

Estudo de Caso: Influência da Menarca no Ganho de Peso Corporal

José Luiz Padilha da Silva
10 de outubro de 2018

Descrição do estudo

Os dados são provenientes de um estudo prospectivo (*MIT Grow and Development Study*) sobre o aumento de gordura corporal em uma coorte de 162 garotas. Estes dados representam um subconjunto dos dados originais do estudo e não devem ser usados para conclusões definitivas.

No início do estudo todas as garotas estavam na pré-menarca e não eram obesas.

- Sabe-se que o percentual de gordura nas garotas tem um aumento considerável no período em torno da menarca (primeira menstruação).
- Parece que este aumento continua significativo por aproximadamente quatro anos depois da menarca, mas este comportamento ainda não havia sido devidamente estudado.
- As meninas foram acompanhadas anualmente até quatro anos depois da menarca. Em cada consulta, foi derivada uma medida da gordura corporal (percentual) por meio de medidas de (1) peso corporal, (2) altura, e (3) impedância bioelétrica.

O objetivo é estudar o comportamento do peso das garotas a partir da menarca. As variáveis são:

- time: tempo em relação à menarca (idade da menina no instante observado menos a idade quando houve a menarca) - pode ser positivo ou negativo.
- pbf: porcentagem de gordura corporal.

Consideraremos a hipótese de que o percentual de gordura aumenta linearmente com a idade, mas com diferentes inclinações antes e após menarca.

Análise Descritiva

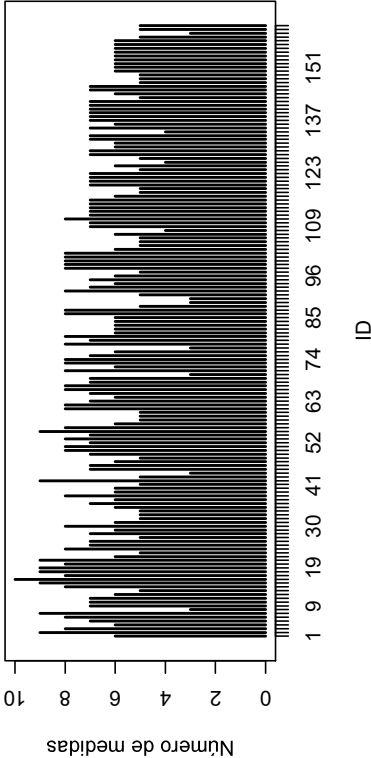
```
library(ggplot2) # gráficos
require(nlme) # usado para ajustar o modelo misto
require(mgcv) # necessário para a função extract.lme.cov
library(car) # envelope no qqplot
options(width=2000)
#
dados<-read.table("fat.txt",header=T)
attach(dados)
head(dados)

## ID age age_menarche time pbf
## 1 1 9.32 13.19 -3.87 7.94
## 2 1 10.33 13.19 -2.86 15.65
## 3 1 11.24 13.19 -1.95 13.51
## 4 1 12.19 13.19 -1.00 23.23
## 5 1 13.24 13.19 0.05 10.52
## 6 1 14.24 13.19 1.05 20.45
```

```
summary(dados, digits=3)

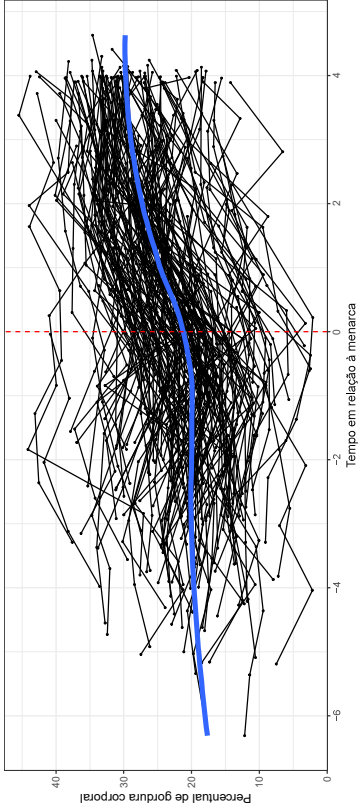
## ID age age_menarche time pbf
## Min. : 1.0 Min. : 8.31 Min. :-6.310 Min. : 2.11
## 1st Qu.: 38.0 1st Qu.:11.26 1st Qu.:12.1 1st Qu.:18.24
## Median : 78.0 Median :12.95 Median :13.0 Median :23.59
## Mean : 78.8 Mean :13.06 Mean :12.9 Mean :23.61
## 3rd Qu.:119.0 3rd Qu.:14.88 3rd Qu.:13.5 3rd Qu.:29.56
## Max. :162.0 Max. :18.36 Max. :4.630 Max. :45.49

plot(table(ID), ylab="Número de medidas")
```



