

# Regressão Linear

Bioestatística em R

---

André M Ribeiro-dos-Santos

10 de Abr, 2017

- Modelar relação entre variáveis através de uma regressão linear.
- Compreender os principais modelos de regressão linear.
- Compreender os coeficientes e seu significado.
- Ilustrar a regressão e resíduos.
- Comparar diferentes modelos de regressão.

## Regressão Linear

---

Em um estudo sobre diabetes, pesquisadores observaram uma relação entre a variação da sensibilidade à insulina e o percentual de ácido graxos entre os pacientes.

## Como qual intensidade o percentual de ácidos graxos afeta a sensibilidade à insulina?

**Table 1:** Medidas de sensibilidade à insulina e composição de ácido graxos em diabéticos

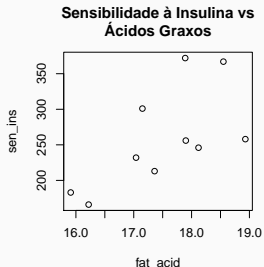
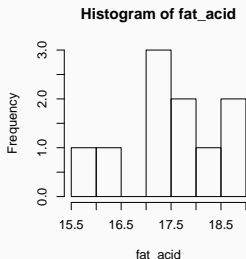
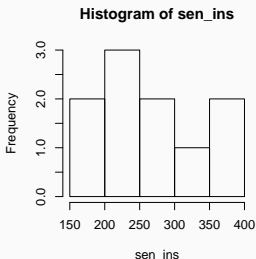
Sensibilidade à Insulina	Ácidos Graxos (%)	Sensibilidade à Insulina	Ácidos Graxos (%)
183	15.91	246	18.12
232	17.04	256	17.90
166	16.22	372	17.89
258	18.93	367	18.55
213	17.36	301	17.15

- As medidas em questão são categóricas ou quantitativas?
- Qual a informação deseja-se obter?
- Qual a distribuição das medidas?
- Qual a relação entre as variáveis?

- As medidas em questão são categóricas ou quantitativas? **Ambas são medidas quantitativas**
- Qual a informação deseja-se obter? **A intensidade que a variação do percentual de ácidos graxos afeta a sensibilidade à insulina altera**

- Qual a distribuição das medidas? **Distribuição Aproximadamente Normal**
- Qual a relação entre as variáveis? **Relação linear**

```
> sen_ins <- c(183, 232, 166, 258, 213, 246, 256, 372, 367, 301)
> fat_acid <- c(15.91, 17.04, 16.22, 18.93, 17.36, 18.12, 17.90,
+               17.89, 18.55, 17.15)
> par(mfrow = c(1,3))
> hist(sen_ins)
> hist(fat_acid)
> plot(fat_acid, sen_ins, main="Sensibilidade à Insulina vs\nÁcidos Graxos")
```



Pode-se *modelar* a relação entre duas variáveis numéricas através da equação a seguir:

$$y = \alpha + \beta x + \epsilon \text{ ou } Y = X\beta + \epsilon$$

- $\alpha$ : o valor basal de  $y$ , quando  $x$  apresenta valor 0.
- $\beta$ : a mudança em  $y$  para cada unidade de  $x$ .
- $\epsilon$ : o erro da predição ou variação aleatória.

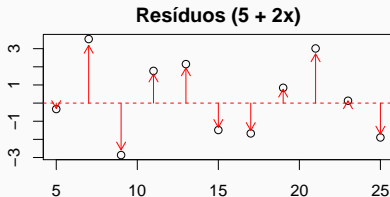
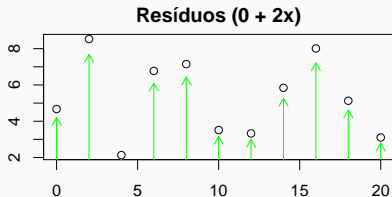
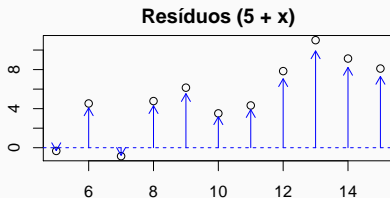
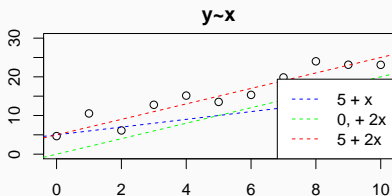
Quando não existe relação entre as variáveis,  $\beta$  segue uma *distribuição t* e podemos utilizar isso para testar a associação das variáveis  $x$  e  $y$ .

$$H_0 : \beta = 0; \quad H_a : \beta \neq 0$$



Como identificar os coeficientes ( $\alpha$  e  $\beta$ ) na relação:

$$y = \alpha + \beta * x + \epsilon$$



O objetivo é minimizar a soma dos erros ou resíduos ( $\epsilon$ ), para tanto existem várias técnicas (e.g. *least-squares*, *ordinal least-square*, *gradiente descendente*, ...).

