

# PRÁTICA 13: REGRESSÃO LOGÍSTICA E LINEAR APLICADAS

## Parte 1: “Entendendo” os dados

- 1) Quais das 13 variáveis são quantitativas e quais delas são categóricas?
  - a. **Categóricas:** Tipo de vinho (branco ou tinto) e Vinho de boa qualidade (sim ou não)
  - b. **Quantitativas:** Acidez fixa, Acidez volátil, Acido citrico, Acucar residual, Cloretos, Dioxido de enxofre livre, Dioxido de enxofre total, Densidade, pH, Sulfatos e Concentracao final de alcool
- 2) Considere a variável “*Vinho de boa qualidade*” como desfecho ou variável resposta ou variável dependente ou variável de interesse. Neste cenário, quais são as variáveis explicativas ou independentes?
  - a. **Resposta:** Tipo de vinho (branco ou tinto), Acidez fixa, Acidez volátil, Acido citrico, Acucar residual, Cloretos, Dioxido de enxofre livre, Dioxido de enxofre total, Densidade, pH, Sulfatos e Concentracao final de alcool
- 3) Considere agora “*Concentracao final de alcool*” como desfecho, quais são as variáveis explicativas?
  - a. **Resposta:** Tipo de vinho (branco ou tinto), Acidez fixa, Acidez volátil, Acido citrico, Acucar residual, Cloretos, Dioxido de enxofre livre, Dioxido de enxofre total, Densidade, pH, Sulfatos, Vinho de boa qualidade (sim ou não)

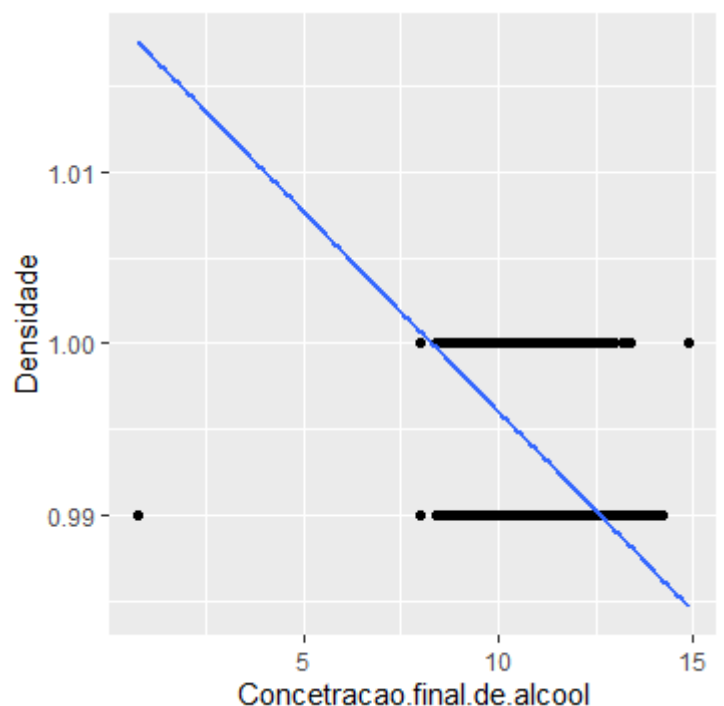
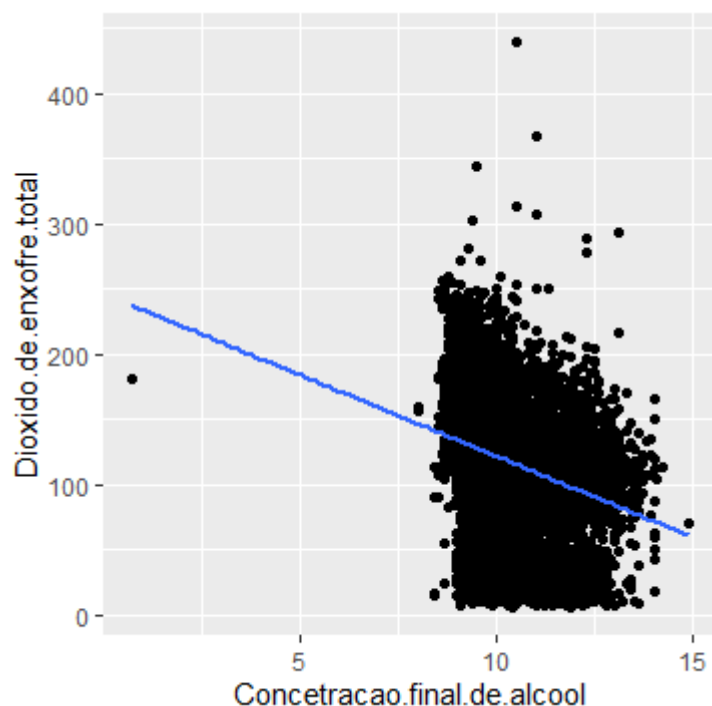
## Parte 2: Modelo de Regressão Linear Múltipla