Vemos que o número de medidas repetidas por menina é variável:

```
range(table(ID)) #mínimo e máximo de avaliações por indivíduo
## [1] 3 10
length(ID)/length(unique(ID)) #número médio de avaliações
## [1] 6.475309
A seguir, gráfico de perfis.
p1=ggplot(dados,aes(x=time,y=pbf)) + geom_line(aes(group=ID)) + theme(legend.position="top") +
scale_x_continuous(breaks=seq(-6,4,2)) + geom_point(size=0.5) + geom_smooth(method="loess",
se=FALSE, size=2) + geom_vline(xintercept=0, col = "red", linetype="dashed") + theme_bw() +
labs(y="Percentual de gordura corporal", x="Tempo em relação à menarca")
p1
```



Ajuste dos Modelos

Modelo 1

O modelo inicialmente proposto considera que cada garota tem uma curva de crescimento *spline* linear com um nó no tempo da menarca. O modelo considera duas inclinações: antes e pós menarca.

Vamos ajustar o seguinte modelo linear de efeitos mistos:

$$E(Y_{ij}|b_i) = \beta_1 + \beta_2 t_{ij} + \beta_3 (t_{ij})_+ + b_{1i} + b_{2i} t_{ij} + b_{3i} (t_{ij})_+,$$

em que

$$(t_{ij})_+ = \begin{cases} t_{ij} & \text{se } t_{ij} > 0 \\ 0 & \text{se } t_{ij} \leq 0. \end{cases}$$

Lembramos que no modelo linear de efeitos mistos, a matriz de variância-covariância de Y_i é dada por:

$$Var(Y_i) = \Sigma_i = Z_i G Z_i' + \sigma^2 I_{n_i},$$

em que Z_i é a matriz de covariáveis relacionadas aos efeitos aleatórios, G é a matriz de covariância dos efeitos aleatórios e n_i é o número de observações do i -ésimo indivíduo, $i = 1, \dots, N$.

Logo, os resíduos transformados podem ser obtidos a partir da decomposição de Cholesky da matriz estimada $Var(Y_i) = Z_i \hat{G} Z_i' + \hat{\sigma}^2 I_{n_i}$.

```
dados$time0 = ifelse(dados$time>0,dados$time,0)
out<-lme(pbf~time+time0, random= ~time+time0|ID,data=dados)
summary(out)
```

```
## Linear mixed-effects model fit by REML
## Data: dados
## AIC      BIC      logLik
## 6082.401 6131.929 -3031.201
```

```
## Random effects:
## Formula: ~time + time0 | ID
## Structure: General positive-definite, Log-Cholesky parametrization
## StdDev   Corr
## (Intercept) 6.777960 (Intr) time
## time      1.277125  0.292
## time0     1.658205 -0.544 -0.827
## Residual   3.077866
## Fixed effects: pbf ~ time + time0
## Value Std.Error DF   t-value p-value
## (Intercept) 21.361383 0.5645526 885 37.83772 0.0000
## time      0.417114 0.1571563 885  2.65413 0.0081
## time0     2.047139 0.2279677 885  8.97995 0.0000
## Correlation:
## (Intr) time
## time      0.351
## time0    -0.515 -0.872
## Standardized Within-Group Residuals:
## Min      Q1      Med      Q3     Yax
## -2.7742100 -0.5900330 -0.0359290  0.5946416  3.3797776
## Number of Observations: 1049
## Number of Groups: 162
```

O principal objetivo do estudo era avaliar se as inclinações populacionais de pbf diferem antes e após a menarca. Baseado na estatística de teste para β_3 podemos concluir que há diferença estatisticamente significativa entre os dois períodos.

