

Trabalho de Dados Reais - Análise de Dados Longitudinais

Helen Lourenço - Maurício Amaral - Vitor Kroeff

06/12/2024

Introdução

ESCREVER

```
# Pacotes utilizados
pacman::p_load(reshape, plyr, ggplot2, gridExtra, mice, geepack, nlme,
               dplyr, GGally, tidyr, modelsummary)
```

Dados

```
dados_brutos <- tibble(read.table('base_de_dados.txt', header = T))
```

Correção na informação da variável **n56hpo** e a criação de um id único para os pacientes

```
# Correção na linha 6 da variável n56hpo
dados_brutos$n56hpo[6] <- 1130

# Criação de um id para os pacientes
dados_brutos$id <- 1:length(dados_brutos$idade)
```

Seleção das variáveis escolhidas para o estudo

```
dados <- dados_brutos %>% select(c(id,sexo, idade,imc, fc,
                                nyha, has, iap, ai, grupo,
                                euroes, fc, creat,
                                t1, t2, t3, t4, t5, t6,
                                #n1anest, n2despin,n3final,
                                #n42hpo, n56hpo,n624hpo
                                ))
```

Tratamento das informações categóricas

```
# Tratamento dos dados categoricos
dados$sexo <- as.factor(ifelse(dados$sexo== 1, 'M', 'F' ))
dados$nyha <- as.factor(dados$nyha)
dados$has <- as.factor(dados$has)
dados$iap <- as.factor(dados$iap)
dados$ai <- as.factor(dados$ai)
dados$grupo <- as.factor(dados$grupo)
```

Descrição da base de dados

Head da base

```
knitr::kable(
head(dados,5))
```

id	sexo	idade	imc	fc	nyha	has	iap	ai	grupo	euroes	creat	t1	t2	t3	t4	t5	t6
1	M	64	20.2	70	2	1	0	1	0	6	1.2	2595.42	2127.32	2871.93	3171.05	4425.33	4592.328
2	M	74	25.6	70	1	1	0	0	0	3	1.1	1251.05	1120.12	1520.17	2625.28	4221.14	2767.505
3	M	20	19.7	90	4	0	0	0	2	5	1.1	279.258	258.436	796.955	590.95	723.514	1258.729
4	F	68	21.0	70	3	1	0	0	2	6	1.1	1394.00	2205.19	3524.08	74095.74	3198.84	3006.303
5	F	51	23.9	70	2	1	0	0	0	2	0.9	728.071	1235.74	1747.37	2573.23	2082.71	2452.582

Descrição da base

```
# Descrições da base
str(dados)
```

```
## tibble [62 x 18] (S3: tbl_df/tbl/data.frame)
## $ id      : int [1:62] 1 2 3 4 5 6 7 8 9 10 ...
## $ sexo    : Factor w/ 2 levels "F","M": 2 2 2 1 1 2 1 1 2 2 ...
## $ idade   : int [1:62] 64 74 20 68 51 60 75 53 75 42 ...
## $ imc     : num [1:62] 20.2 25.6 19.7 21 23.9 19 26.6 36.6 25.9 35.4 ...
## $ fc      : int [1:62] 70 70 90 70 70 70 80 80 70 80 ...
## $ nyha    : Factor w/ 4 levels "1","2","3","4": 2 1 4 3 2 1 3 3 1 1 ...
## $ has     : Factor w/ 2 levels "0","1": 2 2 1 2 2 2 2 2 2 2 ...
## $ iap     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
## $ ai      : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 2 1 1 1 ...
## $ grupo   : Factor w/ 3 levels "0","1","2": 1 1 3 3 1 1 1 1 1 1 ...
## $ euroes  : int [1:62] 6 3 5 6 2 0 8 4 5 0 ...
## $ creat   : num [1:62] 1.2 1.1 1.1 1.1 0.9 1.4 0.8 1.4 0.7 0.8 ...
## $ t1      : num [1:62] 2595 1251 279 1394 728 ...
## $ t2      : num [1:62] 2127 1120 258 2205 1236 ...
## $ t3      : num [1:62] 2872 1520 797 3524 1747 ...
## $ t4      : num [1:62] 3171 2625 591 4096 2573 ...
## $ t5      : num [1:62] 4425 4221 724 3199 2083 ...
## $ t6      : num [1:62] 4592 2768 1259 3006 1453 ...
```

Estrutura de correlação

Correlação entre as medições da citocina Tnf-receptor geral. De maneira geral podemos observar que as correlações entre as medidas da citocina t, apresentam uma queda com o passar do tempo. Essa informação pode ser um indicativo de uma estrutura de correlação do tipo AR(1) ** MELHORAR!!**

```
## Correlação de todos os pacientes
round(cor(dados[,13:16]),2)
```

```
##      t1    t2    t3    t4
## t1 1.00 0.74 0.67 0.61
## t2 0.74 1.00 0.69 0.62
## t3 0.67 0.69 1.00 0.72
## t4 0.61 0.62 0.72 1.00
```

Correlação no Grupo 0

```
## Correlação do grupo 0
round(cor(subset(dados, grupo == 0)[,13:16]),2)
```

```
##      t1  t2  t3  t4
## t1 1.00 0.71 0.71 0.51
## t2 0.71 1.00 0.62 0.49
## t3 0.71 0.62 1.00 0.56
## t4 0.51 0.49 0.56 1.00
```

Correlação no Grupo 1

```
## Correlação do grupo 1
round(cor(subset(dados, grupo == 1)[,13:16]),2)
```

```
##      t1  t2  t3  t4
## t1 1.00 0.74 0.67 0.70
## t2 0.74 1.00 0.80 0.70
## t3 0.67 0.80 1.00 0.82
## t4 0.70 0.70 0.82 1.00
```

Correlação no Grupo 2

```
## Correlação do grupo
round(cor(subset(dados, grupo == 2)[,13:16]),2)
```

```
##      t1  t2  t3  t4
## t1 1.00 0.80 0.58 0.63
## t2 0.80 1.00 0.67 0.74
## t3 0.58 0.67 1.00 0.73
## t4 0.63 0.74 0.73 1.00
```

