

Análise dos dados ‘FEV’

Arthur Cesar Rocha (94361)

8 de outubro de 2018

Introdução

A professora Dra. Isolde Previdelli nos passou um banco de dados FEV para aprimorarmos os conhecimentos nas aplicações de modelos mistos. O banco de dados conta com a variável resposta FEV, referente a uma qualidade respiratória, que foi medida ao longo de 8 horas além da medição pré intervenção das drogas A,C e placebo. Por se tratar de um problema de dados longitudinais (a ordem do tempo importa), foram feitas algumas análises, além da descritiva, com abordagens diferentes, sendo algumas equivocadas (não considerando a dependência) e a parte de modelos mistos. Os modelos lineares mistos são uma extensão dos modelos lineares, em que é possível assumir uma parte de efeitos aleatórios no modelo além da parte de efeitos fixos, permitindo assim a composição de uma estrutura de covariância (dependência) na análise.

Leitura e manejo dos dados

Pacotes

```
library(dplyr)      #Manipulação
library(magrittr)    #Manipulação
library(tidyr)       #Transformar dados para long
library(ggplot2)     #Gráficos
library(corrplot)    #Gráfico matriz de correlação
library(nlme)        #Modelos mistos
```

Análise descritiva

O primeiro passo de uma análise é verificar o comportamento dos dados de forma descritiva, portanto essa sessão é dedicada a esse fim.

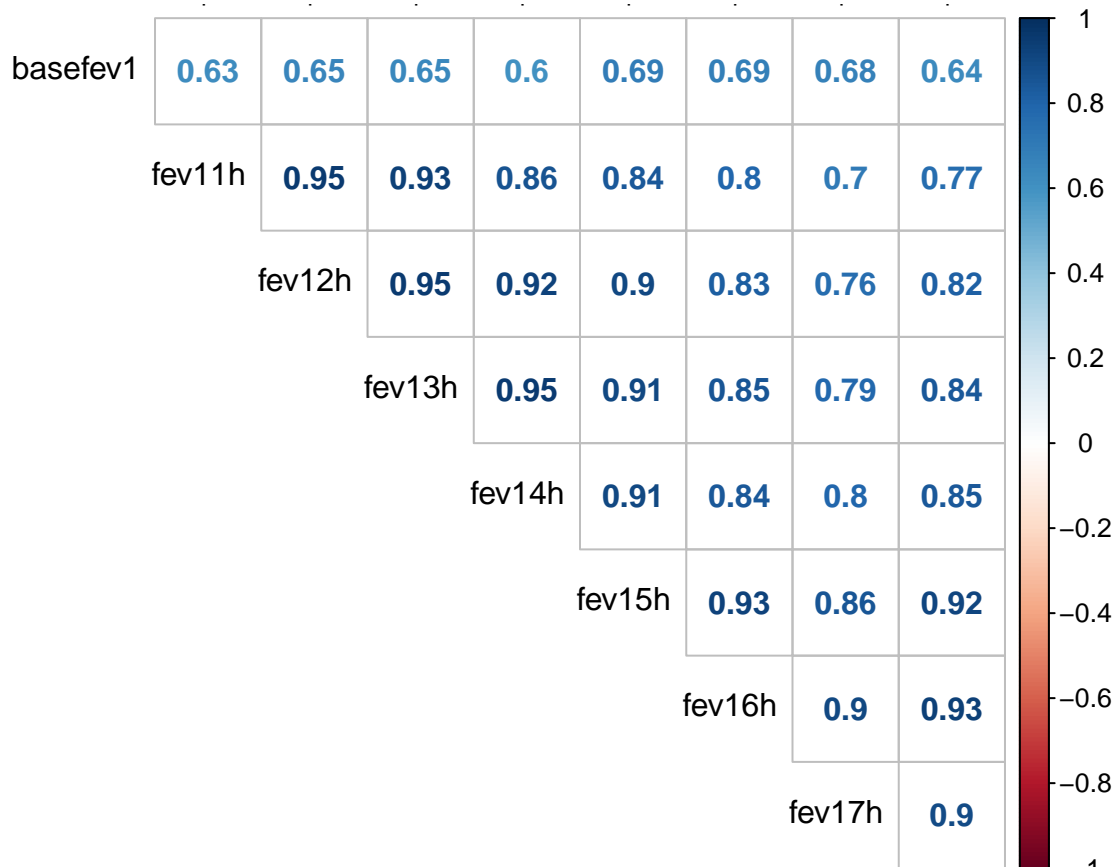
Matriz de correlação amostral:

```
##Matriz de correlação amostral
iniciais %>%
  select(-c(patient,drug)) %>%
  cor() %>%
  round(digits = 2)
```

```
##      basefev1 fev11h fev12h fev13h fev14h fev15h fev16h fev17h fev18h
## basefev1      1.00   0.63   0.65   0.65   0.60   0.69   0.69   0.68   0.64
## fev11h        0.63   1.00   0.95   0.93   0.86   0.84   0.80   0.70   0.77
## fev12h        0.65   0.95   1.00   0.95   0.92   0.90   0.83   0.76   0.82
## fev13h        0.65   0.93   0.95   1.00   0.95   0.91   0.85   0.79   0.84
## fev14h        0.60   0.86   0.92   0.95   1.00   0.91   0.84   0.80   0.85
## fev15h        0.69   0.84   0.90   0.91   0.91   1.00   0.93   0.86   0.92
## fev16h        0.69   0.80   0.83   0.85   0.84   0.93   1.00   0.90   0.93
```

```
## fev17h      0.68  0.70  0.76  0.79  0.80  0.86  0.90  1.00  0.90
## fev18h      0.64  0.77  0.82  0.84  0.85  0.92  0.93  0.90  1.00
```

```
#Gráfico da matriz de correlação
iniciais %>%
  select(-c(patient,drug)) %>%
  cor() %>%
  corrplot(type = "upper",method = "number",
           tl.col = "black",diag = F)
```