- Em particular, a inclinação pré-menarca é relativamente pequena (0,42) e indica que a taxa anual de pbf é menor que 0,5%.
- Em contraste, a inclinação estimada para pós-menarca é 2,46, o que indica que a taxa de crescimento de pbf é aproximadamente 2,5%, quase cinco vezes a taxa do período pré-menarca.

Análise de Resíduos

Baseado na decomposição de Cholesky da matriz de covariância estimada, $\hat{\Sigma}_i$, podemos calcular os resíduos transformados,

$$r_i^* = L_i^{-1} r_i = L_i^{-1} (Y_i - X_i \hat{\beta}),$$

em que $\hat{\Sigma}_i = L_i L_i'$. Para fins ilustrativos, vamos também examinar os resíduos usuais (não transformados) e comparar gráficos de diagnóstico baseados nestes dois tipos de resíduos.

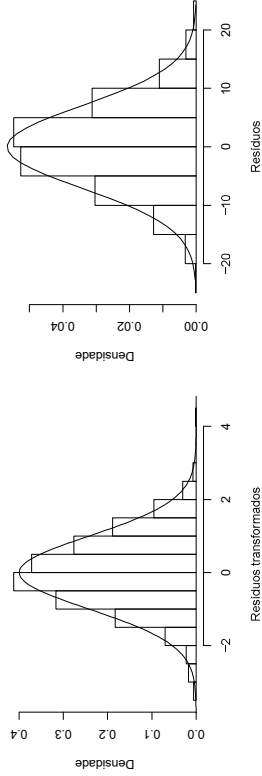
```
res=residuals(out, level=0) #resíduos usuais
pred=fitted(out, level=0) #preditos usuais
#
est.cov <- extract.lme.cov(out, dados)#extraí Cov(Y)
li <- t(chol(est.cov))#decomposição de Choleski de Cov(Y)
rest=solve(Li) %*%res#resíduos transformados
predt=solve(Li) %*%pred#preditos transformados
timet=solve(Li) %*%time#tempos transformados
```

A seguir os histogramas com curva normal

```

par(mfrow=c(1,2))
x = seq(min(rest)-1,max(rest)+1,by=0.05)
hist(rest,prob=T,xlab="Resíduos transformados",ylab="Densidade",main="")
lines(x,dnorm(x,mean=mean(rest),sd=sd(rest)))
x = seq(min(res)-1,max(res)+1,by=1)
hist(res,prob=T,xlab="Resíduos",ylab="Densidade",main="")
lines(x,dnorm(x,mean=mean(res),sd=sd(res)))

```



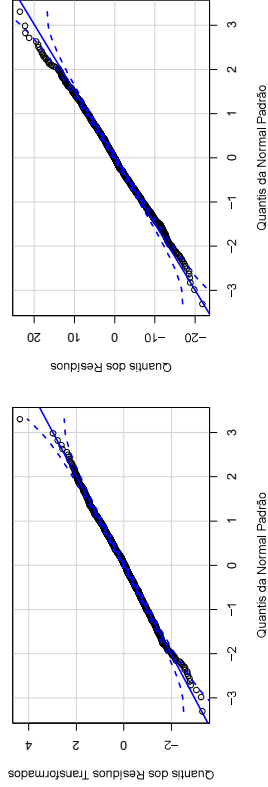
Não há indicação de assimetria nos histogramas dos resíduos e resíduos transformados.

