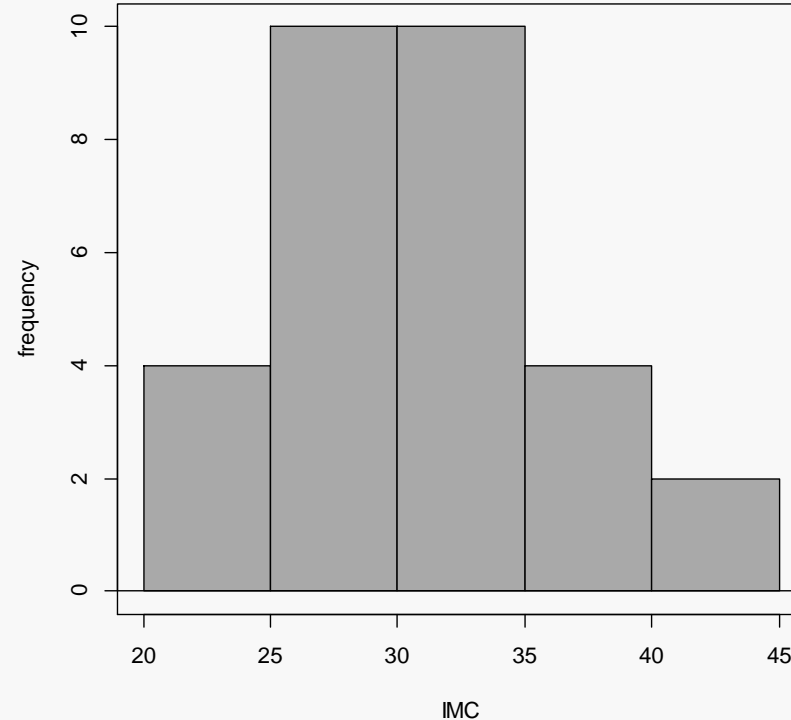
(25; 30]	10
----------	----

(30; 35]	10
----------	----

(35; 40]	4
----------	---

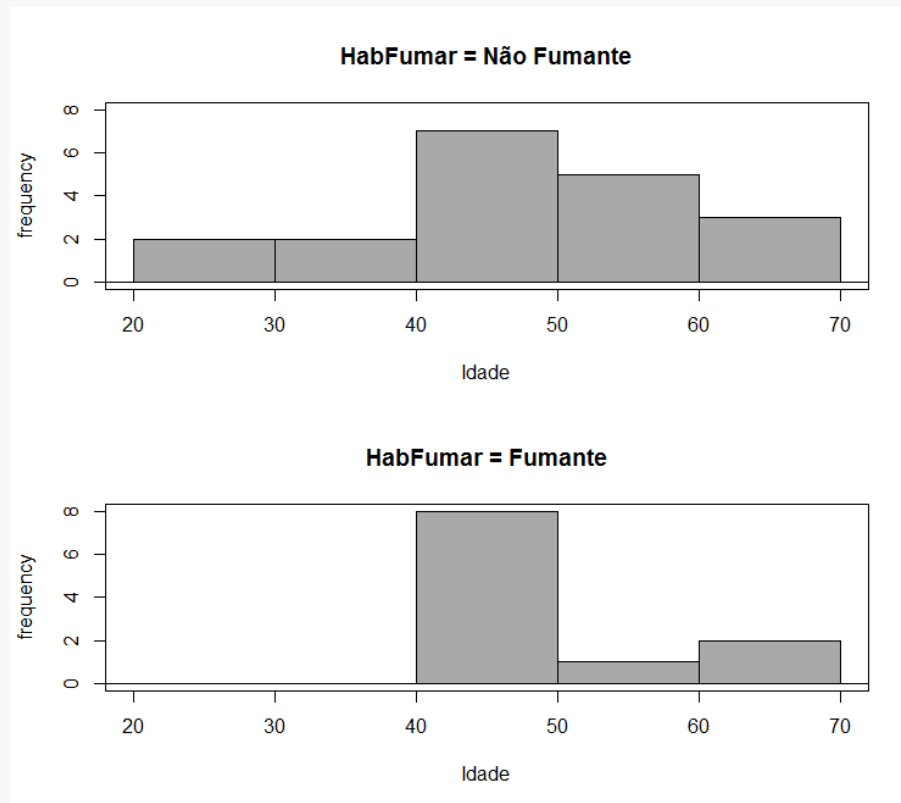
(40; 45]	2
----------	---

Total	30
-------	----



Histograma –Projeto Qualidade de Vida

Distribuição da Idade das pacientes de acordo com o hábito de fumar



O histograma confirma o padrão já evidenciado por meio do boxplot: a distribuição da idade é diferente para os dois grupos de pacientes (fumantes e não fumantes).