Verifica-se que parece ser razoável aceitar que a correlação é aproximadamente constante entre os tempos próximos, ou que decai lentamente conforme o tempo, o que faz sentido do ponto de vista de dados longitudinais.

Medidas descritivas de cada grupo

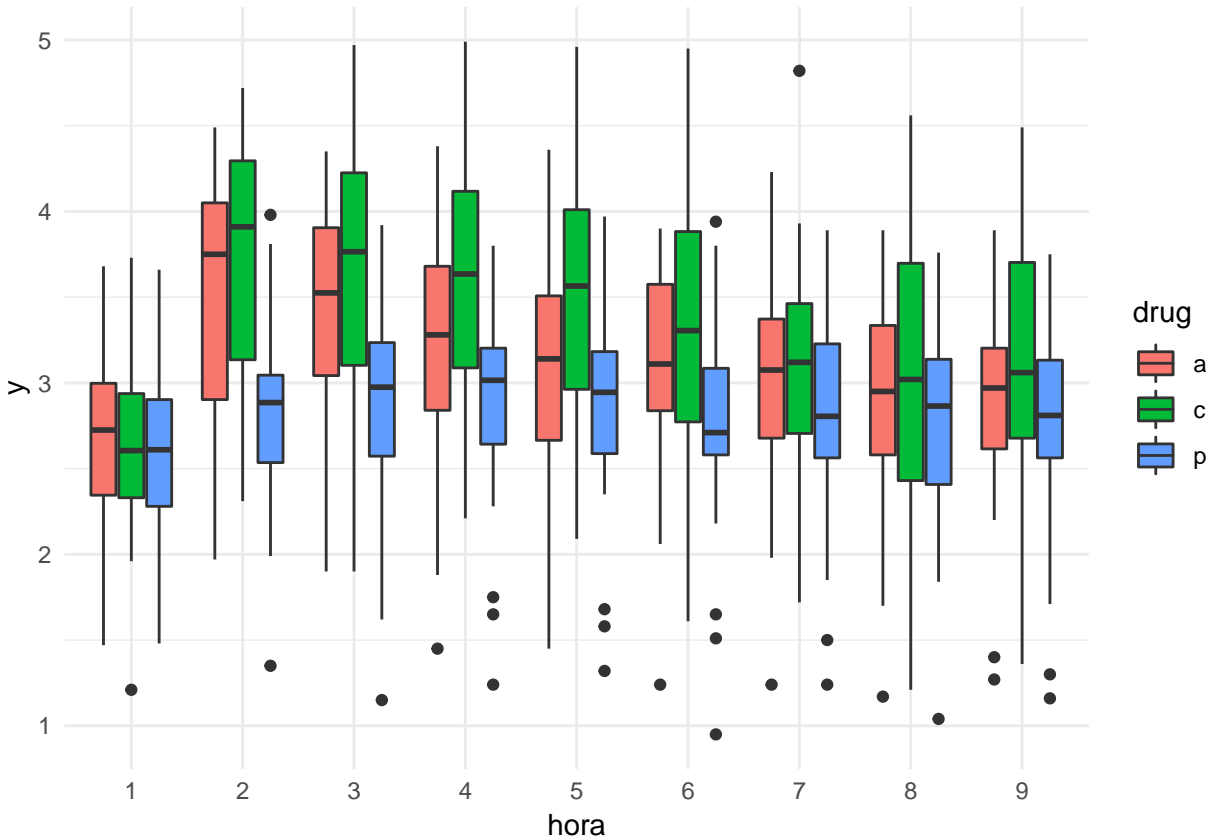
```
dados %$%
  tapply(y, drug, summary)
```

```
## $a
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.170  2.688   3.090   3.072  3.590   4.490
##
## $c
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.210  2.688   3.225   3.251  3.922   4.990
```

```
##
## $p
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.950   2.455   2.845   2.792   3.175   3.980
```