Para regressões lineares, usamos a função `lm`.

```
> ?lm
> ## Fitting Linear Models
> ## Description:
> ##      'lm' is used to fit linear models. It can be used to carry
> ##      out regression, single stratum analysis of variance and
> ##      analysis of covariance (although 'aov' may provide a more
> ##      convenient interface for these).
> ## Usage:
> ##      lm(formula, data, subset, weights, na.action,
> ##          method = "qr", model = TRUE, x = FALSE, y = FALSE,
> ##          qr = TRUE, singular.ok = TRUE, contrasts = NULL,
> ##          offset, ...)
```

Portanto, modelando a relação entre sensibilidade à insulina e o percentual de ácidos graxos. Obtemos:

```
> (model <- lm(sen_ins~fat_acid))  
  
##  
## Call:  
## lm(formula = sen_ins ~ fat_acid)  
##  
## Coefficients:  
## (Intercept)      fat_acid  
##      -558.62         46.73
```

Corresponde a equação:

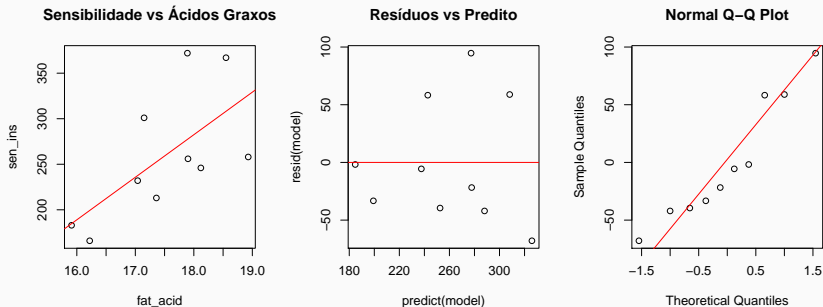
$$\text{sensibilidade insulina} = -558.62 + 46.73 * \text{acido graxo}$$

Indicando que o aumento de 1% do percentual de ácidos graxos resulta no aumento de 46,73 da sensibilidade à insulina.

```

> par(mfrow = c(1,3))
> plot(sen_ins~fat_acid, main = "Sensibilidade vs Ácidos Graxos")
> abline(model, col='red')      ## ou abline(-558.62, 46.73)
> plot(predict(model), resid(model), main="Resíduos vs Predito")
> abline(0, 0, col='red')
> qqnorm(resid(model))          ## qqnorm
> qqline(resid(model), col='red') ## ou plot(model, which=1:2)

```



Para avaliar se a relação obtida é estatisticamente significativa ou espúria, utilizamos a função `summary(model)`.

```
> ?summary
> ## Summarizing Linear Model Fits
> ## Description:
> ##      'summary' method for class '"lm"'.
> ## Usage:
> ##      ## S3 method for class 'lm'
> ##      summary(object, correlation = FALSE, symbolic.cor = FALSE,
> ##              ...)
> ##      ## S3 method for class 'summary.lm'
> ##      print(x, digits = max(3, getOption("digits") - 3),
> ##            symbolic.cor = x$symbolic.cor,
> ##            signif.stars = getOption("show.signif.stars"), ...)
```

```
> summary(model)
```

```
##
## Call:
## lm(formula = sen_ins ~ fat_acid)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -67.89 -37.97 -13.67  43.27  94.70
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -558.62     341.47  -1.636   0.1405
## fat_acid       46.73       19.48   2.399   0.0433 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.33 on 8 degrees of freedom
## Multiple R-squared:  0.4184, Adjusted R-squared:  0.3457
## F-statistic: 5.755 on 1 and 8 DF,  p-value: 0.04325
```

**Table 2:** Funções exploratórias

Função	Retorno
<code>coef</code>	Tabela com os coeficientes e resultados do teste quando <code>summary.lm</code>
<code>resid</code>	Vetor dos resíduos ( $y - \hat{y}$ ) para da valor
<code>predict</code>	Vetor dos valores preditos ( $\hat{y}$ ) para cada valor

## Exercício - Regressão Linear Simples

Pesquisadores buscando entender a elevada variação da sensibilidade à insulina (`sen_ins`) observada numa amostra de 50 diabéticos. Trabalhos anteriores relacionaram esta variação ao teor de colesterol (`chl`) e ao percentual de lipídios (`fat_acid`) no tecido muscular.