Histograma com Classes Desiguais

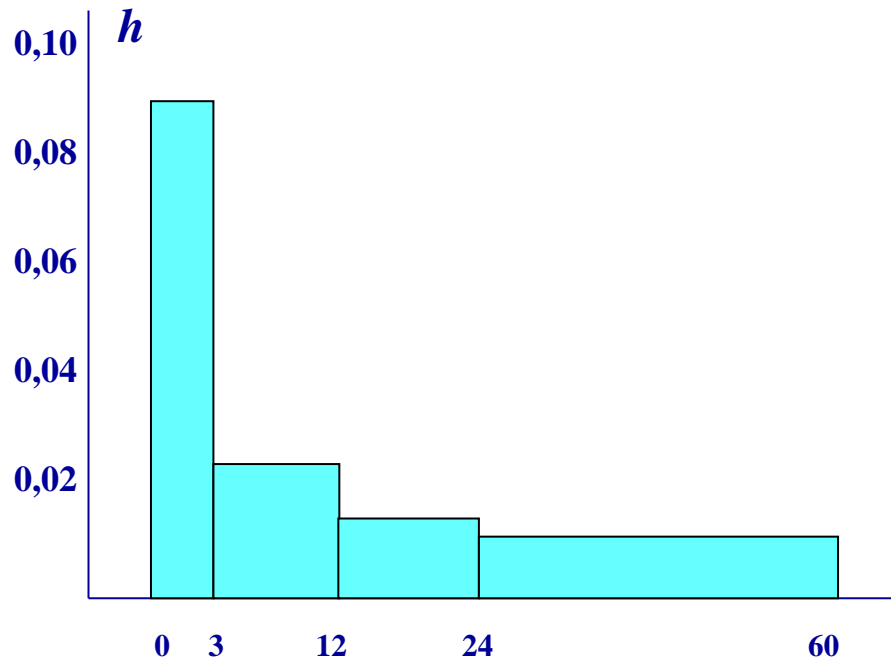
Construção:

- base igual à amplitude do intervalo de classe
- área do retângulo igual a frequência relativa da classe (fr)
- A altura será dada por: $h = fr/base$ (h é a densidade de frequência).

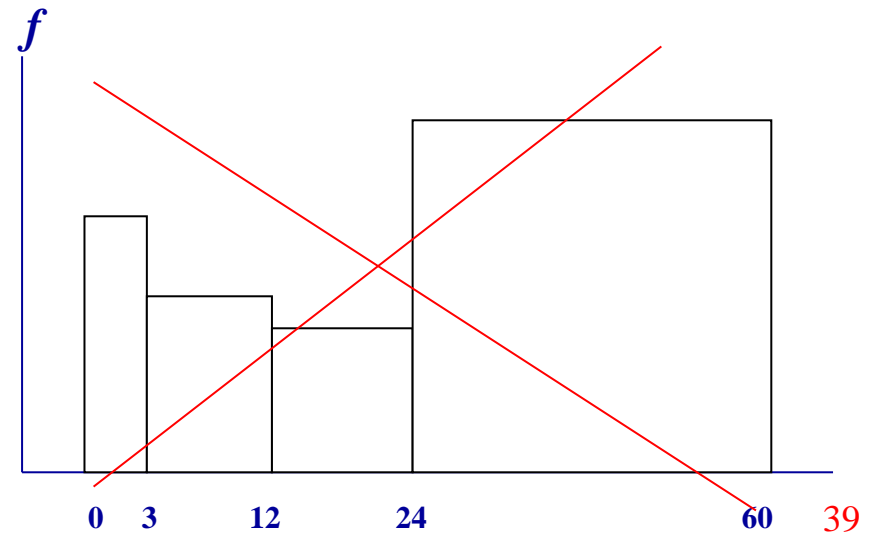
**Distribuição das idades (em meses) de
uma amostra de 500 crianças vacinadas**