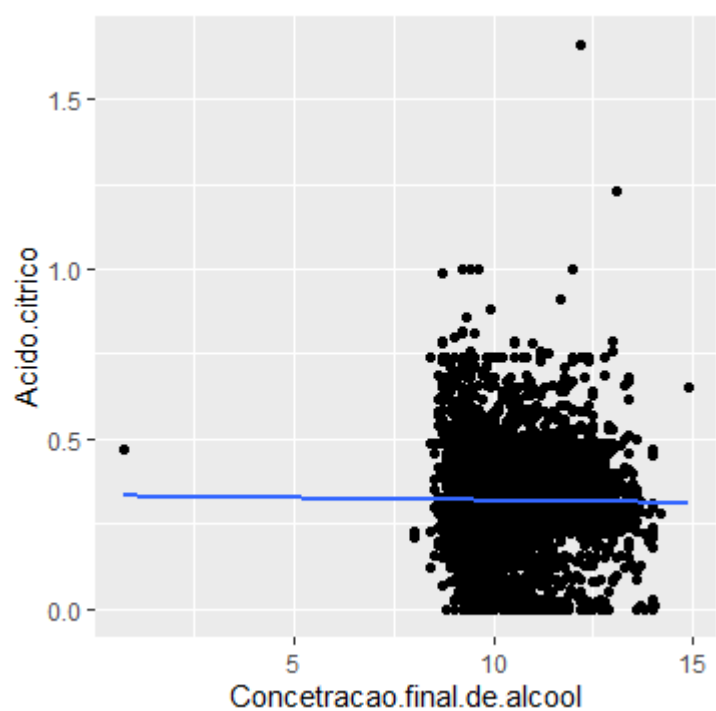
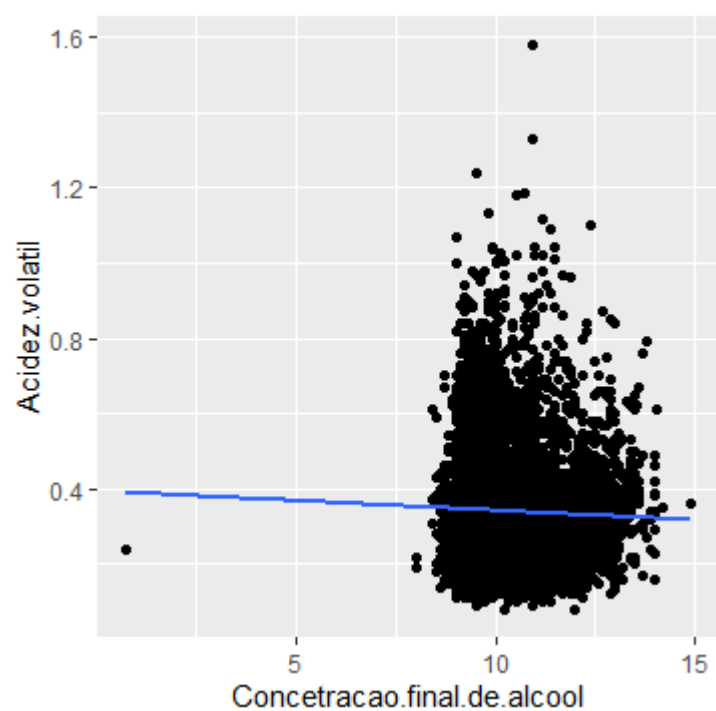
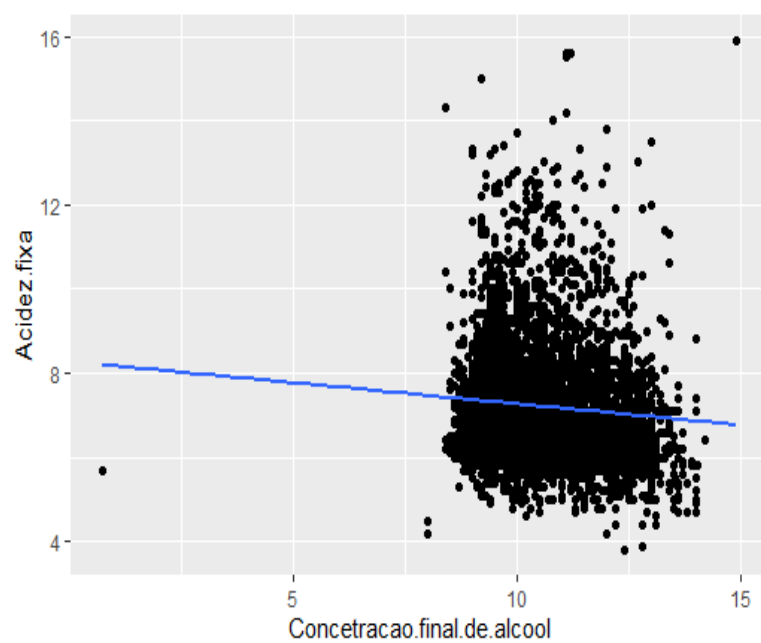
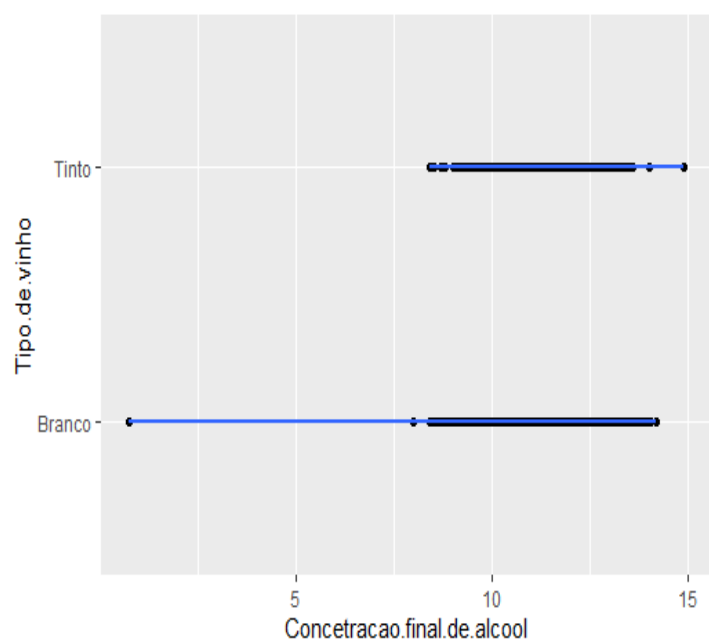
