# relatorio

April 24, 2018

```
In [ ]: # Jailson Nunes Leocadio
        # Exercícios de Regressão Múltipla
```

1   Uma estimativa da incerteza na previsão do modelo

```
In [10]: data = read.table("babies.txt", header=TRUE)

         # Preparação e remoção de NA's
         data[ data$gestation == 999, 2 ] = NA
         data = data[complete.cases(data), ]

         # Conversão do peso para Kg
         bwt = data$bwt * 0.0283495

         model = lm(bwt ~ data$gestation)

         # Cálculo do erro
         ssx = sum((data$gestation - mean(data$gestation))^2)
         vt = seq(min(data$gestation), max(data$gestation), length.out=100)
         se = sqrt((var(bwt)) * ((1/length(bwt)) + (((vt - mean(data$gestation)) ^ 2)/ssx)))

         t = qt(0.975, (length(bwt) - 2))

         vt.model = coef(model)[1] + vt * coef(model)[2]

         vt.model.p = vt.model + (se * t)
         vt.model.n = vt.model - (se * t)

         plot(bwt ~ data$gestation, xlab="Duração da gestação (dias)",
              ylab="Massa ao nascer (kg)", pch=20)
         abline(model, col="blue", lwd=2)

         lines(vt, vt.model.p, col="red", lwd=2)
         lines(vt, vt.model.n, col="green", lwd=2)
```