Transformação para dados Longos

Tempos = t1, t2, ..., t6 discretos **Explicar o tempo**

```
dados_longos <- dados %>% pivot_longer(
  cols = starts_with("t"),
  names_to = "tempo",
  values_to = "citocina_t",
  names_prefix = "t"
)
knitr::kable(
  head(dados_longos), caption = 'Dados em formato longo')
```

Table 2: Dados em formato longo

id	sexo	idade	imc	fc	nyha	has	iap	ai	grupo	euroes	creat	tempo	citocina_t
1	M	64	20.2	70	2	1	0	1	0	6	1.2	1	2595.429
1	M	64	20.2	70	2	1	0	1	0	6	1.2	2	2127.329
1	M	64	20.2	70	2	1	0	1	0	6	1.2	3	2871.933
1	M	64	20.2	70	2	1	0	1	0	6	1.2	4	3171.050
1	M	64	20.2	70	2	1	0	1	0	6	1.2	5	4425.335
1	M	64	20.2	70	2	1	0	1	0	6	1.2	6	4592.328

Gráficos

```
# Gráficos
rotulos <- c('1', '2', '3',
```

```

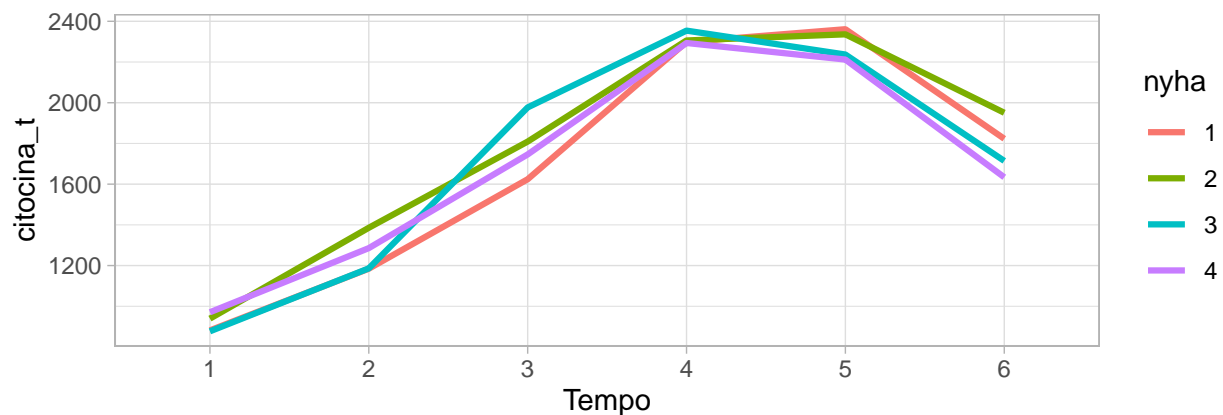
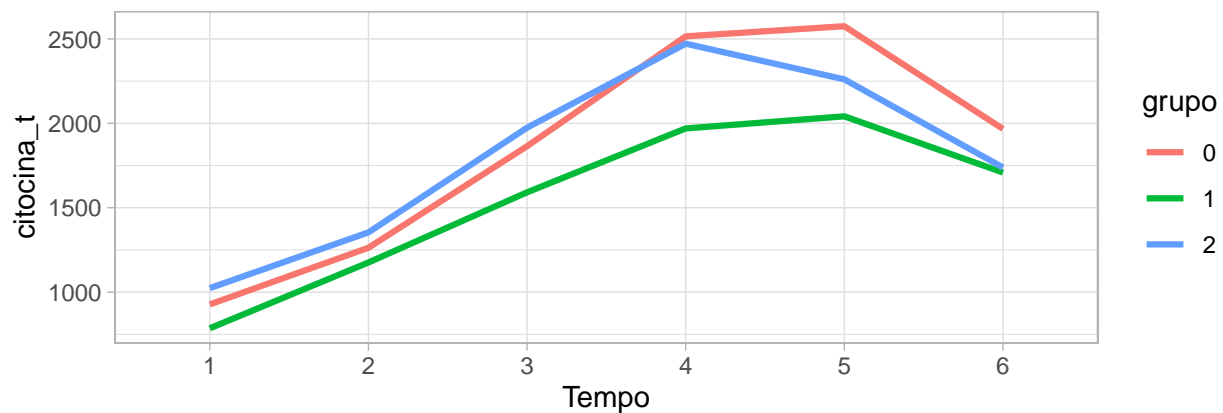
'4', '5', '6')

grafico1 <- ggplot(dados_longos, aes(x = tempo, y = citocina_t, group = grupo, shape = grupo, color = grupo)) +
  stat_summary(fun = "mean", geom = "line", linewidth = 1.1) +
  theme(legend.position = "top") +
  labs(x = "Tempo") +
  theme_light()

grafico2 <- ggplot(dados_longos, aes(x = tempo, y = citocina_t, group = nyha, shape = nyha, color = nyha)) +
  stat_summary(fun = "mean", geom = "line", linewidth = 1.1) +
  theme(legend.position = "top") +
  labs(x = "Tempo") +
  theme_light()

grid.arrange(grafico1, grafico2, ncol = 1)

```



```

grafico3 <-
  ggplot(dados_longos, aes(x = tempo, y = citocina_t, fill = has)) +
  geom_boxplot(notch = TRUE) + theme(legend.position = "top") +
  stat_summary(
    fun = "mean",
    geom = "point",
    size = 2,
    color = "white",
    position = position_dodge(width = 0.75),
    show.legend = FALSE
  ) +