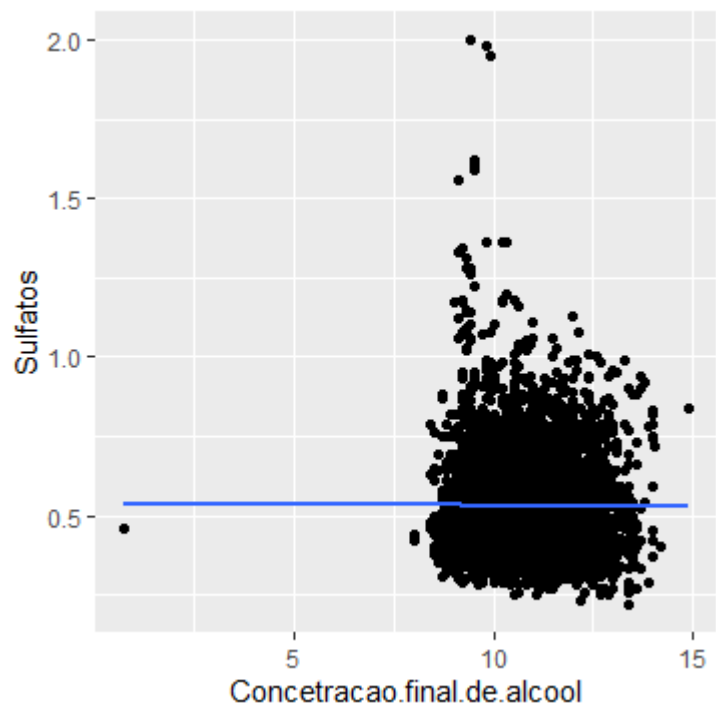
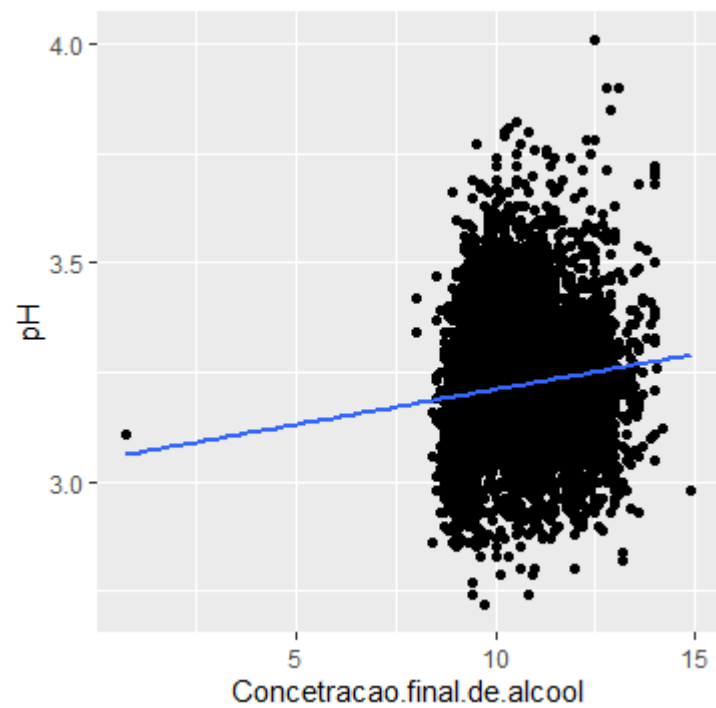
- 1) Ler os dados e fazer um resumo:  
# Resumo estatístico dos dados