1
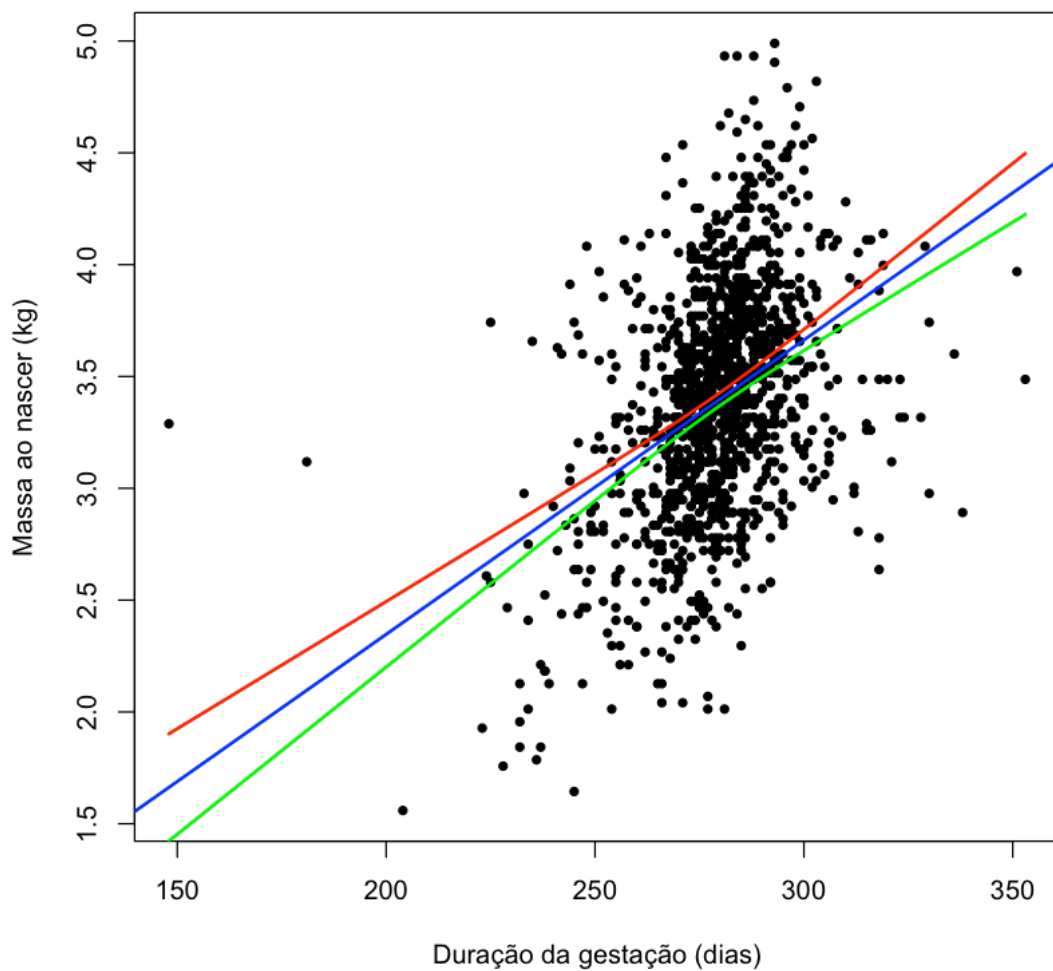
2   Galileu estava Certo?

```
In [19]: init.h = c(600, 700, 800, 950, 1100, 1300, 1500)
         h.d = c(253, 337, 395, 451, 495, 534, 573)

         plot(h.d ~ init.h, pch=19)

         mod1 <- lm(h.d ~ init.h)
         mod2 <- update(mod1, .~. + I(init.h ^ 2))

         # anova(mod1, mod2)
```

```
abline(mod1, lwd=2, col = "green")

cf.m2 <- coef(mod2)
curve(cf.m2[1] +
        cf.m2[2] * x +
        cf.m2[3] * (x ^ 2),
        add = T, lty = 2, col = "blue", lwd=2)

# summary(mod2)

mod3 = lm(h.d ~ init.h + I(init.h ^ 2) + I(init.h ^ 3))
# anova(mod2, mod3)

cf.m3 <- coef(mod3)
curve(cf.m3[1] +
        cf.m3[2] * x +
        cf.m3[3] * (x ^ 2) +
        cf.m3[4] * (x ^ 3),
        col = "red", add = T, lty = 2, lwd=2)
```
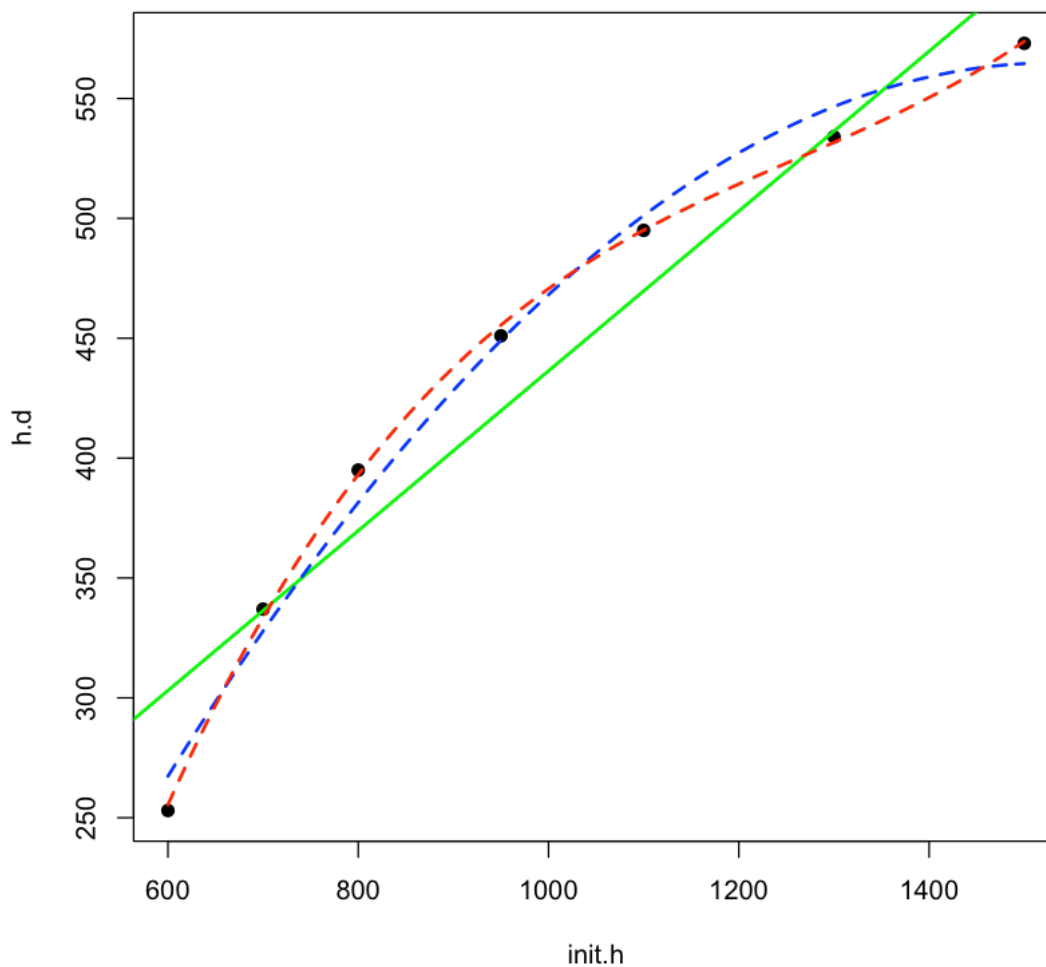
3 Massa de Recém-Nascidos

```
In [25]: data = read.table("babies.txt", header = TRUE)

         data[data$gestation == 999, 2] = NA
         data[data$age == 99, 4] = NA
         data[data$height == 99, 5] = NA
         data[data$weight == 999, 6] = NA
         data[data$smoke == 9, 7] = NA

         data = data[complete.cases(data),]
         data$parity = as.factor(data$parity)
```

```r
data$smoke = as.factor(data$smoke)

writeLines("_____

model1 = lm(bwt ~ gestation + parity + age + height + weight + smoke +
        gestation:age + gestation:smoke +
        I(gestation^2) + I(age^2) + I(height^2) + I(weight^2), data = data)

summary(model1)

writeLines("_____

model2 = lm(bwt ~ gestation + parity + age + height + weight + smoke +
        gestation:age + gestation:smoke +
        I(gestation^2) + I(height^2) + I(weight^2), data = data)

anova(model1, model2)
summary(model2)

writeLines("_____

model3 = lm(bwt ~ gestation + parity + age + height + weight + smoke +
        gestation:age + gestation:smoke +
        I(gestation^2) + I(height^2), data = data)

anova(model2, model3)