```

```

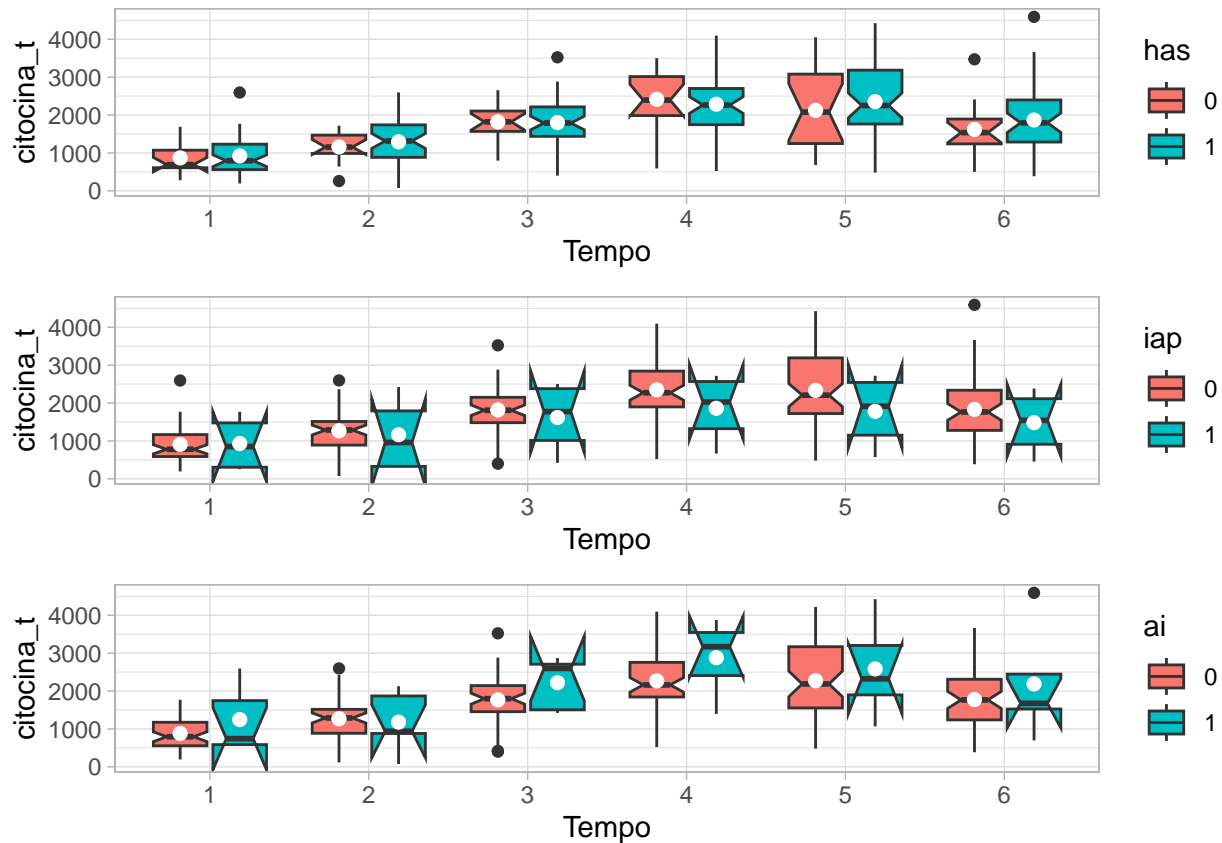
labs(x = "Tempo") + theme_light() +
scale_x_discrete(breaks = seq_along(rotulos), labels = rotulos)+theme(plot.title = element_text(hju

grafico4 <-
  ggplot(dados_longos, aes(x = tempo, y = citocina_t, fill = iap)) +
  geom_boxplot(notch = TRUE) + theme(legend.position = "top") +
  stat_summary(
    fun = "mean",
    geom = "point",
    size = 2,
    color = "white",
    position = position_dodge(width = 0.75),
    show.legend = FALSE
  ) +
  labs(x = "Tempo") + theme_light() +
  scale_x_discrete(breaks = seq_along(rotulos), labels = rotulos)+theme(plot.title = element_text(hju

grafico5 <-
  ggplot(dados_longos, aes(x = tempo, y = citocina_t, fill = ai)) +
  geom_boxplot(notch = TRUE) + theme(legend.position = "top") +
  stat_summary(
    fun = "mean",
    geom = "point",
    size = 2,
    color = "white",
    position = position_dodge(width = 0.75),
    show.legend = FALSE
  ) +
  labs(x = "Tempo") + theme_light() +
  scale_x_discrete(breaks = seq_along(rotulos), labels = rotulos)+theme(plot.title = element_text(hju

grid.arrange(grafico3, grafico4, grafico5, ncol = 1)

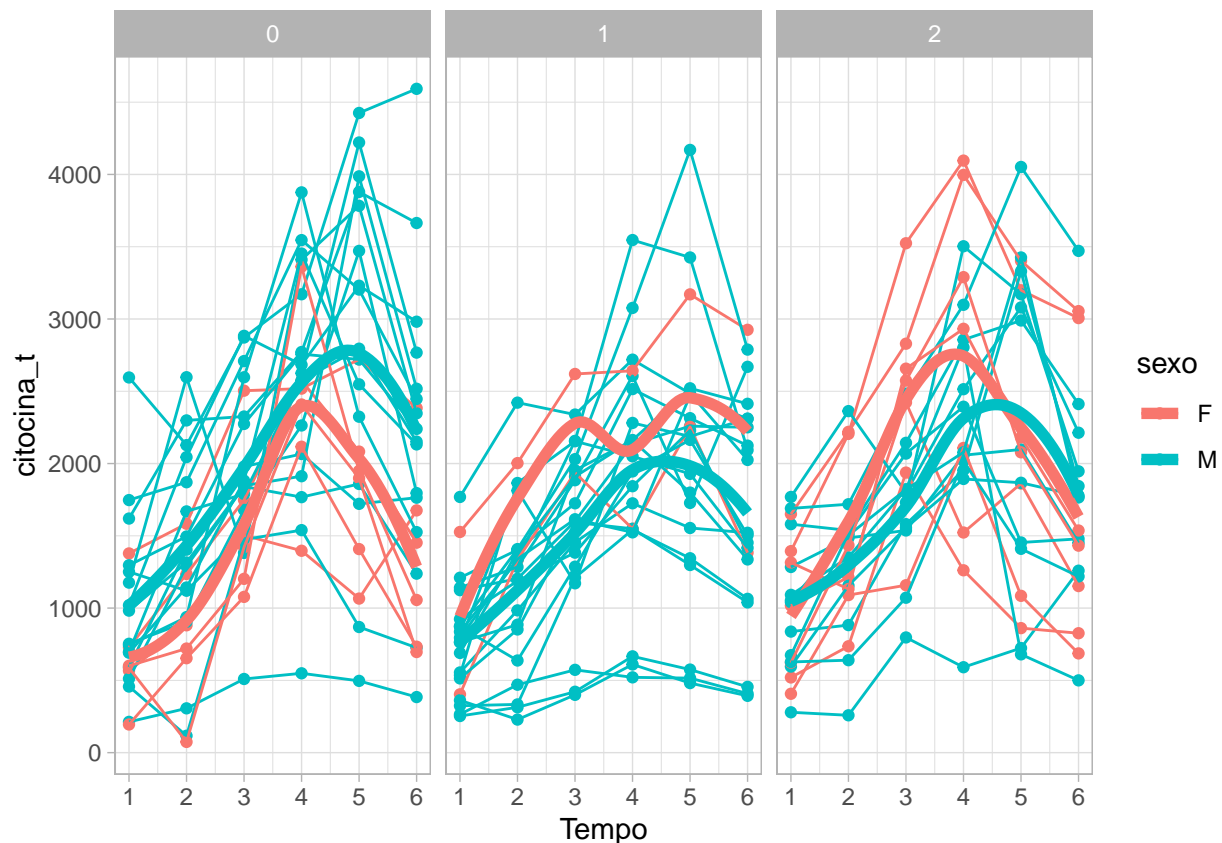
```



```
dados_longos$obs_continua <- as.numeric(dados_longos$tempo) #Criando tempo continuo
grafico6 <-ggplot(dados_longos, aes(x=obs_continua, y= citocina_t, color=sexo))+
  geom_point()+
  geom_line(aes(group=id))+ theme(legend.position="top")+facet_wrap(~grupo) +
  labs(x="Tempo") + theme_light()+
  scale_x_continuous(breaks = seq_along(rotulos), labels = rotulos) + theme(plot.title = element_text(
grafico6 + geom_smooth(method = "loess", se = FALSE, size = 2)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `geom_smooth()` using formula = 'y ~ x'
```



```
### excluimos da base a variável contínua
dados_longos <- dados_longos %>% select(-c(obs_continua))
```