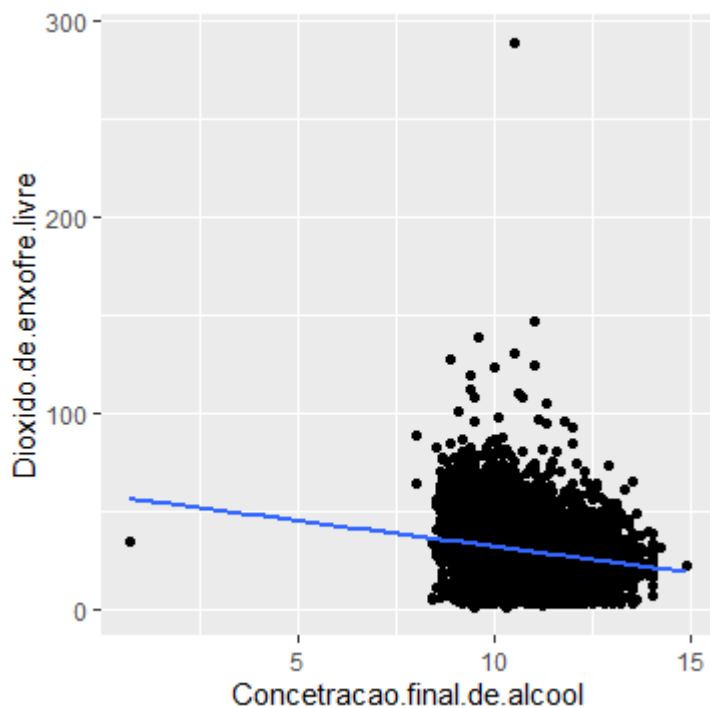
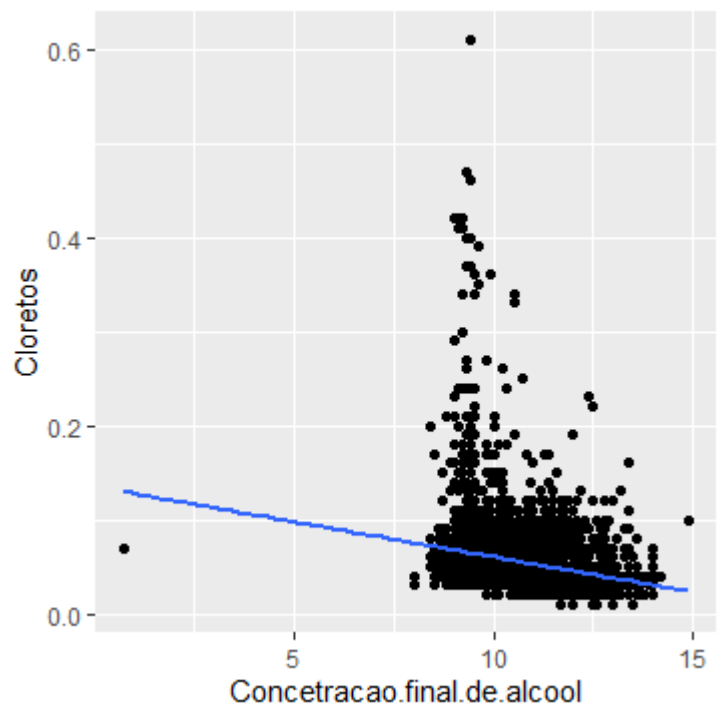
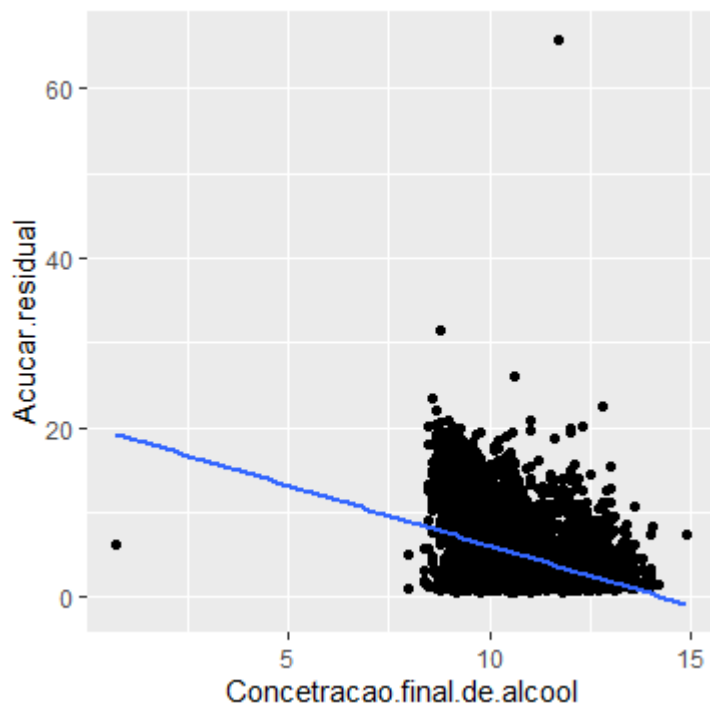
```
> summary(Dados_Vinho)
Tipo.de.vinho  Acidez.fixa  Acidez.volatil  Acido.citrico  Acucar.residual  Cloretos  Dioxido.de.enxofre.livre
Branco:4898   Min.    : 3.800    Min.    :0.0800  Min.    :0.0000  Min.    : 0.600  Min.    :0.0100  Min.    : 1.00
Tinto :1593   1st Qu.: 6.400    1st Qu.:0.2300  1st Qu.:0.2500  1st Qu.: 1.800  1st Qu.:0.0400  1st Qu.: 17.00
              Median : 7.000    Median :0.2900  Median :0.3100  Median : 3.000  Median :0.0500  Median : 29.00
              Mean   : 7.214    Mean   :0.3397  Mean   :0.3187  Mean   : 5.446  Mean   :0.0565  Mean   : 30.54
              3rd Qu.: 7.700    3rd Qu.:0.4000  3rd Qu.:0.3900  3rd Qu.: 8.100  3rd Qu.:0.0700  3rd Qu.: 41.00
              Max.   :15.900    Max.   :1.5800  Max.   :1.6600  Max.   :65.800  Max.   :0.6100  Max.   :289.00

Dioxido.de.enxofre.total  Densidade  pH  Sulfatos  Concentracao.final.de.alcool  vinho.de.boa.qualidade
Min.    : 6.0              Min.    :0.9900  Min.    :2.720  Min.    :0.2200  Min.    : 0.73  Nao:5215
1st Qu.: 78.0             1st Qu.:0.9900  1st Qu.:3.110  1st Qu.:0.4300  1st Qu.: 9.50  Sim:1276
Median :118.0             Median :0.9900  Median :3.210  Median :0.5100  Median :10.30
Mean   :115.8             Mean   :0.9949  Mean   :3.218  Mean   :0.5311  Mean   :10.49
3rd Qu.:156.0             3rd Qu.:1.0000  3rd Qu.:3.320  3rd Qu.:0.6000  3rd Qu.:11.30
Max.   :440.0             Max.   :1.0000  Max.   :4.010  Max.   :2.0000  Max.   :14.90
```