summary(model3)

writeLines("_____

model4 = lm(bwt ~ parity + age + height + weight + smoke +
        gestation:age + gestation:smoke +
        I(gestation^2) + I(height^2), data = data)

anova(model3, model4)
summary(model4)

writeLines("_____

model5 = lm(bwt ~ parity + age + height + weight + smoke +
        gestation:age + gestation:smoke +
        I(height^2), data = data)

anova(model4, model5)
summary(model5)

writeLines("_____
```

```
model6 = lm(bwt ~ parity + age + height + smoke +
            gestation:age + gestation:smoke +
            I(height^2), data = data)

anova(model5, model6)
summary(model6)

writeLines("_____

# Modelo mínimo final selecionado
model7 = lm(bwt ~ parity + age + smoke +
            gestation:age + gestation:smoke +
            I(height^2), data = data)

anova(model6, model7)
summary(model7)

anova(model7)
```

```
Call:
lm(formula = bwt ~ gestation + parity + age + height + weight +
    smoke + gestation:age + gestation:smoke + I(gestation^2) +
    I(age^2) + I(height^2) + I(weight^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-60.102  -9.990  -0.307   9.055  59.441

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.994e+02  2.113e+02   1.890 0.059006 .
gestation        6.690e-01  3.864e-01   1.731 0.083663 .
parity1         -3.426e+00  1.141e+00  -3.002 0.002739 **
age             -4.655e+00  1.608e+00  -2.896 0.003851 **
height          -1.295e+01  6.248e+00  -2.073 0.038432 *
weight           3.311e-01  1.805e-01   1.834 0.066890 .
smoke1          -7.374e+01  1.690e+01  -4.365 1.39e-05 ***
I(gestation^2)  -1.379e-03  6.387e-04  -2.159 0.031072 *
I(age^2)         1.153e-03  1.195e-02   0.096 0.923151
I(height^2)      1.097e-01  4.883e-02   2.247 0.024807 *
I(weight^2)     -9.832e-04  6.210e-04  -1.583 0.113675
gestation:age    1.632e-02  5.114e-03   3.190 0.001459 **
gestation:smoke1 2.336e-01  6.051e-02   3.861 0.000119 ***
---
```