Na sequência são apresentados gráficos de probabilidade normal dos resíduos.

```

par(mfrow=c(1,2))
qqPlot(rest,xlab="Quantis da Normal Padrão",ylab="Quantis dos Resíduos Transformados",
id=FALSE,main="")
qqPlot(res,xlab="Quantis da Normal Padrão",ylab="Quantis dos Resíduos",main="",id=FALSE)

```



Não notamos graves desvios de uma linha reta, com a quase totalidade dos pontos se distribuindo dentro dos limites de confiança. O gráfico dos resíduos transformados mostra, contudo, uma observação bastante extrema. Esta observação corresponde a uma garota (ID=128), obtida antes da menarca.

```
subset(dados,ID==128)
```

```

## ID age age_menarche time pbf time0
## 839 128 9.51 13.31 -3.80 27.60 0.00
## 840 128 10.47 13.31 -2.84 27.08 0.00
## 841 128 11.47 13.31 -1.84 44.14 0.00
## 842 128 12.47 13.31 -0.84 39.94 0.00
## 843 128 13.56 13.31 0.25 41.00 0.25
## 844 128 14.38 13.31 1.07 35.75 1.07
## 845 128 15.38 13.31 2.07 28.46 2.07

```

Aproximadamente dois anos antes da menarca, o percentual de gordura desta garota aumentou para 44% (de 27% um ano antes). Este é um valor extremamente alto, acima de 3 desvios padrões da média para o percentual de gordura pré-menarca. Esta observação não é um erro de medida. A garota teve percentual de gordura de 40% e 41% nas duas ocasiões seguintes.

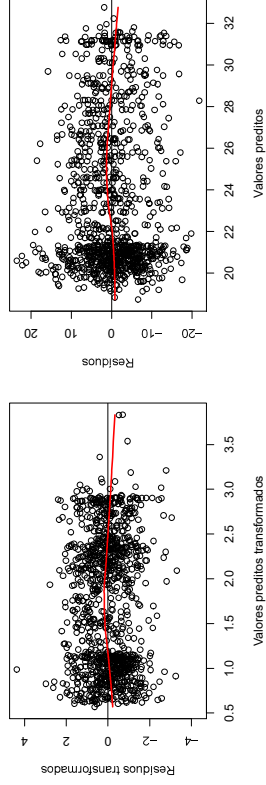
De modo geral, o número de resíduos extremos na figura não é maior do que esperaríamos ocorrer ao acaso, dado o total de 1049 observações. Assim, do exame dos resíduos não temos evidências de assimetria e a suposição de normalidade parece razoável.

A seguir, gráfico de dispersão dos resíduos (transformados e não transformados) versus valores ajustados (transformados e não transformados).

```

par(mfrow=c(1,2))
plot(predt,rest,xlab="Valores preditos transformados",ylab="Resíduos transformados",
ylim=c(-max(rest),max(rest)))
abline(h=0)
lines(lowess(predt,rest),type="l",lwd=2,col="red")
plot(pred,res,xlab="Valores preditos",ylab="Resíduos")
abline(h=0)
lines(lowess(pred,res),type="l",lwd=2,col="red")

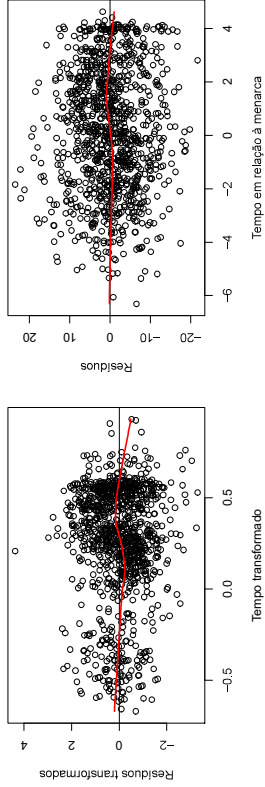
```



Não notamos padrões sistemáticos, com uma dispersão aleatória constante em torno do zero. Contudo, a curva suavizada revela alguma curvatura. Focando nos resíduos transformados, parece haver uma tendência quadrática, embora a queda na curva nos maiores valores dos dados preditos transformados tenha que ser interpretada com cuidado porque a curva ajustada é baseada em poucas observações nos extremos e é, portanto, instável naquela região.