Classes (meses)	f	fr	$base$	h
0 - 3	140	0,28	3	0,093
3 - 12	100	0,20	9	0,022
12 -24	80	0,16	12	0,013
24 -60	180	0,36	36	0,010
Total	500	1,00		

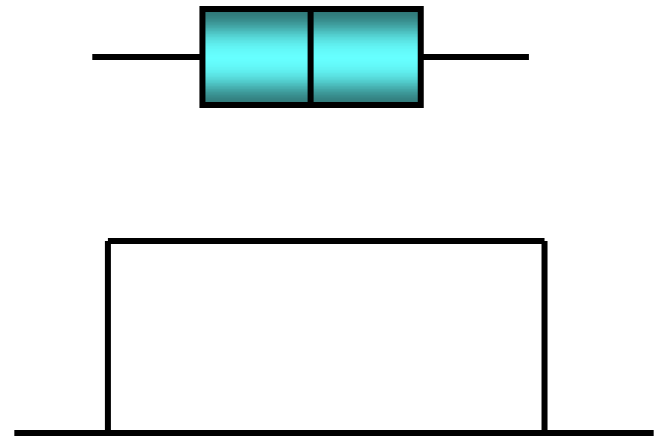
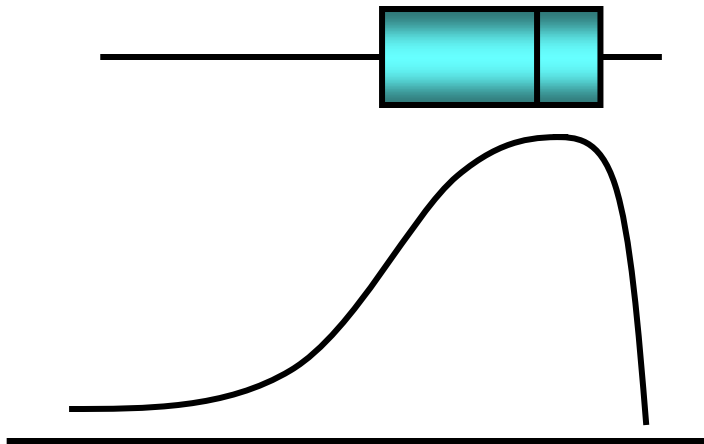
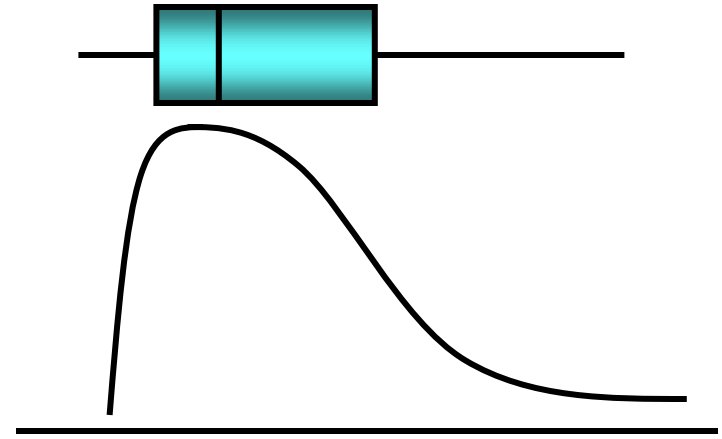
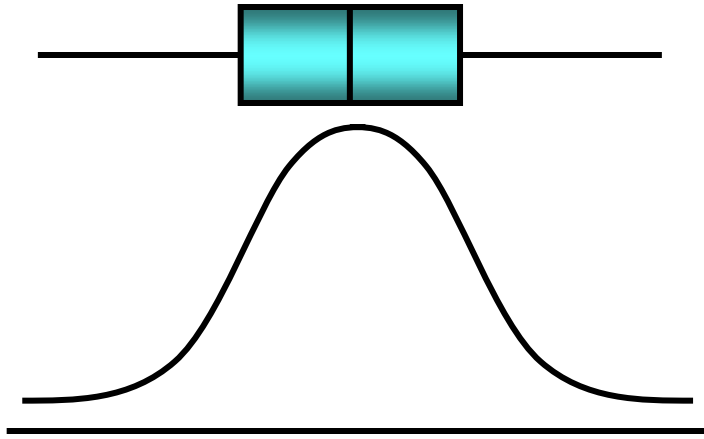
Histograma com Classes Desiguais