Modelos Marginais (GEE)

Para iniciar a modelagem dos dados, optamos por utilizar os Modelos de Equações de Estimação Generalizadas (GEE) como modelo marginal. Ajustaremos quatro modelos GEE, cada um assumindo uma estrutura de correlação diferente: Independente, Simetria Composta, AR(1) e Não Estruturada. Nesse processo, utilizaremos todas as variáveis selecionadas, sem incluir, neste momento, termos de interação entre as variáveis.

A equação dos modelos abaixo modelo ser definida como:

$$ESCREVERALGO\beta$$

Independente

```
ajuste_gee_indep <- geeglm(citocina_t ~ tempo + ai + sexo + nyha + grupo + idade + has + euroes + imc +
                           data = dados_longos, corstr = 'independence', id = id, family = 'gaussian')

round(coef(summary(ajuste_gee_indep)), 3)
```

	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-402.140	893.669	0.202	0.653
## tempo2	352.444	50.909	47.929	0.000
## tempo3	896.709	60.078	222.779	0.000

## tempo4	1408.408	85.044	274.266	0.000
## tempo5	1389.107	98.021	200.834	0.000
## tempo6	900.528	71.824	157.199	0.000
## ai1	406.664	370.763	1.203	0.273
## sexoM	127.000	192.801	0.434	0.510
## nyha2	212.307	216.756	0.959	0.327
## nyha3	49.942	240.360	0.043	0.835
## nyha4	122.510	266.263	0.212	0.645
## grupo1	-291.072	206.225	1.992	0.158
## grupo2	459.941	329.615	1.947	0.163
## idade	16.913	8.088	4.373	0.037
## has1	294.079	231.393	1.615	0.204
## euroes	-39.966	68.361	0.342	0.559
## imc	-1.596	27.753	0.003	0.954
## iap1	-105.313	467.494	0.051	0.822

Simetria Composta

```
ajuste_gee_simetria <- geeglm(citocina_t ~ tempo + ai + sexo + nyha + grupo + idade + has + euroes + imc + iap1,
                             data = dados_longos, corstr = 'exchangeable', id = id, family = 'gaussian')
round(coef(summary(ajuste_gee_simetria)),3)
```

##	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-402.140	893.669	0.202	0.653
## tempo2	352.444	50.909	47.929	0.000
## tempo3	896.709	60.078	222.779	0.000
## tempo4	1408.408	85.044	274.266	0.000
## tempo5	1389.107	98.021	200.834	0.000
## tempo6	900.528	71.824	157.199	0.000
## ai1	406.664	370.763	1.203	0.273
## sexoM	127.000	192.801	0.434	0.510
## nyha2	212.307	216.756	0.959	0.327
## nyha3	49.942	240.360	0.043	0.835
## nyha4	122.510	266.263	0.212	0.645
## grupo1	-291.072	206.225	1.992	0.158
## grupo2	459.941	329.615	1.947	0.163
## idade	16.913	8.088	4.373	0.037
## has1	294.079	231.393	1.615	0.204
## euroes	-39.966	68.361	0.342	0.559
## imc	-1.596	27.753	0.003	0.954
## iap1	-105.313	467.494	0.051	0.822

AR(1)

```
ajuste_gee_ar1 <- geeglm(citocina_t ~ tempo + ai + sexo + nyha + grupo + idade + has + euroes + imc + iap1,
                         data = dados_longos, corstr = 'ar1', id = id, family = 'gaussian')
round(coef(summary(ajuste_gee_ar1)),3)
```

##	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-382.116	887.771	0.185	0.667
## tempo2	352.444	50.909	47.929	0.000
## tempo3	896.709	60.078	222.779	0.000
## tempo4	1408.408	85.044	274.266	0.000

## tempo5	1389.107	98.021	200.834	0.000
## tempo6	900.528	71.824	157.199	0.000
## ai1	381.205	375.076	1.033	0.309
## sexoM	260.163	201.637	1.665	0.197
## nyha2	208.722	210.302	0.985	0.321
## nyha3	30.979	232.160	0.018	0.894
## nyha4	92.926	250.706	0.137	0.711
## grupo1	-224.679	208.395	1.162	0.281
## grupo2	413.806	321.093	1.661	0.197
## idade	12.384	7.410	2.793	0.095
## has1	359.633	214.950	2.799	0.094
## euros	1.753	68.765	0.001	0.980
## imc	-2.778	28.239	0.010	0.922
## iap1	-110.470	462.678	0.057	0.811