Por causa da sugestão de curvatura, vamos examinar os gráficos de dispersão dos resíduos versus o tempo.

```
par(mfrow=c(1,2))
plot(timet,rest,xlab="Tempo transformado",ylab="Resíduos transformados")
abline(h=0)
lines(lowess(timet,rest),type="l",lwd=2,col="red")
plot(time,rest,xlab="Tempo em relação à menarca",ylab="Resíduos")
abline(h=0)
lines(lowess(time,rest),type="l",lwd=2,col="red")
```



Estes gráficos dos resíduos transformados e não transformados sugerem curvatura nos tempos (não transformados) pós-menarca. O padrão é mais aparente no gráfico dos resíduos transformados e não podem mais ser atribuídos à escassez de observações nos extremos. A curvatura sugere que o modelo para a resposta média pode ser melhorado pela inclusão de um termo quadrático para a tendência no período pós-menarca.

Modelo 2

Consideremos agora um modelo com termo quadrático no período pós-menarca. O modelo é dado por

$$E(Y_{ij}|b_i) = \beta_1 + \beta_2 t_{ij} + \beta_3 (t_{ij})_+ + \beta_4 (t_{ij})^2 + b_{2i} t_{ij} + b_{3i} (t_{ij})_+ + b_{4i} (t_{ij})^2,$$

em que

$$(t_{ij})_+^2 = \begin{cases} t_{ij}^2 & \text{se } t_{ij} > 0 \\ 0 & \text{se } t_{ij} \leq 0. \end{cases}$$

Neste modelo, cada garota tem uma curva de crescimento que pode ser descrita em termos de uma tendência linear para mudanças na resposta antes da menarca, e uma tendência quadrática para mudanças na resposta após a menarca.

```
dados$time2 = ifelse(dados$time>0,dados$time^2,0)
out2<-lme(pbf~time+time2, random= ~time+time0+time2|ID,data=dados)
summary(out2)
```

```
## Linear mixed-effects model fit by REML
## Data: dados
## AIC      BIC
## 6011.494 6085.771 -2990.747
##
## Random effects:
## Formula: ~time + time0 + time2 | ID
```

```
## Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev    Corr
## (Intercept) 6.932433 (Intr) time    time0
## time        1.316266  0.332
## time0       2.316876 -0.597 -0.503
## time2       0.342296  0.273 -0.385 -0.554
## Residual    2.833270
##
## Fixed effects: pbf ~ time + time0 + time2
##          Value Std.Error DF   t-value p-value
## (Intercept) 20.420151 0.5817186 884 35.10315 0.0000
## time       -0.015505 0.1612239 884 -0.09617 0.9234
## time0      4.843885 0.4055211 884 11.94484 0.0000
## time2     -0.646878 0.0771528 884 -8.38437 0.0000
## Correlation:
##          (Intr) time    time0
## time      0.399
## time0    -0.500 -0.660
## time2    0.263  0.212 -0.824
##
## Standardized Within-Group Residuals:
##          Min      Q1      Med      Q3      Max
## -2.957046827 -0.536979474 -0.007809063  0.536818414  3.495308391
##
## Number of Observations: 1049
## Number of Groups: 162
```

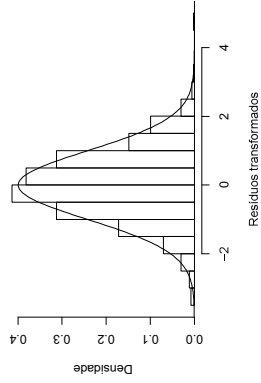
Vemos que o termo quadrático é significativo no ajuste. Procedemos com a análise dos resíduos.

Análise de Resíduos

```
res=residuals(out2, level=0)#resíduos usuais
pred=fitted(out2, level=0)#preditos usuais
#
est.cov <- extract.lme.cov(out2, dados)#extraí Cov(Y)
Li <- t(chol(est.cov))#decomposição de Choleski de Cov(Y)
rest=solve(Li) %*%res#resíduos transformados
predt=solve(Li) %*%pred#preditos transformados
timet=solve(Li) %*%time#tempos transformados
```