ERRADO!



Forma da Distribuição



Variáveis Qualitativas

Os dados podem ser resumidos por meio de tabelas de distribuição de frequências, que quantificam a frequência das distintas categorias.

Variáveis qualitativas do Projeto Qualidade de Vida

Etnia

Fumante

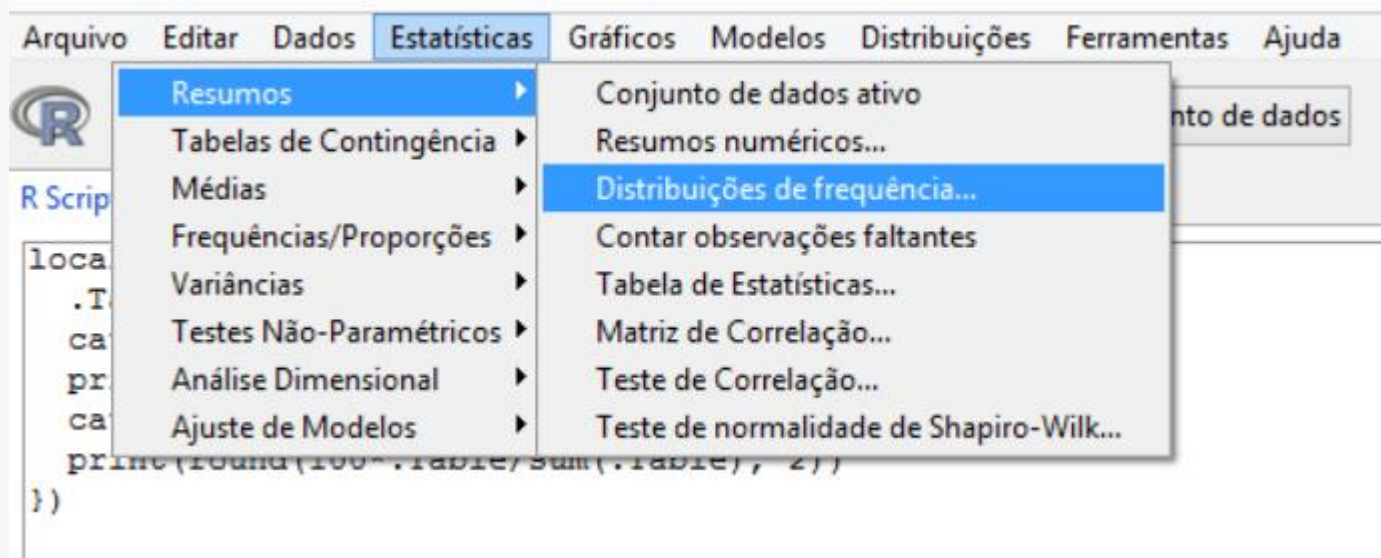
Escolaridade

Estado Civil

Inicialmente, converter variável numérica para Fator



Cálculo das frequências observadas e relativas



Variáveis Qualitativas – Projeto Qualidade de Vida

counts: **HabFumar**

Não Fumante	Fumante
19	11

percentages: **HabFumar**

Não Fumante	Fumante
63.33	36.67

counts: **Escolaridade**

1	2	3	4	5	6
20	2	1	4	2	1

percentages: **Escolaridade**

1	2	3	4	5	6
66.67	6.67	3.33	13.33	6.67	3.33