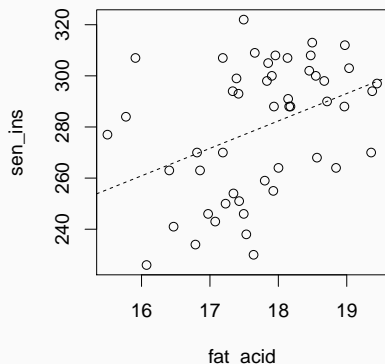
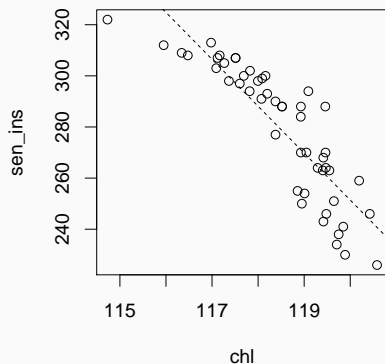
1. Ilustre a relação entre `sen_ins` com `chl` e `fat_acid`.
2. Aplique uma regressão linear e explique como as variáveis `chl` e `fat_acid` se relacionam com `sen_ins`. Avalie se a relação entre as medidas é real ou aleatória.
3. Faça os principais gráficos de diagnóstico (scatter, resíduos e qqplot) para as regressões anteriores. Para alguma das regressões não seria recomendado o modelo utilizado? Justifique.
4. Existe alguma estratégia que você recomendaria para tratar deste caso?

```
> db <- read.table('db.tsv', header=T)
```



1. Ilustre a relação entre `sen_ins` com `chl` e `fat_acid`.

```
> par(mfrow = c(1,2))  
> plot(sen_ins~chl, data=db)  
> abline(model_chl <- lm(sen_ins~chl, data=db), lty=2)  
> plot(sen_ins~fat_acid, data=db)  
> abline(model_fat <- lm(sen_ins~fat_acid, data=db), lty=2)
```



2. Aplique uma regressão linear e explique como as variáveis `chl` e `fat_acid` se relacionam com `sen_ins`. Avalie se a relação entre as medidas é real ou aleatória.

```
> coef(summary(model_chl))
```

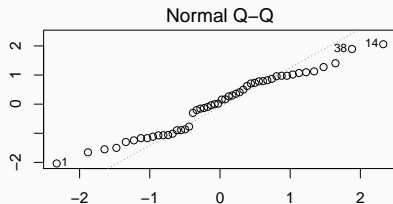
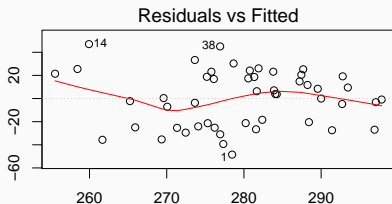
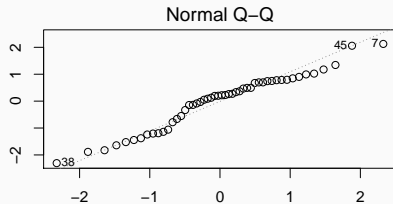
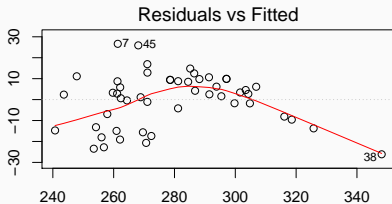
##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	2453.76745	175.342711	13.99412	1.458882e-18
## chl	-18.35297	1.479988	-12.40076	1.406545e-16

```
> coef(summary(model_fat))
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	89.14280	64.79262	1.375817	0.175264336
## fat_acid	10.73416	3.64859	2.942003	0.005008373

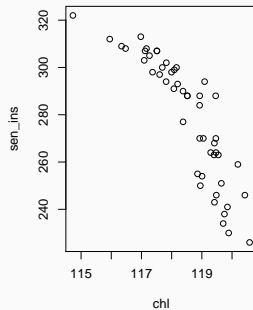
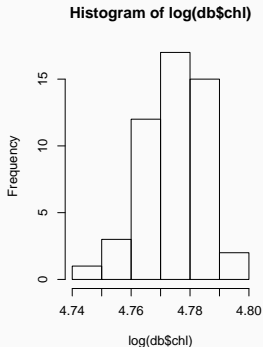
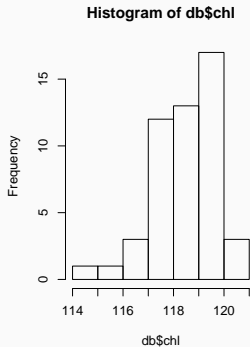
3. Faça os principais gráficos de diagnóstico (scatter, resíduos e qqplot) para as regressões anteriores. Para alguma das regressões não seria recomendado o modelo utilizado? Justifique.

```
> par(mfrow = c(2,2), mar=rep(2,4))
> plot(model_chl, which = c(1,2))
> plot(model_fat, which = c(1,2))
```



4. Existe alguma estratégia que você recomendaria para tratar deste caso?

```
> par(mfrow = c(1,3))
> hist(db$chl)
> hist(log(db$chl))
> plot(sen_ins~chl, data=db)
```





Outro fator de risco para diabetes é a pressão sanguínea, a qual é relacionada a fatores como índice de massa corpórea (IMC ou BMI) e o teor de colesterol.

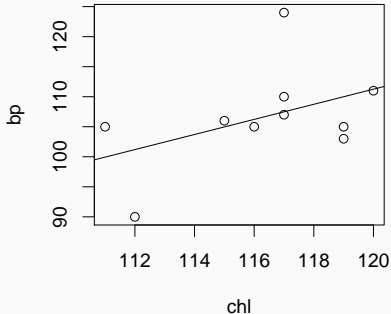
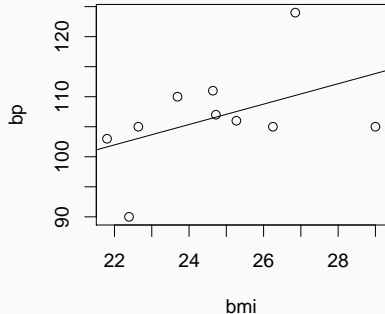
### Qual o efeito dessas variáveis sobre a pressão sanguínea?

**Table 3:** IMC, Teor de Colesterol e Pressão Sanguínea em Diabeticos

IMC	Colesterol	Pressão Sanguínea	IMC	Colesterol	Pressão Sanguínea
26.85	117	124	21.80	119	103
24.64	120	111	23.69	117	110
22.46	119	105	24.72	115	107
22.39	112	90	26.25	115	105
29.00	112	105	25.27	115	106

Continuamos avaliando a **relação entre variáveis quantitativas** de forma a **modelar seus efeitos**, falta verificar se existe uma relação linear entre as medidas.

```
> bmi <- c(26.85, 24.64, 22.64, 22.39, 29.00, 21.80, 23.69,  
+         24.72, 26.25, 25.27)  
> chl <- c(117, 120, 119, 112, 111, 119, 117, 117, 116, 115)  
> bp <- c(124, 111, 105, 90, 105, 103, 110, 107, 105, 106)  
> par(mfrow = c(1,2))  
> plot(bp~bmi); abline(model_bmi <- lm(bp~bmi))  
> plot(bp~chl); abline(model_chl <- lm(bp~chl))
```



Os modelos são estatisticamente significativos?

```
> coef(summary(model_bmi))
```

```
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 64.645050   29.332219  2.203892 0.05863691
## bmi         1.696863    1.182011  1.435574 0.18904513
```

```
> summary(model_bmi)$adj.r.squared
```

```
## [1] 0.1054453
```

```
> coef(summary(model_chl))
```

```
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -39.631242 104.7869674 -0.3782077 0.7151147
## chl          1.257362   0.9007457  1.3959127 0.2002616
```

```
> summary(model_chl)$adj.r.squared
```

```
## [1] 0.09534757
```

```
> summary(model_both <- lm(bp ~ bmi + chl))
```

```
##
```

```
## Call:
```

```
## lm(formula = bp ~ bmi + chl)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -5.6104 -2.2335 -1.0212  0.6937  9.2192
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -239.3843     83.5265  -2.866  0.02413 *
```

```
## bmi           3.0859      0.8206   3.760  0.00707 **
```

```
## chl           2.3189      0.6219   3.729  0.00737 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

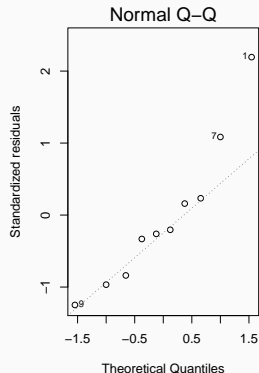
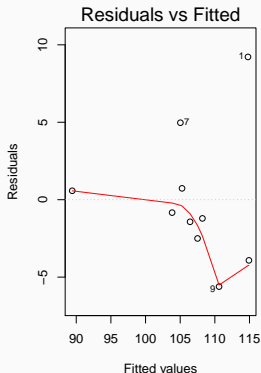
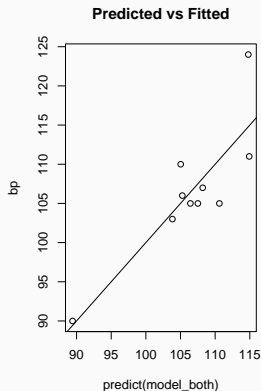
```
##
```

```
## Residual standard error: 4.897 on 7 degrees of freedom
```

```
## Multiple R-squared:  0.7337, Adjusted R-squared:  0.6577
```

```
## F-statistic: 9.645 on 2 and 7 DF, p-value: 0.00974
```

```
> par(mfrow = c(1, 3))  
> plot(predict(model_both), bp, main="Predicted vs Fitted")  
> abline(0, 1)  
> plot(model_both, which = 1:2)
```



Usando apenas BMI, o modelo obteve um  $R^2$  ajustado de ~ 10%, no entanto quando consideramos ambos BMI e colesterol obteve-se ~ 67%. **Podemos assumir que o modelo utilizando ambas variáveis é melhor do que o modelo com apenas BMI?** Para responder essa pergunta, devemos primeiro avaliar se a diferença entre os modelos é significativa ou aleatória.

```
> anova(model_bmi, model_both)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: bp ~ bmi
```

```
## Model 2: bp ~ bmi + chl
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1         8 501.27
```

```
## 2         7 167.85  1    333.42 13.905 0.007369 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como podemos comparar o impacto de cada medida para identificar o fator mais importante na regulação da pressão sanguínea? Uma vez que as variáveis estão em escalas diferentes, precisamos equalizar elas numa escala única (e.g. **z-scale**).

```
> model_scaled <- lm(bp~scale(bmi) + scale(chl))  
> coef(summary(model_scaled))
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	106.600000	1.548488	68.841326	3.587867e-11
## scale(bmi)	6.888660	1.831858	3.760478	7.070462e-03
## scale(chl)	6.830942	1.831858	3.728970	7.368697e-03

E só para provar que essa transformação não afeta o modelo em si.

```
> anova(model_both, model_scaled)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: bp ~ bmi + chl
```

```
## Model 2: bp ~ scale(bmi) + scale(chl)
```

```
##   Res.Df    RSS Df    Sum of Sq  F Pr(>F)
```

```
## 1      7 167.85
```

```
## 2      7 167.85  0 -1.7053e-12
```



Sabendo que o sexo do paciente (*feminino* ou *masculino*) afeta a pressão sanguínea média.

**A inclusão do sexo melhora a predição do modelo?**

**Table 4:** Pressão Sanguínea e Sexo em diabeticos

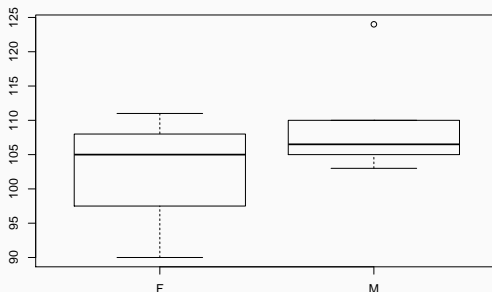
Pressão Sanguínea	Sexo	Pressão Sanguínea	Sexo
124	M	103	M
111	F	110	M
105	M	107	M
90	F	105	F
105	F	106	M

A primeira questão é avaliar se a pressão sanguínea difere entre os grupos (sexo masculino e feminino). Como estamos comparando a distribuição de uma variável quantitativa entre dois grupos, utilizamos um *teste t*.

```
> sex <- c('M', 'F', 'M', 'F', 'F', 'M', 'M', 'M', 'F', 'M')  
> t.test(bp~sex)$p.value
```

```
## [1] 0.2855744
```

```
> boxplot(bp~sex)
```



Já tendo observado a associação entre pressão sanguínea, BMI e Colesterol.

## Será o efeito do sexo mediado pelo BMI?

```
> model_add <- lm(bp ~ bmi + sex)
> coef(summary(model_add))
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	40.374916	26.329642	1.533440	0.16903533
## bmi	2.439385	1.021906	2.387093	0.04837505
## sexM	9.852134	4.417477	2.230263	0.06094724

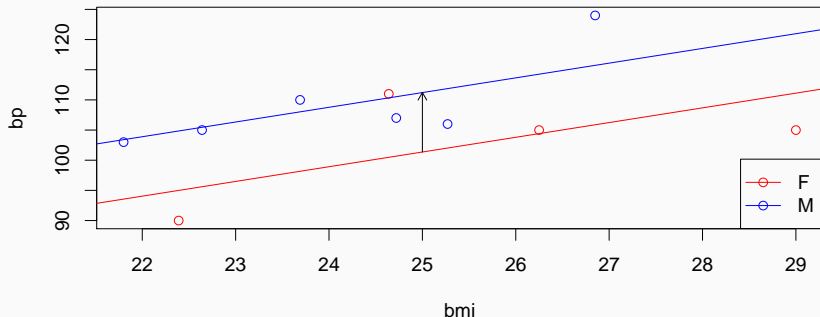
```
> summary(model_add)$adj.r.squared
```

```
## [1] 0.4023389
```

```

> cf <- coef(model_add)
> plot(bp~bmi, col = c(F='red', M='blue')[sex])
> abline(cf[1], cf[2], col='red')
> abline(cf[1] + cf[3], cf[2], col='blue')
> arrows(25, predict(model_add, data.frame(bmi=25, sex='F')),
+       25, predict(model_add, data.frame(bmi=25, sex='M')),
+       length=.1)
> legend("bottomright", c('F', 'M'), pch = 1, lty=1, col=c('red', 'blue'))

```



```
> summary(model_int <- lm(bp~bmi*sex))
```

```
##
```

```
## Call:
```

```
## lm(formula = bp ~ bmi * sex)
```

```
##
```

```
## Residuals:
```

##	Min	1Q	Median	3Q	Max
##	-6.9778	-3.9998	0.9333	2.1966	9.9381

```
##
```

```
## Coefficients:
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	56.336	35.870	1.571	0.167
## bmi	1.815	1.397	1.300	0.241
## sexM	-26.840	53.472	-0.502	0.634
## bmi:sexM	1.482	2.152	0.689	0.517

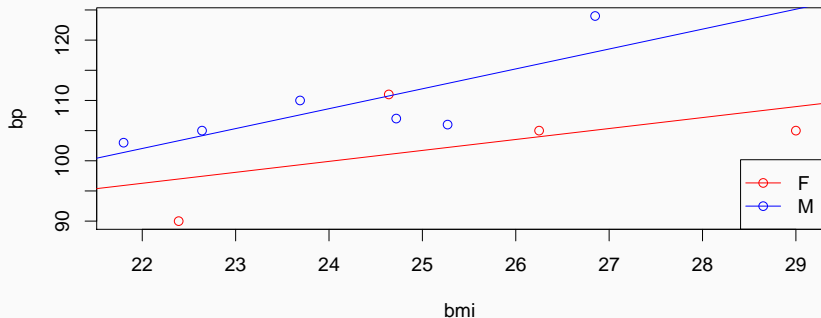
```
##
```

```
## Residual standard error: 6.728 on 6 degrees of freedom
```

```
## Multiple R-squared: 0.5692, Adjusted R-squared: 0.3538
```

```
## F-statistic: 2.643 on 3 and 6 DF, p-value: 0.1436
```

```
> cf <- coef(model_int)
> plot(bp~bmi, col = c(F='red', M='blue')[sex])
> abline(cf[1], cf[2], col='red')
> abline(cf[1] + cf[3], cf[2] + cf[4], col='blue')
> legend("bottomright", c('F', 'M'), pch = 1, lty=1, col=c('red', 'blue'))
```



```
> anova(model_bmi, model_add, model_int)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: bp ~ bmi
```

```
## Model 2: bp ~ bmi + sex
```

```
## Model 3: bp ~ bmi * sex
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      8 501.27
```

```
## 2      7 293.04  1    208.23 4.6006 0.07564 .
```

```
## 3      6 271.57  1     21.47 0.4743 0.51673
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 5: Principais Modelos de Regressão Linear

Modelo	Equação	Formula
Nula	$y = \alpha + \epsilon$	$y \sim 1$
Linear	$y = \alpha + \beta x + \epsilon$	$y \sim 1 + x$ ou $y \sim x$
Discreto	$y = \alpha + \beta x + \epsilon; x \in 0, 1$	$y \sim 1 + x$ ou $y \sim x$
Polinomial	$y = \alpha + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \epsilon$	$y \sim \text{poly}(x, k)$
Múltipla	$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$	$y \sim x_1 + x_2 + x_3$
Iteração	$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \epsilon$	$y \sim x_1 + x_2 + x_1 : x_2$ ou $y \sim x_1 * x_2$



## Exercícios - Regressão Linear

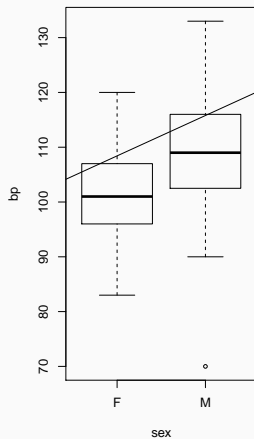
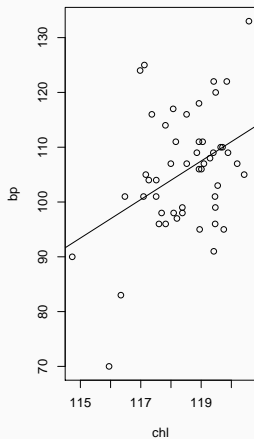
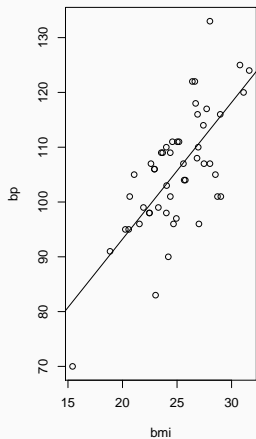
Agora utilizando a amostra de 50 pacientes investigados, desenvolva um modelo estatístico relacionando a pressão sanguínea (**bp**) com o índice de massa corpórea (**bmi**), taxa de colesterol (**chl**) e o sexo (**sexo**).

1. Ilustre a relação das variáveis investigadas e a pressão sanguínea.
2. Considerando um modelo de efeito apenas aditivo (sem interações), descreva o efeito das variáveis sobre a pressão sanguínea, avalie se a relação é significativa ou espúria e ilustre e avalie os principais gráficos de diagnóstico.
3. Considerando um modelo onde o efeito do BMI depende do sexo, descreva o efeito das variáveis sobre a pressão sanguínea, avalie se a relação é significativa ou espúria e ilustre e avalie os principais gráficos de diagnóstico.
4. Compare o modelo aditivo ao com interações. Qual prevê melhor a pressão sanguínea nessa amostra? Justifique

```
> db <- read.table('db.tsv', header=T)
```

1. Ilustre a relação das variáveis investigadas e a pressão sanguínea.

```
> par(mfrow = c(1, 3))  
> plot(bp~bmi, data=db); abline(lm(bp~bmi, data=db))  
> plot(bp~chl, data=db); abline(lm(bp~chl, data=db))  
> plot(bp~sex, data=db); abline(lm(bp~sex, data=db))
```



2. Considerando um modelo de efeito apenas aditivo (sem interações), descreva o efeito das variáveis sobre a pressão sanguínea, avalie se a relação é significativa ou espúria e ilustre e avalie os principais gráficos de diagnóstico.

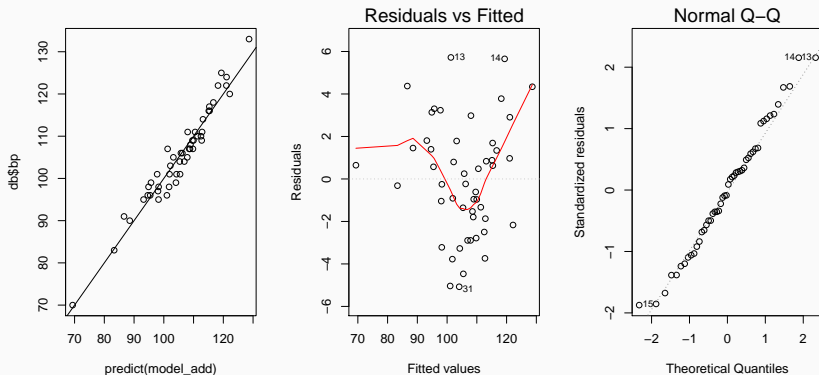
```