Nota-se um desempenho aparentemente superior da droga a e c, não tendo muita diferença em suas distribuições, ao contrário da droga p, que apresentou as medidas de posição um pouco abaixo das demais.

```
dados %>%
  ggplot(aes(hora,y,fill=drug))+
  geom_boxplot()+
  theme_minimal()
```

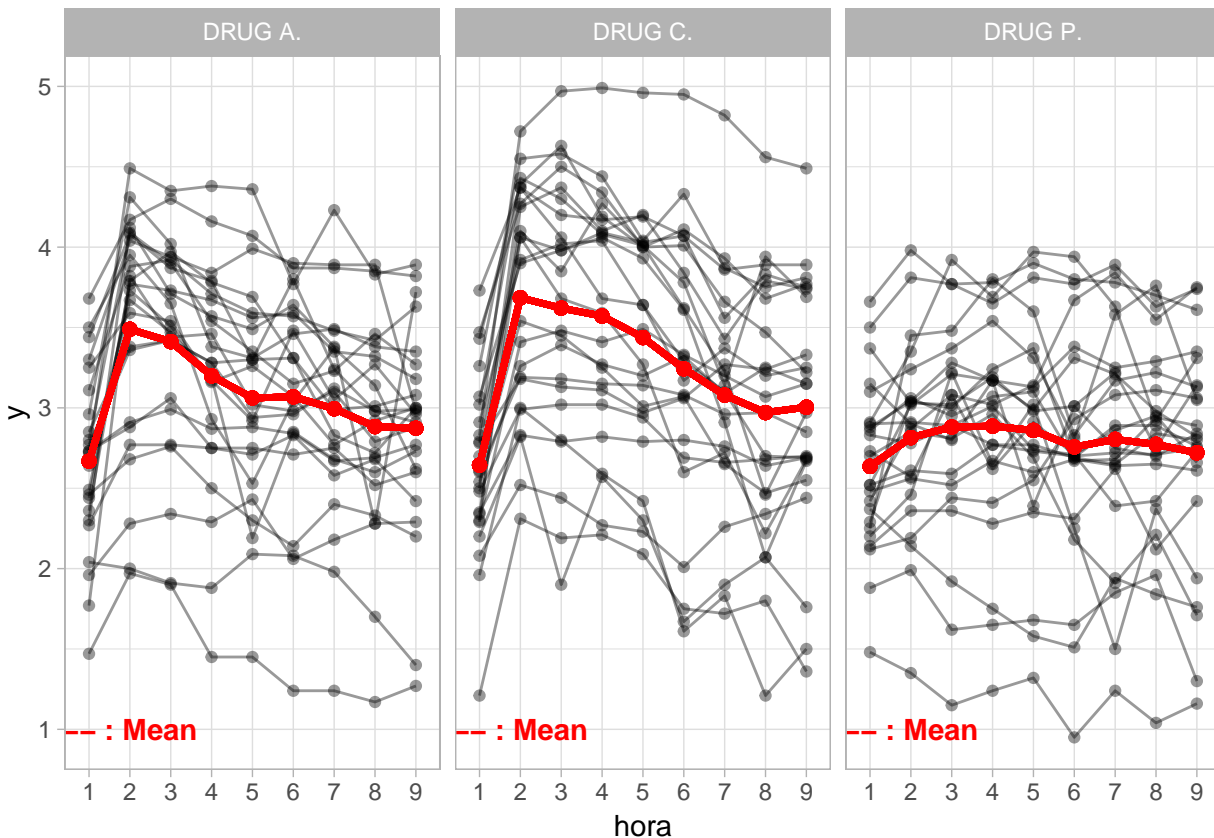


Ao se analisar o comportamento da resposta conforme a droga e a hora, nota-se que há um pico na resposta na hora de aplicação (hora 2) e depois há uma queda e estabilização.

##Gráfico de perfis

```
dados %>%
  mutate(drug=factor(drug,levels = levels(drug),
                     labels = c("DRUG A.", "DRUG C.", "DRUG P."))) %>%
  group_by(drug,hora) %>%
  mutate(MEAN=mean(y,na.rm = T)) %>%
  ggplot(aes(x=hora,y=y,group=patient))+
  geom_point(alpha=.4)+
  geom_line(alpha=.4)+
  geom_line(aes(x=hora,y=MEAN),col=2, size=1.2,show.legend = T)+
  geom_point(aes(x=hora,y=MEAN),col=2, size=2)+
```

```
theme_light()+
facet_wrap(~drug)+
annotate(geom="text",x=2,y=1,label="-- : Mean",col=2,fontface=2)
```



A partir dos gráficos de perfis é possível perceber que o intercepto é diferente para cada indivíduo, isto é, cada paciente começa com uma resposta distinta. Além disso, constata-se o que foi visto anteriormente de que as maiores médias da resposta são de indivíduos que utilizaram a droga C e A, respectivamente.