counts: **EstCivil**

1	2	3	4
16	6	5	3

percentages: **EstCivil**

1	2	3	4
53.33	20.00	16.67	10.00

counts: **Etnia**

1	2	3
25	1	4

percentages: **Etnia**

1	2	3
83.33	3.33	13.33

Variáveis Qualitativas

Gráficos

- Gráfico de setores
- Gráfico de barras

GRÁFICO DE SETORES

Um círculo é dividido em tantos setores quantas forem as categorias da variável.

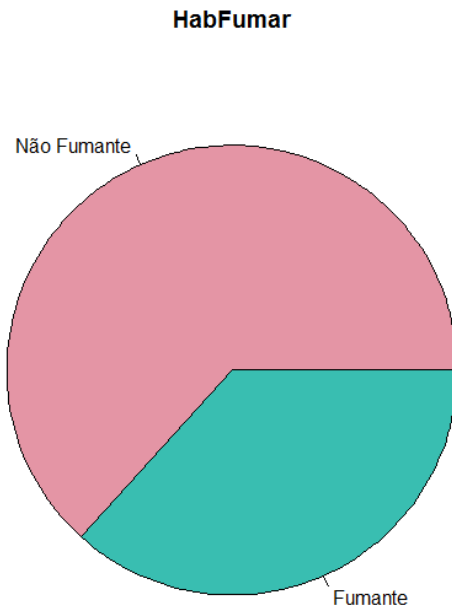
A área de cada setor é proporcional à frequência da categoria.

Rcmdr: Gráficos → Gráfico de Pizza

Projeto *QV*- Gráfico de setores para a **variável** **Hábito de Fumar (HabFumar)**

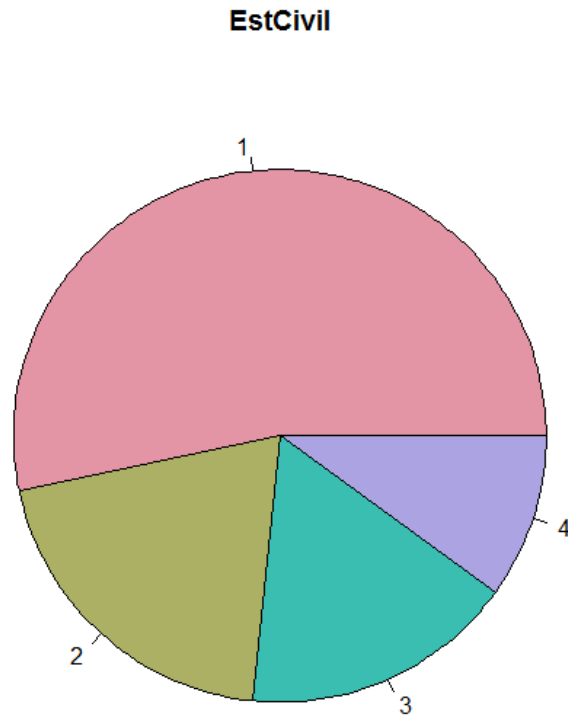
HabFumar

	Frequências Absolutas	Frequências Relativas
Não Fumante	19	63,33
Fumante	11	36,67