Construa diagramas de dispersão e avalie a possível relação entre as variáveis explicativas em relação ao desfecho *Concentracao final de alcool*.







2) Faça a regressão linear múltipla, considerando o desfecho *Concetracao final de álcool*.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.508e+02	3.575e+00	42.193	< 2e-16	***
Tipo.de.vinhoTinto	3.057e-02	6.290e-02	0.486	0.627006	
Acidez.fixa	4.504e-02	1.307e-02	3.445	0.000575	***
Acidez.volatil	8.236e-01	9.850e-02	8.362	< 2e-16	***
Acido.citrico	9.624e-01	9.676e-02	9.946	< 2e-16	***
Acucar.residual	1.629e-02	3.587e-03	4.542	5.67e-06	***
Cloretos	-7.880e+00	3.931e-01	-20.046	< 2e-16	***
Dioxido.de.enxofre.livre	5.078e-03	9.304e-04	5.458	4.99e-08	***
Dioxido.de.enxofre.total	-6.733e-03	3.854e-04	-17.469	< 2e-16	***
Densidade	-1.439e+02	3.670e+00	-39.200	< 2e-16	***
pH	7.543e-01	8.835e-02	8.538	< 2e-16	***
Sulfatos	7.801e-01	9.011e-02	8.658	< 2e-16	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