Não Estruturada

```
ajuste_gee_unstructured <- geeglm(citocina_t ~ tempo + ai + sexo + nyha + grupo + idade + has + euros + imc + iap1,
                                   data = dados_longos, corstr = 'unstructured', id = id, family = 'gaussian')
round(coef(summary(ajuste_gee_unstructured)), 3)
```

##	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-583.926	699.755	0.696	0.404
## tempo2	352.444	50.909	47.929	0.000
## tempo3	896.709	60.078	222.779	0.000
## tempo4	1408.408	85.044	274.266	0.000
## tempo5	1389.107	98.021	200.834	0.000
## tempo6	900.528	71.824	157.199	0.000
## ai1	619.833	302.804	4.190	0.041
## sexoM	-277.163	133.451	4.313	0.038
## nyha2	105.922	158.306	0.448	0.503
## nyha3	-179.649	198.084	0.823	0.364
## nyha4	-46.952	226.081	0.043	0.835
## grupo1	-141.162	136.601	1.068	0.301
## grupo2	907.326	258.257	12.343	0.000
## idade	16.431	7.404	4.925	0.026
## has1	220.758	220.274	1.004	0.316
## euros	-82.647	50.076	2.724	0.099
## imc	23.438	19.122	1.502	0.220
## iap1	164.805	360.280	0.209	0.647

Seleção de Variáveis e Interações

De maneira geral, observamos que os modelos estimaram coeficientes muito próximos, tanto em termos de magnitude quanto em relação à significância estatística de cada variável. Diante disso, optamos por adotar o modelo GEE com estrutura de correlação **AR(1)**. Essa escolha se baseia na análise das correlações das medidas da citocina TNF-receptor nos diferentes tempos, que indicaram um padrão de decaimento conforme o intervalo entre as medições aumenta.

Seleção de Variáveis

Com base no resultado da função `summary()` aplicada ao modelo GEE com estrutura de correlação **AR(1)** (exibido novamente abaixo), observamos que as seguintes variáveis não apresentam significância estatística a

um nível de 5%: **Variável A**, **Variável C**, **Variável D**, **Variável E** e **Variável G**.

No entanto, as variáveis **A**, **C** e **D** possuem relevância clínica no contexto do estudo. Por esse motivo, decidimos mantê-las no modelo, mesmo que não tenham demonstrado significância no modelo ajustado

```
round(coef(summary(ajuste_gee_ar1)),3)
```

##	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-382.116	887.771	0.185	0.667
## tempo2	352.444	50.909	47.929	0.000
## tempo3	896.709	60.078	222.779	0.000
## tempo4	1408.408	85.044	274.266	0.000
## tempo5	1389.107	98.021	200.834	0.000
## tempo6	900.528	71.824	157.199	0.000
## ai1	381.205	375.076	1.033	0.309
## sexoM	260.163	201.637	1.665	0.197
## nyha2	208.722	210.302	0.985	0.321
## nyha3	30.979	232.160	0.018	0.894
## nyha4	92.926	250.706	0.137	0.711
## grupo1	-224.679	208.395	1.162	0.281
## grupo2	413.806	321.093	1.661	0.197
## idade	12.384	7.410	2.793	0.095
## has1	359.633	214.950	2.799	0.094
## euros	1.753	68.765	0.001	0.980
## imc	-2.778	28.239	0.010	0.922
## iap1	-110.470	462.678	0.057	0.811

Interação de Variáveis

Com base no modelo inicial e nas variáveis previamente selecionadas, decidimos testar termos de interação para avaliar se eles possuem relevância no contexto do estudo. O primeiro termo de interação a ser analisado é o efeito entre **Grupo** e o **tempo**.

Abaixo, apresentamos a equação do modelo ajustado e o resultado do `summary()` correspondente.

Equação2

```
ajuste_gee_ar1_grupotempo <- geeglm(citocina_t ~ tempo*grupo + sexo + imc + nyha + idade + euros,
                                   data = dados_longos, corstr = 'ar1', id = id, family = 'gaussian')
round(coef(summary(ajuste_gee_ar1_grupotempo)),4)
```

##	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	258.4699	826.9088	0.0977	0.7546
## tempo2	335.7087	100.9778	11.0528	0.0009
## tempo3	937.1624	93.2124	101.0840	0.0000
## tempo4	1588.1250	146.3053	117.8282	0.0000
## tempo5	1647.9664	159.7611	106.4032	0.0000
## tempo6	1039.5211	118.6682	76.7358	0.0000
## grupo1	-183.6637	151.8076	1.4637	0.2263
## grupo2	207.2921	243.4156	0.7252	0.3944
## sexoM	209.5120	220.5135	0.9027	0.3421
## imc	-12.0736	25.8231	0.2186	0.6401
## nyha2	212.4329	220.5006	0.9282	0.3353
## nyha3	16.8317	227.9247	0.0055	0.9411
## nyha4	7.9476	258.7087	0.0009	0.9755
## idade	11.2772	8.6664	1.6933	0.1932

```
## euros          16.3856  74.8018  0.0480  0.8266
## tempo2:grupo1  54.2587 131.0613  0.1714  0.6789
## tempo3:grupo1 -131.5640 133.2377  0.9750  0.3234
## tempo4:grupo1 -404.0468 190.3561  4.5054  0.0338
## tempo5:grupo1 -392.1963 226.3615  3.0019  0.0832
## tempo6:grupo1 -116.6034 166.6298  0.4897  0.4841
## tempo2:grupo2   -5.3615 124.3806  0.0019  0.9656
## tempo3:grupo2   13.4085 153.8346  0.0076  0.9305
## tempo4:grupo2 -139.8685 215.9243  0.4196  0.5171
## tempo5:grupo2 -411.2185 236.3418  3.0274  0.0819
## tempo6:grupo2 -324.6783 174.3459  3.4680  0.0626
```

Com base nos resultados do modelo, podemos observar que, embora a interação entre o **Tempo 4** e o **Grupo 2** seja significativa, os efeitos nos demais níveis de tempo e grupo não parecem ser significativos. Isso pode indicar a ausência de um efeito conjunto entre **tempo** e **grupo** no modelo analisado.