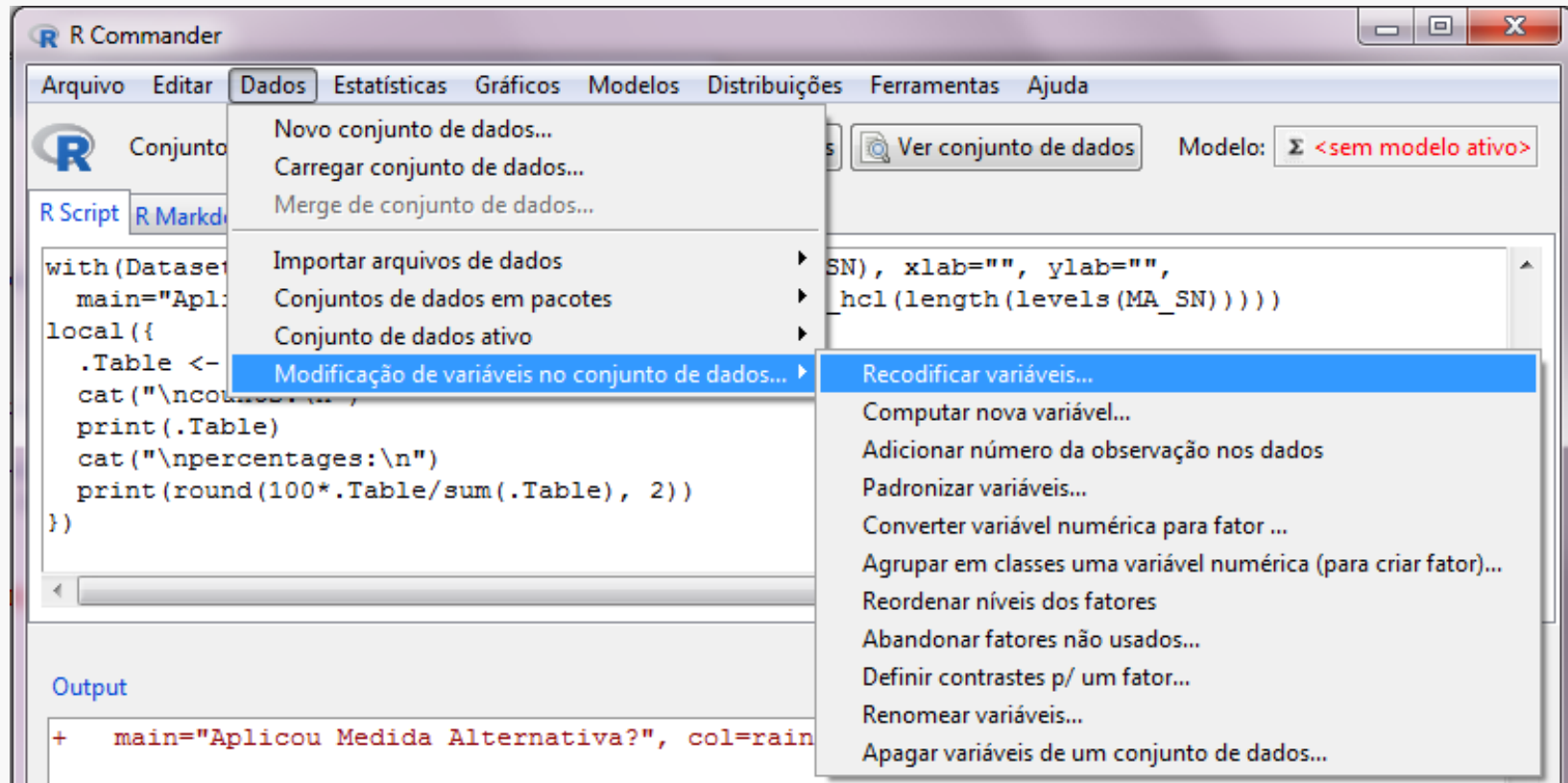
**Gráficos →
Gráfico de Pizza**

Projeto QV- Gráfico de setores para a **variável** **Estado Civil**



Observe que as
categorias estão
representadas por
números

Renomeando fatores para o gráfico de setores



Renomeando fatores para o gráfico de setores

Recodificar variáveis

Variável(s) p/ recodificar (uma ou mais)

- Altura
- Escolaridade
- EstCivil
- Etnia
- Fumante
- Idade

Novo nome de variável para recodificação múltipla: EstadoCivil

☒ Faça de cada nova variável um fator

Definições p/ recodificação

"1"="Casada"
"2"="Solteira"
"3"="Separada"
"4"="Viúva"

Ajuda Resetar OK Cancelar Apply

Projeto QV- Gráfico de setores para a **variável** **Estado Civil**

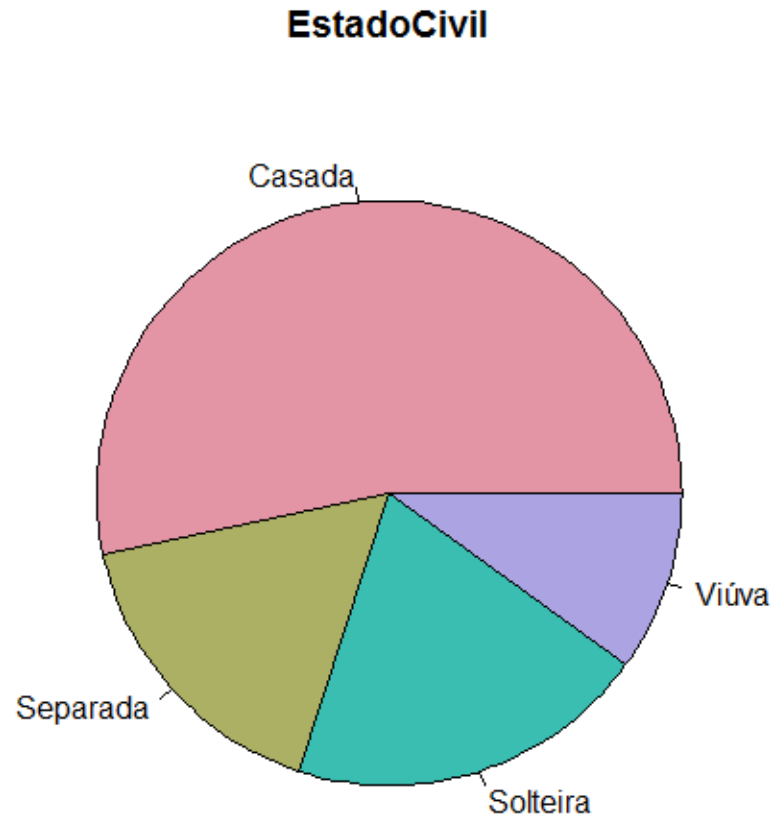


GRÁFICO DE BARRAS

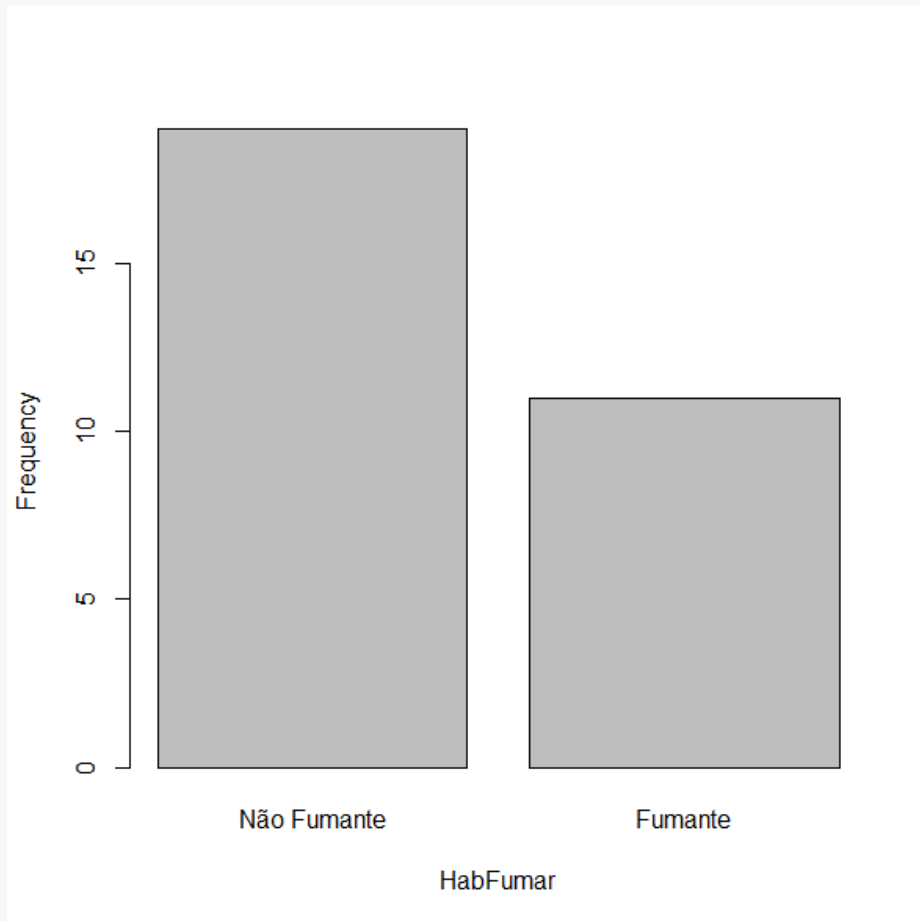
Sobre um eixo, são representados retângulos, um para cada categoria da variável.