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.63 on 1161 degrees of freedom
Multiple R-squared: 0.28,Adjusted R-squared: 0.2725
F-statistic: 37.62 on 12 and 1161 DF, p-value: < 2.2e-16

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1161	283739.2	NA	NA	NA	NA
1162	283741.5	-1	-2.275194	0.009309607	0.923151

Call:
lm(formula = bwt ~ gestation + parity + age + height + weight +
    smoke + gestation:age + gestation:smoke + I(gestation^2) +
    I(height^2) + I(weight^2), data = data)

Residuals:
    Min   1Q Median   3Q   Max
-59.999 -9.968 -0.312   9.052 59.422

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.975e+02 2.103e+02   1.890 0.058976 .
gestation   6.698e-01 3.861e-01   1.735 0.083077 .
parity1   -3.403e+00 1.117e+00 -3.046 0.002371 **
age   -4.584e+00 1.426e+00 -3.214 0.001343 **
height   -1.293e+01 6.240e+00 -2.071 0.038555 *
weight   3.319e-01 1.802e-01   1.841 0.065812 .
smoke1   -7.367e+01 1.687e+01 -4.366 1.38e-05 ***
I(gestation^2) -1.379e-03 6.384e-04 -2.161 0.030915 *
I(height^2)   1.096e-01 4.877e-02   2.246 0.024878 *
I(weight^2)   -9.855e-04 6.203e-04 -1.589 0.112415
gestation:age   1.630e-02 5.109e-03   3.190 0.001459 **
gestation:smoke1 2.334e-01 6.043e-02   3.862 0.000119 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.63 on 1162 degrees of freedom
Multiple R-squared: 0.2799,Adjusted R-squared: 0.2731
F-statistic: 41.07 on 11 and 1162 DF, p-value: < 2.2e-16

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1162	283741.5	NA	NA	NA	NA
1163	284357.8	-1	-616.2614	2.523761	0.1124149

```
Call:
lm(formula = bwt ~ gestation + parity + age + height + weight +
    smoke + gestation:age + gestation:smoke + I(gestation^2) +
    I(height^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-59.920  -9.853  -0.316   9.170  59.301

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.870e+02  2.103e+02   1.840 0.066012 .
gestation        6.533e-01  3.862e-01   1.691 0.091049 .
parity1         -3.343e+00  1.117e+00  -2.991 0.002835 **
age             -4.571e+00  1.427e+00  -3.203 0.001396 **
height          -1.202e+01  6.218e+00  -1.933 0.053519 .
weight           4.833e-02  2.501e-02   1.933 0.053518 .
smoke1          -7.379e+01  1.688e+01  -4.371 1.35e-05 ***
I(gestation^2)  -1.349e-03  6.385e-04  -2.113 0.034815 *
I(height^2)      1.032e-01  4.864e-02   2.122 0.034022 *
gestation:age    1.629e-02  5.113e-03   3.187 0.001476 **
gestation:smoke1 2.337e-01  6.047e-02   3.865 0.000117 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.64 on 1163 degrees of freedom
Multiple R-squared:  0.2784,Adjusted R-squared:  0.2722
F-statistic: 44.87 on 10 and 1163 DF,  p-value: < 2.2e-16
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1163	284357.8	NA	NA	NA	NA
1164	285057.2	-1	-699.396	2.860473	0.09104874

```
Call:
lm(formula = bwt ~ parity + age + height + weight + smoke + gestation:age +
    gestation:smoke + I(gestation^2) + I(height^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-59.621  -9.999  -0.227   9.268  48.050