Como segunda hipótese, decidimos testar a interação entre as variáveis **Sexo** e **Grupo** para avaliar se há relevância dessa relação.

EQUAÇÃO3

```
# interação entre sexo e grupo
ajuste_gee_ar1_sexgrup <- geeglm(citocina_t ~ tempo*grupo*sexo+ imc + nyha + idade + euros,
                                data = dados_longos, corstr = 'ar1', id = id, family = 'gaussian')

summary(ajuste_gee_ar1_sexgrup)

##
## Call:
## geeglm(formula = citocina_t ~ tempo + grupo * sexo + imc + nyha +
##       idade + euros, family = "gaussian", data = dados_longos,
##       id = id, corstr = "ar1")
##
## Coefficients:
##              Estimate Std. err    Wald Pr(>|W|)
## (Intercept)  -103.852   790.268    0.017   0.8954
## tempo2         352.444    50.909   47.929 4.42e-12 ***
## tempo3         896.709    60.078  222.779 < 2e-16 ***
## tempo4        1408.408    85.044  274.266 < 2e-16 ***
## tempo5        1389.107    98.021  200.834 < 2e-16 ***
## tempo6         900.528    71.824  157.199 < 2e-16 ***
## grupo1         544.532   418.351    1.694   0.1930
## grupo2         496.622   308.529    2.591   0.1075
## sexoM          729.238   316.117    5.322   0.0211 *
## imc            -9.351    26.500    0.125   0.7242
## nyha2          238.572   221.781    1.157   0.2821
## nyha3          136.243   205.378    0.440   0.5071
## nyha4           5.831   283.721    0.000   0.9836
## idade           9.882     8.226    1.443   0.2296
## euros          32.264    72.981    0.195   0.6584
## grupo1:sexoM  -995.282   471.336    4.459   0.0347 *
## grupo2:sexoM -717.545   361.164    3.947   0.0469 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation structure = ar1
## Estimated Scale Parameters:
##
##           Estimate Std.err
## (Intercept)  503877   63973
##   Link = identity
##
## Estimated Correlation Parameters:
##           Estimate Std.err
## alpha    0.7899 0.03438
## Number of clusters:   62   Maximum cluster size: 6
```

Diferentemente do ajuste anterior, a interação entre os diferentes níveis de **Sexo** e **Grupo**, são significativas. Indicando possivelmente indicando um efeito combinado dessas duas variáveis na variável resposta.

Modelos Mistos

Uma abordagem alternativa para essa análise seria a utilização de **modelos mistos**, que permitem a inclusão de interceptos e/ou inclinações aleatórias para capturar a variabilidade intra e interpacientes

Ajuste 1: Intercepto Aleatório

Como primeiro passo, ajustaremos um modelo misto com **intercepto aleatório** para todas as variáveis inicialmente selecionadas. Em seguida, realizaremos uma seleção de variáveis para os ajustes posteriores

Podemos observar que os efeitos de grupo e das variáveis A, B e C não parecem ser significativos para o ajuste. Porém, da mesma forma que o modelos GEE, mantivemos as variáveis que são clinicamente significativas para o modelo do estudo.

```
## Intercepto aleatório

ajuste_misto_1 <- lme(citocina_t ~ tempo + ai + sexo + nyha + grupo + idade + has + euroes + imc + iap,
                     data = dados_longos, random = ~1|id)

round(coef(summary(ajuste_misto_1)),4)
```

```
##           Value Std.Error   DF t-value p-value
## (Intercept) -402.140    991.59  305  -0.4055  0.6854
## tempo2      352.444     83.07  305   4.2427  0.0000
## tempo3      896.709     83.07  305  10.7945  0.0000
## tempo4     1408.408     83.07  305  16.9542  0.0000
## tempo5     1389.107     83.07  305  16.7219  0.0000
## tempo6      900.528     83.07  305  10.8405  0.0000
## ai1         406.664    383.70   49   1.0598  0.2944
## sexoM       127.000    243.59   49   0.5214  0.6045
## nyha2       212.307    229.15   49   0.9265  0.3587
## nyha3        49.942    299.37   49   0.1668  0.8682
## nyha4       122.510    401.17   49   0.3054  0.7614
## grupo1     -291.072    227.15   49  -1.2814  0.2061
## grupo2      459.941    377.69   49   1.2178  0.2291
## idade        16.913     10.47   49   1.6155  0.1126
## has1        294.079    331.35   49   0.8875  0.3791
## euroes      -39.966     74.99   49  -0.5329  0.5965
## imc          -1.596     25.70   49  -0.0621  0.9508
## iap1       -105.313    397.41   49  -0.2650  0.7921
```

Ajuste 2: Intercepto Aleatório e Variáveis Clínicas

Como alternativa ao modelo anterior, decidimos ajustar um modelo apenas com as variáveis clinicamente significativas, mas mantendo o efeito aleatório no intercepto. Abaixo temos um resumo do modelo:

```
ajuste_misto_2<- lme(citocina_t ~ tempo+grupo +sexo+ imc + nyha + idade + euroes,  
                     data = dados_longos, random = ~1|id)  
  
round(coef(summary(ajuste_misto_2)),4)
```

##		Value	Std.Error	DF	t-value	p-value
##	(Intercept)	282.28	873.936	305	0.3230	0.7469
##	tempo2	352.44	83.071	305	4.2427	0.0000
##	tempo3	896.71	83.071	305	10.7945	0.0000
##	tempo4	1408.41	83.071	305	16.9542	0.0000
##	tempo5	1389.11	83.071	305	16.7219	0.0000
##	tempo6	900.53	83.071	305	10.8405	0.0000
##	grupo1	-362.37	209.933	52	-1.7261	0.0903
##	grupo2	127.44	287.701	52	0.4430	0.6596
##	sexoM	83.48	231.067	52	0.3613	0.7194
##	imc	-10.74	24.056	52	-0.4466	0.6570
##	nyha2	215.95	225.720	52	0.9567	0.3431
##	nyha3	42.04	289.012	52	0.1455	0.8849
##	nyha4	45.85	394.550	52	0.1162	0.9079
##	idade	15.26	9.166	52	1.6650	0.1019
##	euroes	-23.54	64.444	52	-0.3653	0.7164

Ajuste 3: Intercepto Aleatório e Interação

Um segundo ajuste a ser testado, seria o modelo misto com intercepto aleatório e a interação entre entre **Sexo** e **Grupo**.

