

# LABORATÓRIO 1: Regressão Linear Simples

Fernando Bispo

# Sumário

Questão 1 . . . . .	1
Running Code . . . . .	2
Questão 2 . . . . .	2
Item G . . . . .	4

```
<!-- html: -->
<!-- toc: true -->
<!-- toc-title: Sumário -->
<!-- colorlinks: true -->
<!-- documentclass: report -->
<!-- papersize: letter -->
<!-- number-sections: false -->
<!-- lang: pt-BR -->
  <!-- # geometry: -->
  <!-- # - top=30mm -->
  <!-- # - left=30mm -->
  <!-- # - right=20mm -->
  <!-- # - bottom=20mm -->
  <!-- # - heightrounded -->
  <!-- # fig-pos: "H" -->
  <!-- # fig-align: center -->
  <!-- # fontsize: 12pt -->
```

## Questão 1

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

- sexo (**sex**);
- largura do crânio (**skullw**);
- comprimento total (**totlngth**).

Tabela 1: Tabela 1: Apresentação parcial do conjunto de dados.

pregnant	glucose	diastolic	triceps	insulin	bmi	diabetes	age	test
6	148	72	35	0	33,6	0,627	50	1
1	85	66	29	0	26,6	0,351	31	0
8	183	64	0	0	23,3	0,672	32	1
1	89	66	23	94	28,1	0,167	21	0
0	137	40	35	168	43,1	2,288	33	1
5	116	74	0	0	25,6	0,201	30	0

Tabela 2: Tabela 1: Medidas Resumo para o sexo feminino.

	Min	Q1	Med	Média	Q3	Max	Desvio Padrão	CV
age	21,00	24,00	29,00	33,24	41,00	81,00	11,76	0,35
bmi	0,00	27,30	32,00	31,99	36,60	67,10	7,88	0,25
diabetes	0,08	0,24	0,37	0,47	0,63	2,42	0,33	0,70
diastolic	0,00	62,00	72,00	69,11	80,00	122,00	19,36	0,28
glucose	0,00	99,00	117,00	120,89	140,50	199,00	31,97	0,26
insulin	0,00	0,00	30,50	79,80	127,50	846,00	115,24	1,44
pregnant	0,00	1,00	3,00	3,85	6,00	17,00	3,37	0,88
test	0,00	0,00	0,00	0,35	1,00	1,00	0,48	1,37
triceps	0,00	0,00	23,00	20,54	32,00	99,00	15,95	0,78

## Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

## Questão 2

Para a resolução da segunda parte da atividade foi disponibilizado um conjunto de dados originalmente do *National Institute of Diabetes and Digestive and Kidney Diseases*. (Instituto Nacional de Diabetes e Doenças Digestivas e Renais). O objetivo do conjunto de dados é prever diagnosticamente se um paciente tem ou não diabetes, com base em determinadas medidas de diagnóstico incluídas no conjunto de dados. Várias restrições foram impostas à seleção dessas instâncias em um banco de dados maior. Em particular, todos os pacientes aqui são mulheres com pelo menos 21 anos de idade, de ascendência indígena Pima.