Modelos

Primeiramente será feita uma análise considerando modelos mais simples e posteriormente modelos mistos serão aplicados nos dados

Análises de variâncias

```
## - ANOVA one way - errado -
```

```
modelo1=aov(y~drug,data = dados1)
modelo1 %>%
summary()
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## drug       2  25.78  12.891   24.51 6.1e-11 ***
## Residuals 573 301.35   0.526
```

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

```
## - ANOVA two way - errado -
```

```
modelo2=aov(y~drug+hora,data = dados1)
modelo2 %>% summary()
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## drug           2  25.78   12.891   25.675 2.12e-11 ***
## hora           7   17.17    2.453    4.885 2.28e-05 ***
## Residuals     566 284.18    0.502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Esses dois modos equivocados de análise, por conta de não considerar a dependência presente nos dados, tiveram resultados similares, em que foi possível averiguar significância no efeito da droga e da hora.

```
## - ANOVA com indivíduo como bloco -
```

```
modelo3=aov(y~drug*hora+patient,data = dados1)
modelo3 %>% summary()
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## drug           2  25.78   12.891 204.212 < 2e-16 ***
## hora           7   17.17    2.453  38.857 < 2e-16 ***
## patient        69 247.41    3.586   56.801 < 2e-16 ***
## drug:hora      14   6.28    0.449    7.106 1.92e-13 ***
## Residuals     483  30.49    0.063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ao se considerar o paciente como bloco, é possível retirar um possível efeito de dependência.

```
## - ANOVA para medidas repetidas -
```

```
modelo4=aov(y ~ basefev1+ drug*hora + Error(patient/hora),data=dados1)
modelo4 %>% summary()
```

```
##
## Error: patient
##              Df Sum Sq Mean Sq F value    Pr(>F)
## basefev1      1 131.89   131.89  76.984 8.64e-13 ***
## drug          2   24.81    12.41   7.241 0.00141 **
## Residuals    68 116.50     1.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: patient:hora
##              Df Sum Sq Mean Sq F value    Pr(>F)
## hora          7   17.17    2.4529  38.857 < 2e-16 ***
## drug:hora     14   6.28    0.4486   7.106 1.92e-13 ***
## Residuals    483  30.49    0.0631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Considerando que a estrutura de covariância é esférica, torna-se viável considerar esse modelo. É possível

perceber significância nos efeitos fixos e também no erro composto.

Modelos mistos:

Modelo com estrutura de covariância Compound Symmetry

```
mixed1=lme(y ~ basefev1+hora*drug, random= ~1 | patient, method="REML",correlation = corCompSymm(),data=dados)
anova(mixed1)
```

##	numDF	denDF	F-value	p-value
## (Intercept)	1	483	3204.098	<.0001
## basefev1	1	68	76.984	<.0001
## hora	7	483	38.857	<.0001
## drug	2	68	7.241	0.0014
## hora:drug	14	483	7.106	<.0001

```
cat("\n AIC:",AIC(mixed1))
```

```
##
## AIC: 403.2902
```

Considerando que a estrutura de covariância é esférica, torna-se viável considerar esse modelo. É possível perceber significância nos efeitos fixos e aleatórios.

Modelo com estrutura de covariância AR(1)

```
mixed2=lme(y ~ basefev1+hora*drug, random= ~1 | patient, method="REML",correlation = corAR1(),data=dados)
anova(mixed2)
```

##	numDF	denDF	F-value	p-value
## (Intercept)	1	483	3245.828	<.0001
## basefev1	1	68	76.574	<.0001
## hora	7	483	17.101	<.0001
## drug	2	68	7.783	9e-04
## hora:drug	14	483	3.942	<.0001

```
cat("\n AIC:",AIC(mixed2))
```

```
##
## AIC: 303.0306
```

Modelo com estrutura de covariância não estruturada

Não convergiu

Modelo com estrutura de covariância não estruturada

Não convergiu

Teste da razão entre verossimilhanças

Para verificar qual a estrutura de covariância pode ser escolhida é usual utilizar o teste da razão entre verossimilhanças, desde que se tratem de estruturas “aninhadas”, isto é, uma pode ser escrita como caso particular da outra.

```
a=anova.lme(mixed1,mixed2,test = T)
```

Adotando a classe de modelos mistos com estruturas de covariância CS e AR(1), verifica-se que o segundo tem menor AIC e pela forma da estrutura da matriz de correlação amostral parece ser mais adequado.