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.998e+02  1.996e+02   2.503 0.012436 *
parity1         -3.256e+00  1.117e+00  -2.914 0.003633 **
age             -5.649e+00  1.278e+00  -4.421 1.07e-05 ***
height          -1.229e+01  6.221e+00  -1.976 0.048356 *
weight           4.844e-02  2.503e-02   1.936 0.053170 .
smoke1          -7.262e+01  1.688e+01  -4.302 1.84e-05 ***
I(gestation^2)  -3.474e-04  2.386e-04  -1.456 0.145746
I(height^2)      1.054e-01  4.866e-02   2.165 0.030556 *
age:gestation    2.018e-02  4.572e-03   4.414 1.11e-05 ***
smoke1:gestation 2.297e-01  6.047e-02   3.798 0.000154 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.65 on 1164 degrees of freedom
Multiple R-squared:  0.2766,Adjusted R-squared:  0.271
F-statistic: 49.45 on 9 and 1164 DF,  p-value: < 2.2e-16
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1164	285057.2	NA	NA	NA	NA
1165	285576.1	-1	-518.9456	2.119058	0.1457456

```
Call:
lm(formula = bwt ~ parity + age + height + weight + smoke + gestation:age +
    gestation:smoke + I(height^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-59.555  -9.969  -0.229   9.286  47.897

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     475.244278 199.011363   2.388 0.017098 *
parity1          -3.319584   1.116772  -2.972 0.003015 **
age              -3.863853   0.358395 -10.781  < 2e-16 ***
height          -12.361828   6.223788  -1.986 0.047243 *
weight            0.047062   0.025020   1.881 0.060226 .
smoke1          -66.881984  16.423584  -4.072 4.97e-05 ***
I(height^2)       0.105817   0.048686   2.173 0.029947 *
age:gestation     0.013780   0.001261  10.931  < 2e-16 ***
```

```
smoke1:gestation   0.209324   0.058865   3.556 0.000392 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.66 on 1165 degrees of freedom
Multiple R-squared:  0.2753,Adjusted R-squared:  0.2703
F-statistic: 55.32 on 8 and 1165 DF,  p-value: < 2.2e-16
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1165	285576.1	NA	NA	NA	NA
1166	286443.4	-1	-867.2742	3.538021	0.0602262

```
Call:
lm(formula = bwt ~ parity + age + height + smoke + gestation:age +
    gestation:smoke + I(height^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-60.503 -10.134  -0.047   9.594  50.895

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     483.913925 199.174399   2.430 0.015266 *
parity1          -3.486627   1.114446  -3.129 0.001800 **
age              -3.820584   0.358045 -10.671  < 2e-16 ***
height          -12.628447   6.228942  -2.027 0.042850 *
smoke1          -68.620482  16.415393  -4.180 3.13e-05 ***
I(height^2)       0.109228   0.048705   2.243 0.025108 *
age:gestation     0.013695   0.001261  10.859  < 2e-16 ***
smoke1:gestation  0.215119   0.058848   3.656 0.000268 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.67 on 1166 degrees of freedom
Multiple R-squared:  0.2731,Adjusted R-squared:  0.2687
F-statistic: 62.58 on 7 and 1166 DF,  p-value: < 2.2e-16
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1166	286443.4	NA	NA	NA	NA
1167	287453.1	-1	-1009.745	4.110279	0.04285043

```
Call:
lm(formula = bwt ~ parity + age + smoke + gestation:age + gestation:smoke +
    I(height^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-58.717 -10.092   0.020   9.509  56.271

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      80.314176   6.321909  12.704  < 2e-16 ***
parity1          -3.421734   1.115470  -3.068 0.002208 **
age              -3.788709   0.358176 -10.578  < 2e-16 ***
smoke1          -69.489712  16.431646  -4.229 2.53e-05 ***
I(height^2)       0.010526   0.001424   7.393 2.73e-13 ***
age:gestation     0.013613   0.001262  10.785  < 2e-16 ***
smoke1:gestation  0.218495   0.058903   3.709 0.000218 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.69 on 1167 degrees of freedom
Multiple R-squared:  0.2705,Adjusted R-squared:  0.2668
F-statistic: 72.13 on 6 and 1167 DF,  p-value: < 2.2e-16
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
parity	1	759.71505	759.71505	3.0842853	7.931427e-02
age	1	60.15788	60.15788	0.2442285	6.212616e-01
smoke	1	24042.55690	24042.55690	97.6077887	3.671286e-22
I(height^2)	1	17674.37136	17674.37136	71.7542777	7.244930e-17
age:gestation	1	60678.64473	60678.64473	246.3426980	1.630659e-50
smoke:gestation	1	3389.27989	3389.27989	13.7597726	2.175101e-04
Residuals	1167	287453.12512	246.31802	NA	NA