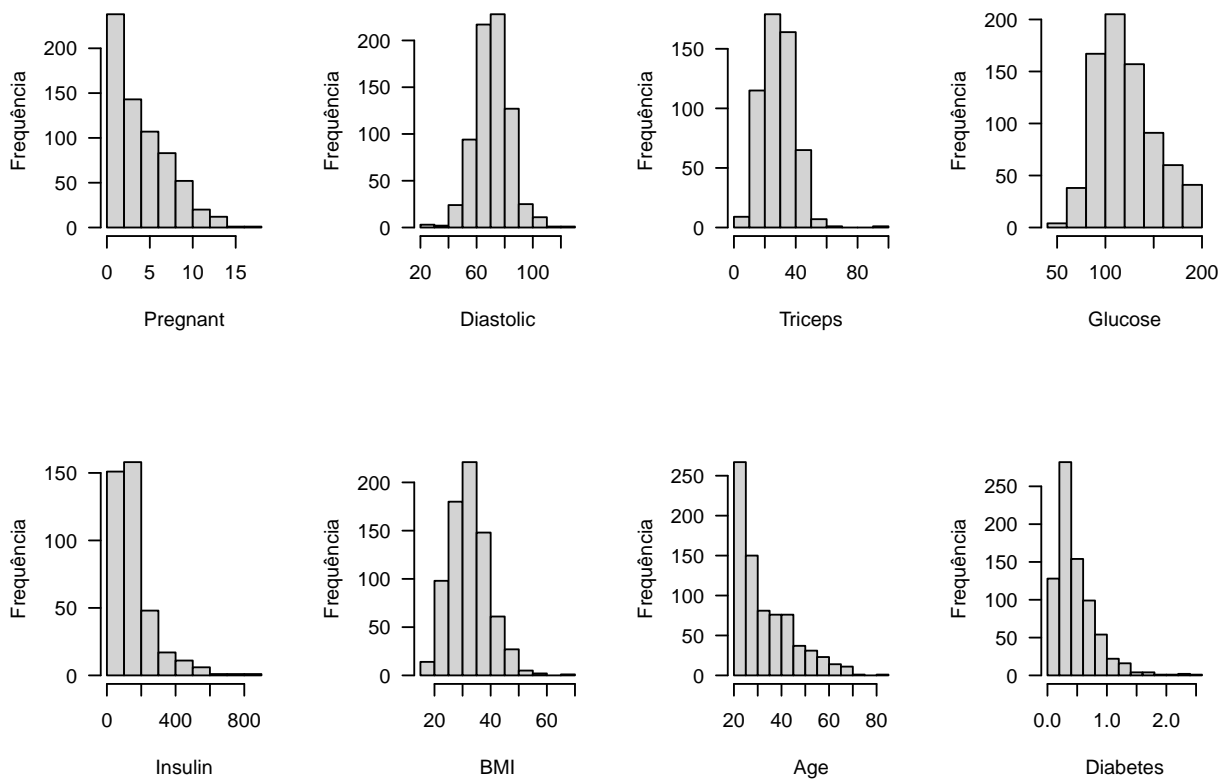
A fim de conhecer o conjunto de dados, segue a Tabela 2 contendo as principais medidas resumo do conjunto de dados disponibilizado.

Tabela 3: Tabela 1: Medidas Resumo para o sexo feminino.

	Min	Q1	Med	Média	Q3	Max	Desvio Padrão	CV
age	21,00	24,00	29,00	33,24	41,00	81,00	11,76	0,35
bmi	18,20	27,50	32,30	32,46	36,60	67,10	6,92	0,21
diabetes	0,08	0,24	0,37	0,47	0,63	2,42	0,33	0,70
diastolic	24,00	64,00	72,00	72,41	80,00	122,00	12,38	0,17
glucose	44,00	99,00	117,00	121,69	141,00	199,00	30,54	0,25
insulin	14,00	76,00	125,00	155,55	190,00	846,00	118,78	0,76
pregnant	1,00	2,00	4,00	4,49	7,00	17,00	3,22	0,72
test	0,00	0,00	0,00	0,35	1,00	1,00	0,48	1,37
triceps	7,00	22,00	29,00	29,15	36,00	99,00	10,48	0,36

Levando em consideração o fato de haverem muitos dados faltantes constatou-se que esses dados foram substituídos por valores zero, o que influencia negativamente na interpretação dos resultados das principais medidas resumo calculadas, logo essas observações serão removidas e novos cálculos serão realizados a fim de se obter medidas mais fidedignas da realidade.

É importante ressaltar que a variável *test* é uma variável categórica, logo não faz sentido haver interpretações acerca das medidas resumo dessa variável, sendo expressa a Figura X para possibilitar uma melhor visualização do comportamento dessa variável.



## Item G

Construção de boxplots comparando a distribuição das variáveis por grupo de resultado do teste de diabetes: