

Bioestatística em R - Exercícios

Nome: _____

Data: _____

Matricula: _____

Correlação e Regressão Linear

Em um estudo de hipertensão, os pesquisadores constataram uma elevada variabilidade da pressão sanguínea em repouso (bp) dos pacientes. Para investigar a causa desta variabilidade, os pesquisadores coletaram dados referentes ao sexo dos pacientes (sex), índice de massa corporea (bmi), taxa de metabolismo basal (bmr), taxa de colesterol no sangue (chl) e genotiparam uma mutação do gene eNOS (enos).

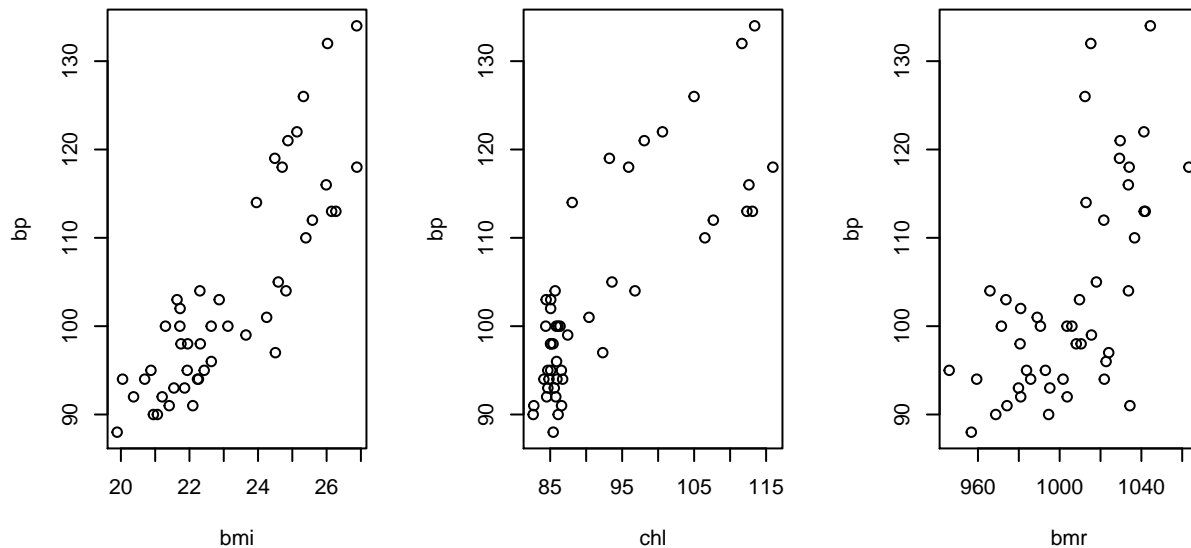
```
dt <- read.table('bp.tsv', header=T)
summary(dt)
```

```
##      bmi      bmr      chl      bp      enos
## Min.   :19.89  Min.   : 945.9  Min.   : 82.64  Min.   : 88.0  aa: 3
## 1st Qu.:21.68  1st Qu.: 982.4  1st Qu.: 85.08  1st Qu.: 94.0  Aa:16
## Median :22.43  Median :1008.1  Median : 86.11  Median :100.0  AA:28
## Mean   :23.11  Mean   :1005.1  Mean   : 91.93  Mean   :103.3
## 3rd Qu.:24.76  3rd Qu.:1026.7  3rd Qu.: 96.34  3rd Qu.:112.5
## Max.   :26.88  Max.   :1063.3  Max.   :115.94  Max.   :134.0
## sex
## F:29
## M:18
##
##
##
##
```

Sobre os dados obtidos responda:

1. Para as medidas numéricas, avalie a correlação destas com a pressão sanguínea. Ilustre a relação entre as variáveis e indique qual a medida de correlação mais indicada para cada situação.

```
## Plotar um scatter plot para as medidas
par(mfrow = c(1,3))
plot(bp~bmi, data=dt)
plot(bp~chl, data=dt)
plot(bp~bmr, data=dt)
```



```
## ou plot(bp~bmi+chl+bmr, data=dt)
```

```
## Testar a correlação
cor.test(dt$bp, dt$bmi)
```

```
##
## Pearson's product-moment correlation
##
## data: dt$bp and dt$bmi
## t = 10.963, df = 45, p-value = 2.698e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.7493864 0.9158254
## sample estimates:
## cor
## 0.8529906
```

```
cor.test(dt$bp, dt$chl, method="spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: dt$bp and dt$chl
## S = 4891.1, p-value = 1.414e-08
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.717213
```