A altura do retângulo é proporcional à frequência da categoria.

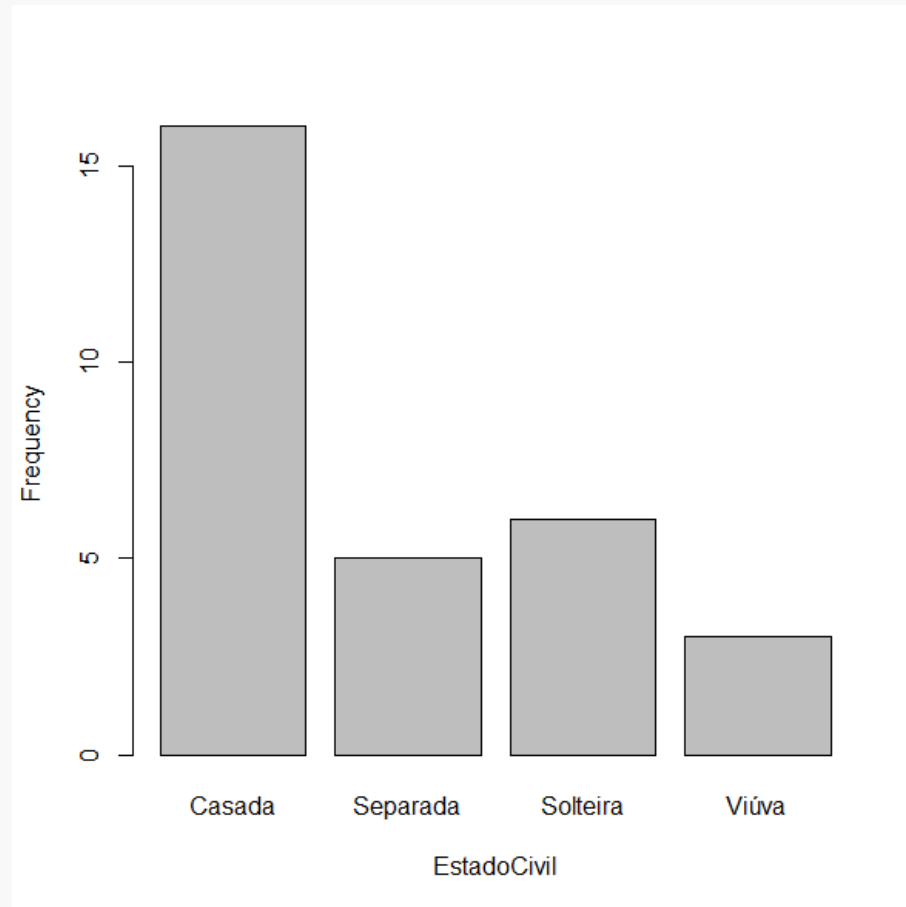
Rcmdr: Gráficos → Gráfico de Barras

Projeto QV- Gráfico de Barras

Hábito de Fumar



Estado Civil



Gráficos → Gráfico de Barras

RESUMO