a. Responda às perguntas:

- O tipo de vinho (tinto versus branco) afeta a sua concentração alcóolica final? Não
- Quais variáveis têm correlação positiva significativa com a concentração alcóolica final do vinho de tal forma que, quando esta variável aumenta, o teor alcóolico do vinho também

aumenta? Analisando o “Estimate”: Cloretos e densidade

- Quais variáveis têm correlação negativa significativa com a concentração alcóolica final do vinho de tal forma que, quando esta variável aumenta, o teor alcóolico do vinho também aumenta? Analisando o “Estimate”: Tipo de vinho, acidez fixa, acidez volátil, acido cítrico, acido residual, Dioxido de enxofre livre, Dioxido de enxofre total, ph e sulfatos

### Parte 3: Modelo de Regressão Logística

a) Construa um modelo de regressão para a qualidade do vinho.

```
> table(modelo.pred, vinho.de.boa.qualidade)
      vinho.de.boa.qualidade
modelo.pred  Nao  Sim
      Nao 4982  953
      Sim  233  323
> mean(modelo.pred == vinho.de.boa.qualidade)
[1] 0.8172855
```

Analisando a matriz de confusão verificamos que, em 4982 casos foi dito que o vinho não era de boa qualidade e realmente não era, em contrapartida tivemos erro em 953 casos onde foi dito que eles eram de boa qualidade porém na realidade não são. 233 foi o numero de casos onde foi dito que o vinho era de boa qualidade porém na realidade não são e 323 casos o vinho é de boa qualidade e foi classificado de forma correta, Analisando a taxa de erro, tivemos uma media de 81% de acertos.