Abaixo temos um resumo do novo modelo ajustado

```
## Intercepto aleatório: Interação sexo grupo  
ajuste_misto_3<- lme(citocina_t ~ tempo+grupo*sexo - grupo+ imc + nyha + idade + euroes,  
                     data = dados_longos, random = ~1|id)  
  
round(coef(summary(ajuste_misto_3)),4)
```

##		Value	Std.Error	DF	t-value	p-value
##	(Intercept)	-192.995	900.037	305	-0.2144	0.8304
##	tempo2	352.444	83.071	305	4.2427	0.0000
##	tempo3	896.709	83.071	305	10.7945	0.0000
##	tempo4	1408.408	83.071	305	16.9542	0.0000
##	tempo5	1389.107	83.071	305	16.7219	0.0000
##	tempo6	900.528	83.071	305	10.8405	0.0000
##	sexoM	645.748	361.572	50	1.7859	0.0802
##	imc	-7.851	23.644	50	-0.3320	0.7413
##	nyha2	244.944	222.811	50	1.0993	0.2769
##	nyha3	168.333	290.961	50	0.5785	0.5655
##	nyha4	47.787	400.239	50	0.1194	0.9054
##	idade	13.542	9.018	50	1.5016	0.1395
##	euroes	-3.623	64.106	50	-0.0565	0.9552
##	grupo1:sexoF	499.392	552.086	50	0.9046	0.3700
##	grupo2:sexoF	600.060	386.387	50	1.5530	0.1267
##	grupo1:sexoM	-530.510	221.768	50	-2.3922	0.0205

```
## grupo2:sexoM -173.961 340.450 50 -0.5110 0.6116
```

Olhando para o resumo do modelo, não parece haver um efeito claro do termo de interação para esse modelo.

Ajuste 4: Efeito Aleatório no Grupo e Interação

Como outra alternativa, iremos ajustar um modelo com o efeito aleatório na variável de grupo, afim de verificar se produz um ajuste melhor em comparação ao anterior. Novamente, decidimos optar pelo termo de interação entre **Sexo** e **Grupo**. Abaixo temos um resumo do modelo ajustado:

```
ajuste_misto_4 <- lme(citocina_t ~ tempo+grupo * sexo - grupo+ imc + nyha + idade + euros,
                      data = dados_longos, random = ~1|grupo)

round(coef(summary(ajuste_misto_4)),4)
```

##	Value	Std.Error	DF	t-value	p-value
## (Intercept)	-192.995	460.732	353	-0.4189	0.6756
## tempo2	352.444	129.931	353	2.7125	0.0070
## tempo3	896.709	129.931	353	6.9014	0.0000
## tempo4	1408.408	129.931	353	10.8396	0.0000
## tempo5	1389.107	129.931	353	10.6911	0.0000
## tempo6	900.528	129.931	353	6.9308	0.0000
## sexoM	645.748	168.464	353	3.8331	0.0001
## imc	-7.851	11.016	353	-0.7126	0.4765
## nyha2	244.944	103.812	353	2.3595	0.0188
## nyha3	168.333	135.565	353	1.2417	0.2152
## nyha4	47.787	186.480	353	0.2563	0.7979
## idade	13.542	4.202	353	3.2229	0.0014
## euros	-3.623	29.868	353	-0.1213	0.9035
## grupo1:sexoF	499.392	355.235	353	1.4058	0.1607
## grupo2:sexoF	600.060	304.031	353	1.9737	0.0492
## grupo1:sexoM	-530.510	265.898	353	-1.9952	0.0468
## grupo2:sexoM	-173.961	291.868	353	-0.5960	0.5515

Não parece haver uma diferença muito expressiva em relação ao modelo anterior ao observarmos os termos de interação.

Comparação dos Ajustes

Ao todo, ajustamos quatro modelos mistos, descritos a seguir:

1. **Ajuste 1:** Inclui todas as variáveis e um **intercepto aleatório**.
2. **Ajuste 2:** Inclui as variáveis clinicamente relevantes e um **intercepto aleatório**.
3. **Ajuste 3:** Inclui **variáveis selecionadas**, um **intercepto aleatório** e um **termo de interação**.
4. **Ajuste 4:** Inclui um **efeito aleatório** na variável de **grupo** e um **termo de interação**.

Para comparar esses ajustes, utilizaremos como critérios as medidas de **AIC** e **BIC** dos três modelos.

```
knitr::kable(
  data.frame(
    Modelos = c('Ajuste 1', 'Ajuste 2', 'Ajuste 3', 'Ajuste 4'),
    AIC = c(AIC(ajuste_misto_1), AIC(ajuste_misto_2), AIC(ajuste_misto_3), AIC(ajuste_misto_4)),
    BIC = c(BIC(ajuste_misto_1), BIC(ajuste_misto_2), BIC(ajuste_misto_3), BIC(ajuste_misto_4))
  ))
```