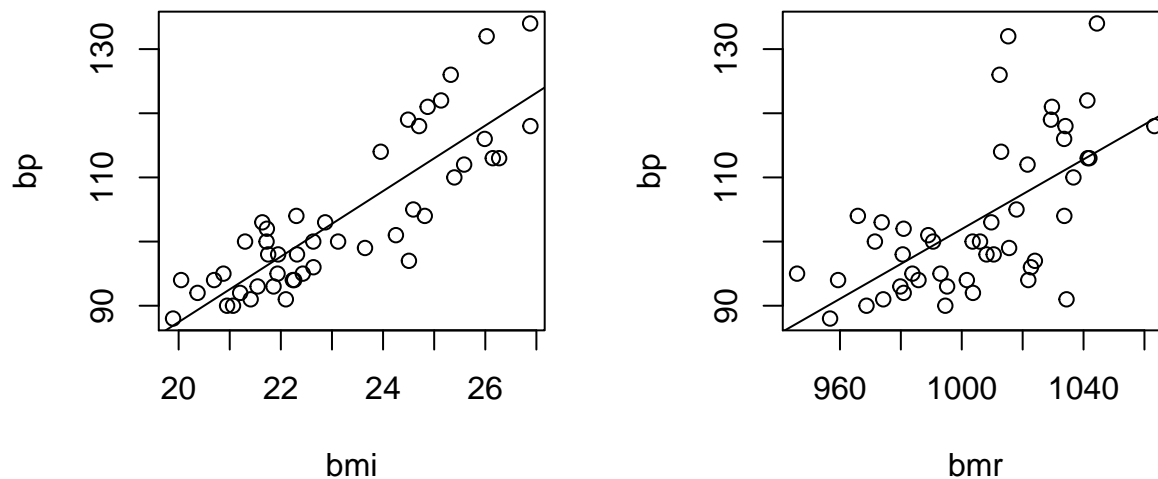
```
cor.test(dt$bp, dt$bmr)
```

```
##
```

```
## Pearson's product-moment correlation
##
## data: dt$bp and dt$bmr
## t = 5.4269, df = 45, p-value = 2.194e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.4171209 0.7759667
## sample estimates:
## cor
## 0.62895
```

2. Para as medidas que apresentam uma relação linear, modele a influência dela sobre a pressão sanguínea. Ilustre a relação, indicando a intensidade do efeito e se ele é aleatória ou significativa. Por último compare os modelos indicando qual melhor modela a pressão sanguínea segundo o coeficiente de determinação?

```
## modelar as relações
model_bmi <- lm(bp~bmi, data=dt)
model_bmr <- lm(bp~bmr, data=dt)
## ilustrar a relação entre as medidas
par(mfrow = c(1,2))
plot(bp~bmi, data=dt)
abline(model_bmi)
plot(bp~bmr, data=dt)
abline(model_bmr)
```



```
## avaliar a associação indicando se a associação é espúria ou
## significativa.
summary(model_bmi)
```

```
##
## Call:
## lm(formula = bp ~ bmi, data = dt)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.455  -4.858  -1.596   5.262  13.781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -14.6559    10.8009  -1.357   0.182
## bmi          5.1053     0.4657  10.963 2.7e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.209 on 45 degrees of freedom
## Multiple R-squared:  0.7276, Adjusted R-squared:  0.7215
## F-statistic: 120.2 on 1 and 45 DF,  p-value: 2.698e-14
```

```
summary(model_bmr)
```

```
##
## Call:
## lm(formula = bp ~ bmr, data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.292  -5.600  -1.829   5.524  25.908
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -169.74419    50.33858  -3.372  0.00154 **
## bmr          0.27169     0.05006   5.427 2.19e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.249 on 45 degrees of freedom
## Multiple R-squared:  0.3956, Adjusted R-squared:  0.3821
## F-statistic: 29.45 on 1 and 45 DF,  p-value: 2.194e-06
```

```
## comparando o r^2 ajustado
```

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

```
## [1] 0.7215394
```

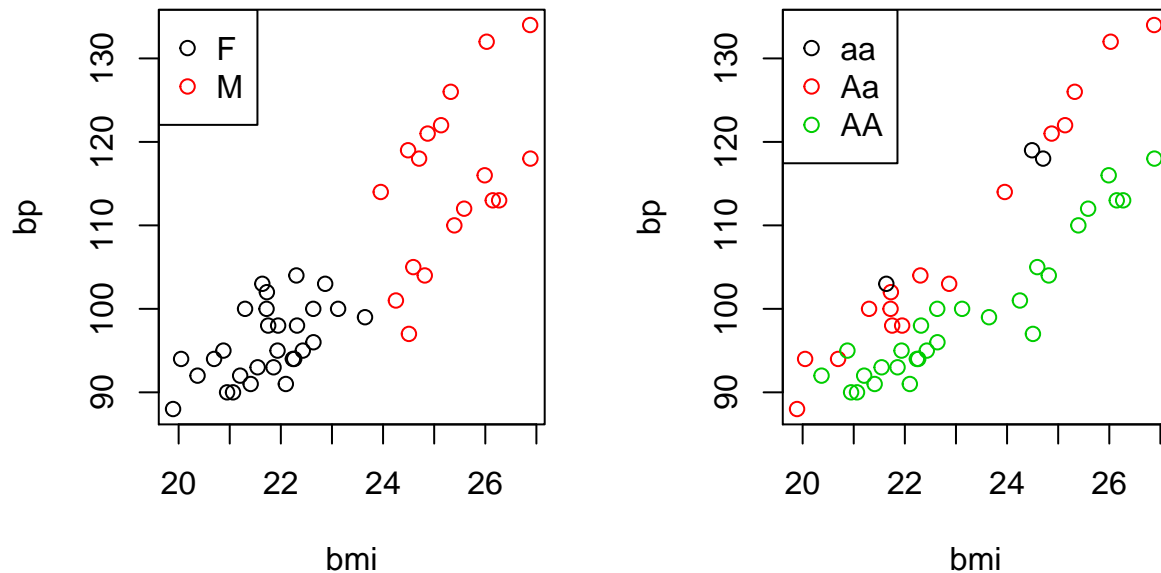
```
summary(model_bmr)$adj.r.squared
```

```
## [1] 0.3821465
```

3. Para o melhor modelo identificado, ilustre a relação entre as medidas destacando os

casos com cores diferentes de acordo com o sexo e genótipo do eNOS. Essas variáveis apareçam estar relacionadas com a pressão sanguínea?

```
## ilustrar a relação, usando cores diferentes para cada categoria
par(mfrow=c(1,2))
plot(bp~bmi, data=dt, col=sex)
legend("topleft", levels(dt$sex), col=1:2, pch = 1)
plot(bp~bmi, data=dt, col=enos)
legend("topleft", levels(dt$enos), col=1:3, pch = 1)
```



4. Atualize o modelo obtido na questão (2) para incluir a variável identificada na questão (3) seguindo um modelo aditivo e interativo. Ilustre essa relação utilizando curvas distintas para cada categoria, a seguir compare os modelos obtidos indicando qual o melhor resultado.

```
## atualizar o modelo para incluir o efeito do eNOS em recessivo
## (aa/Aa vs AA).
dt$enos_rec <- factor(dt$enos == "AA", labels=c("aa/Aa", "AA"))

model_add <- lm(bp~bmi+enos_rec, data=dt)
summary(model_add)
```

```
##
## Call:
## lm(formula = bp ~ bmi + enos_rec, data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.453  -2.103  -0.009   1.990   7.471
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.0311      6.1197  -1.803   0.0783 .
## bmi          5.2132      0.2636  19.778 < 2e-16 ***
## enos_recAA  -10.2720      1.0447  -9.833 1.12e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.512 on 44 degrees of freedom
## Multiple R-squared:  0.9148, Adjusted R-squared:  0.9109
## F-statistic: 236.2 on 2 and 44 DF,  p-value: < 2.2e-16
```

```
model_int <- lm(bp~bmi*enos_rec, data=dt)
summary(model_int)
```

```
##
## Call:
## lm(formula = bp ~ bmi * enos_rec, data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.1192 -1.3183  0.0158  1.7417  5.1562
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -40.1510      7.0765  -5.674 1.09e-06 ***
## bmi            6.4785      0.3063  21.151 < 2e-16 ***
## enos_recAA     42.1410      9.5064   4.433 6.32e-05 ***
## bmi:enos_recAA -2.2702      0.4103  -5.533 1.74e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.715 on 43 degrees of freedom
## Multiple R-squared:  0.9502, Adjusted R-squared:  0.9468
## F-statistic: 273.7 on 3 and 43 DF,  p-value: < 2.2e-16
```

```
## ilustre as relações
par(mfrow=c(1,2))

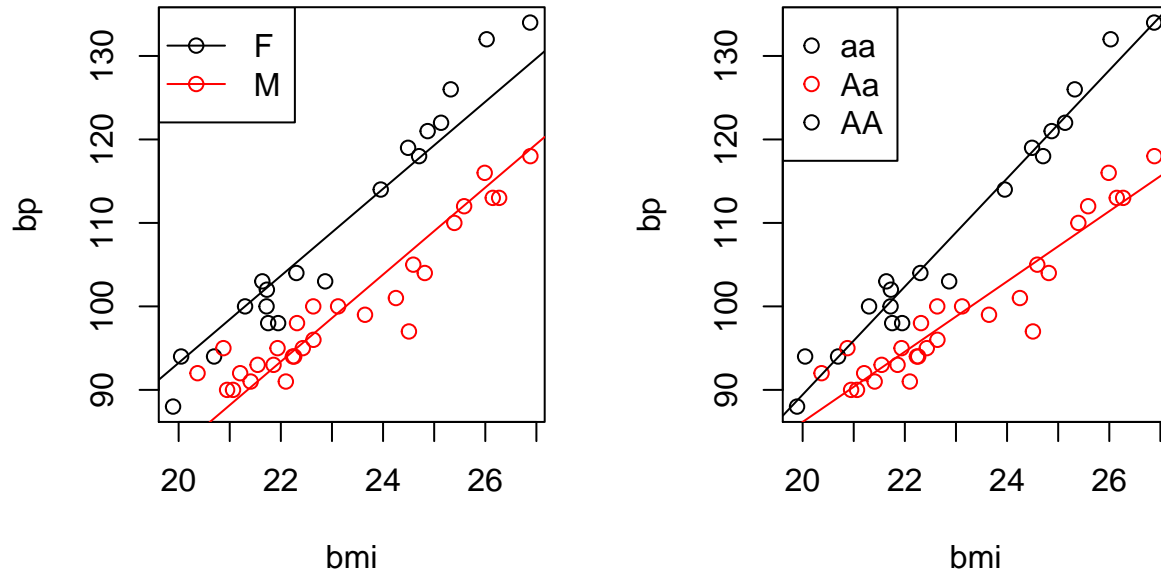
plot(bp~bmi, data=dt, col=enos_rec)
legend("topleft", levels(dt$sex), col=1:2, pch = 1, lty = 1)

cf <- coef(model_add)
abline(cf[1], cf[2], col=1)
abline(cf[1] + cf[3], cf[2], col=2)

plot(bp~bmi, data=dt, col=enos_rec)
```

```
legend("topleft", levels(dt$enos), col=1:2, pch = 1)
```

```
cf <- coef(model_int)
abline(cf[1], cf[2], col=1)
abline(cf[1] + cf[3], cf[2] + cf[4], col=2)
```



```
## comparar os modelos
anova(model_bmi, model_add, model_int)
```

```
## Analysis of Variance Table
##
## Model 1: bp ~ bmi
## Model 2: bp ~ bmi + enos_rec
## Model 3: bp ~ bmi * enos_rec
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      45 1734.84
## 2      44  542.57  1   1192.27 161.773 3.619e-16 ***
## 3      43  316.91  1    225.66 30.619 1.737e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```