> summary(model_add <- lm(bp~bmi+chl+sex, data=db))

##
## Call:
## lm(formula = bp ~ bmi + chl + sex, data = db)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.0745 -1.8502  0.0071  1.6322  5.7203
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -562.2738    39.6765  -14.17  < 2e-16 ***
## bmi          2.8780     0.1262   22.81  < 2e-16 ***
## chl          4.9794     0.3293   15.12  < 2e-16 ***
## sexM          9.8891     0.8195   12.07  7.46e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.789 on 46 degrees of freedom
## Multiple R-squared:  0.9416, Adjusted R-squared:  0.9377
## F-statistic: 247 on 3 and 46 DF, p-value: < 2.2e-16

```

```
> par(mfrow = c(1,3))
> plot(db$bp~predict(model_add))
> abline(0, 1)
> plot(model_add, which = 1:2)
```



3. Considerando um modelo onde o efeito do eNOS interage com BMI, descreva o efeito das variáveis sobre a pressão sanguínea, avalie se a relação é significativa ou espúria e ilustre e avalie os principais gráficos de diagnóstico.

```
> summary(model_int <- lm(bp~bmi*sex+chl, data=db))
```

```
##
```

```
## Call:
```

```
## lm(formula = bp ~ bmi * sex + chl, data = db)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -4.4414 -1.2784  0.0318  1.0547  4.1842
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -524.3073    31.8531 -16.460  < 2e-16 ***
## bmi          2.1935     0.1595  13.756  < 2e-16 ***
## sexM        -17.8531     5.1116  -3.493  0.00108 **
## chl          4.8050     0.2600  18.483  < 2e-16 ***
## bmi:sexM      1.1045     0.2019   5.471  1.89e-06 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 2.186 on 45 degrees of freedom
```

```
## Multiple R-squared:  0.9649, Adjusted R-squared:  0.9618
```

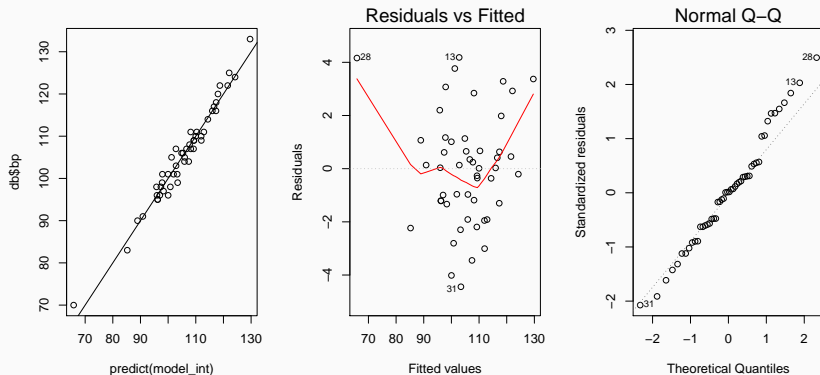
```
## F-statistic: 309.2 on 4 and 45 DF,  p-value: < 2.2e-16
```



```

> par(mfrow = c(1,3))
> plot(db$bp~predict(model_int))
> abline(0, 1)
> plot(model_int, which = 1:2)

```



4. Compare o modelo aditivo ao com interações. Qual prevê melhor a pressão sanguínea nessa amostra? Justifique

```
> anova(model_add, model_int)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: bp ~ bmi + chl + sex
```

```
## Model 2: bp ~ bmi * sex + chl
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1      46 357.93
```

```
## 2      45 214.96  1    142.96 29.928 1.893e-06 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ao final

---

- Modelar relação entre variáveis através de uma regressão linear.
- Compreender os principais modelos de regressão linear.
- Compreender os coeficientes e seu significado.
- Ilustrar a regressão e resíduos.
- Comparar diferentes modelos de regressão.

Até a próxima

---