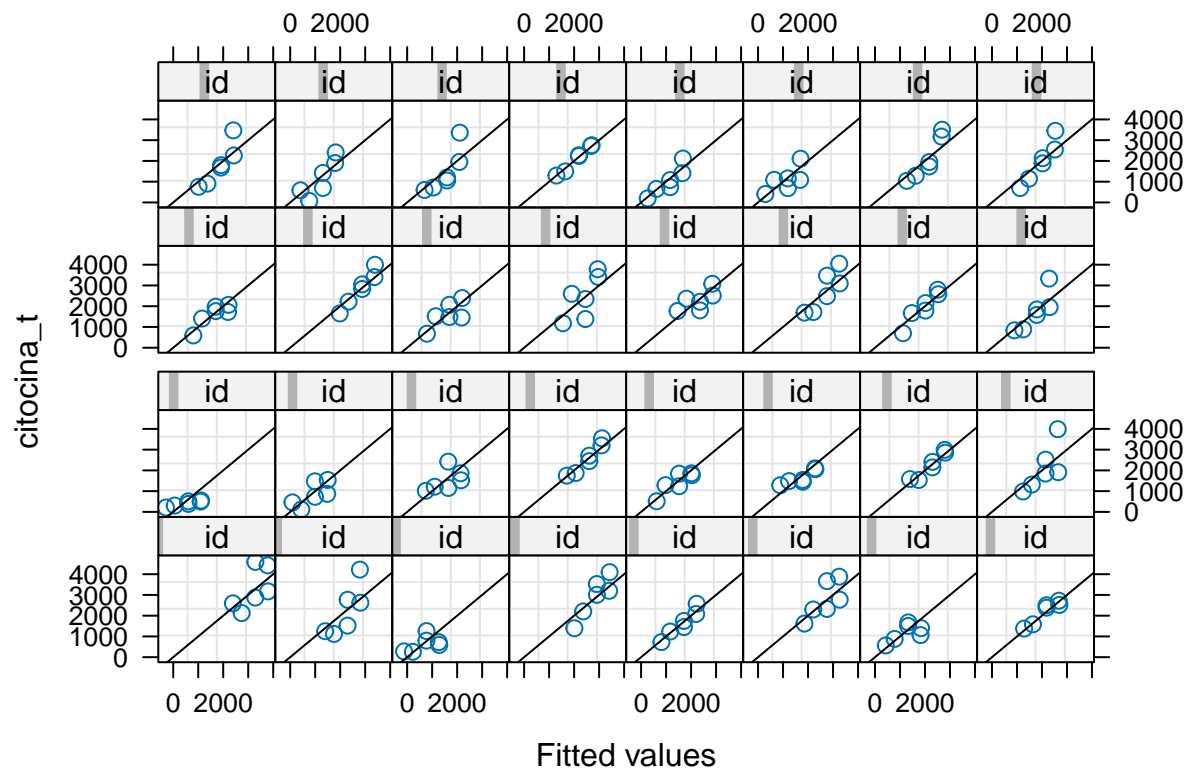
Modelos	AIC	BIC
Ajuste 1	5594	5671
Ajuste 2	5631	5697
Ajuste 3	5602	5676
Ajuste 4	5799	5872

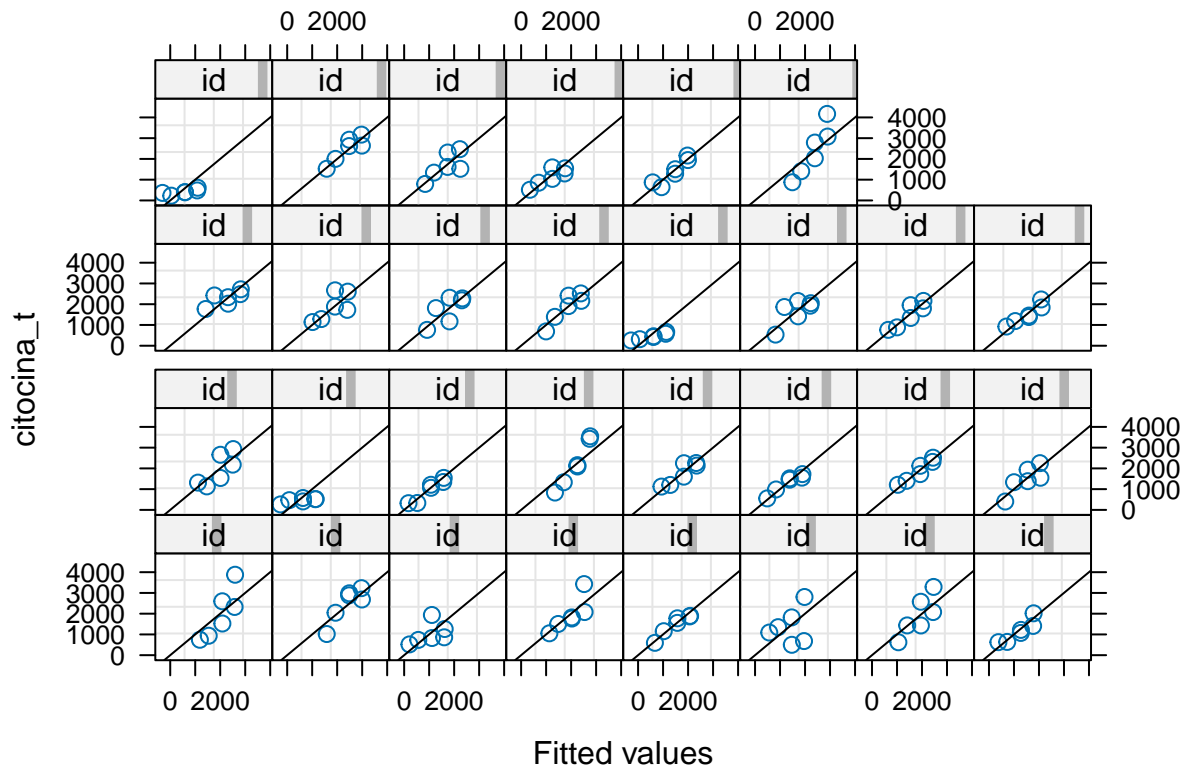
Podemos observar que o modelo que produz o menor AIC e BIC, parece ser o modelo que inclui todas as variáveis e um **intercepto aleatório** (Ajuste 1).

Análise de Resíduos

Com base no modelo misto selecionado, **Ajuste**, trazemos abaixo um gráfico de resíduos das medições da citocina Tnf-receptor ajustadas pelo modelo, nos separados por pacientes.

```
plot(ajuste_misto_1, citocina_t ~ fitted(.) | id, layout= c(8,4), between= list(y= c(0, 0.5, 0)),
     abline= c(0,1))
```





Com base nos gráficos acima, podemos observar que o modelo aparenta apresentar um bom ajuste para os indivíduos. Porém, podemos ver que para alguns pacientes, os pontos se afastam bastante da reta identidade, indicando que, para esse pacientes, o modelo não se ajusta tão bem assim.

Outra análise gráfica que podemos realizar, são os gráficos de resíduos padronizados por tempo, que são exibidos abaixo:

```
plot(ajuste_misto_1, residuals(., type= "p") ~ fitted(.) | tempo)
```