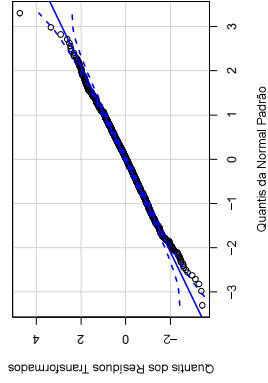
A seguir os histogramas com curva normal.

```
par(mfrow=c(1,2))
x = seq(min(rest)-1,max(rest)+1,by=0.05)
hist(rest,prob=T,xlab="Resíduos transformados",ylab="Densidade",main="")
lines(x,dnorm(x,mean=mean(rest),sd=sd(rest)))
x = seq(min(res)-1,max(res)+1,by=1)
hist(res,prob=T,xlab="Resíduos",ylab="Densidade",main="")
lines(x,dnorm(x,mean=mean(res),sd=sd(res)))
```



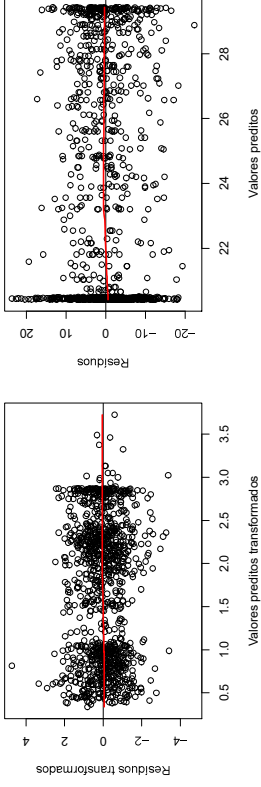
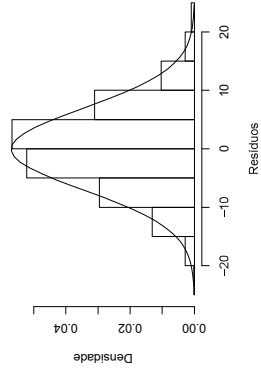
Gráficos de probabilidade normal dos resíduos:

```
par(mfrow=c(1,2))
qqPlot(rest,xlab="Quantis da Normal Padrão",ylab="Quantis dos Resíduos Transformados",
       id=FALSE, main="")
qqPlot(res,xlab="Quantis da Normal Padrão",ylab="Quantis dos Resíduos",main="", id=FALSE)
```



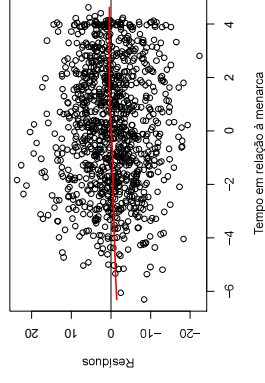
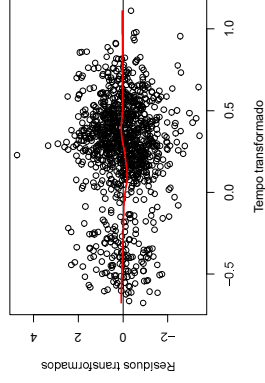
Gráficos de dispersão dos resíduos:

```
par(mfrow=c(1,2))
plot(predt,rest,xlab="Valores preditos transformados",ylab="Resíduos transformados",
     ylim=c(-max(rest),max(rest)))
abline(h=0)
lines(lowess(predt,rest),type="l",lwd=2,col="red")
plot(pred,res,xlab="Valores preditos",ylab="Resíduos")
abline(h=0)
lines(lowess(pred,res),type="l",lwd=2,col="red")
```



Não notamos curvatura nos resíduos transformados. Partimos para os gráficos de dispersão dos resíduos versus o tempo.

```
par(mfrow=c(1,2))
plot(timet,rest,xlab="Tempo transformado",ylab="Resíduos transformados")
abline(h=0)
lines(lowess(timet,rest),type="l",lwd=2,col="red")
plot(time,res,xlab="Tempo em relação à menarca",ylab="Resíduos")
abline(h=0)
lines(lowess(time,res),type="l",lwd=2,col="red")
```



A curvatura observada anteriormente foi eliminada com a inclusão do termo quadrático no modelo. Pela análise dos resíduos, concluímos pela adequação do segundo modelo proposto.