

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)
```

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	4592.32
2	M	74	25.6	70	1	1	0	0	0	3	1.1	1251.05	1120.12	1520.17	2625.28	4221.14	2767.50
3	M	20	19.7	90	4	0	0	0	2	5	1.1	279.25	258.43	6796.95	5590.95	723.51	1258.72
4	F	68	21.0	70	3	1	0	0	2	6	1.1	1394.00	2205.19	3524.08	74095.74	3198.84	3006.30
5	F	51	23.9	70	2	1	0	0	0	2	0.9	728.07	1235.74	1747.37	2573.23	2082.71	2452.58

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[,13:18]),2)
```

```
##      t1    t2    t3    t4    t5    t6
## t1 1.00 0.74 0.67 0.61 0.71 0.79
## t2 0.74 1.00 0.69 0.62 0.67 0.74
## t3 0.67 0.69 1.00 0.72 0.59 0.66
## t4 0.61 0.62 0.72 1.00 0.67 0.60
## t5 0.71 0.67 0.59 0.67 1.00 0.85
## t6 0.79 0.74 0.66 0.60 0.85 1.00
```

Correlação no Grupo 0

```
## Correlação do grupo 0
round(cor(subset(dados, grupo == 0)[,13:18]),2)
```

```
##      t1  t2  t3  t4  t5  t6
## t1 1.00 0.71 0.71 0.51 0.78 0.90
## t2 0.71 1.00 0.62 0.49 0.72 0.81
## t3 0.71 0.62 1.00 0.56 0.55 0.72
## t4 0.51 0.49 0.56 1.00 0.56 0.45
## t5 0.78 0.72 0.55 0.56 1.00 0.85
## t6 0.90 0.81 0.72 0.45 0.85 1.00
```

Correlação no Grupo 1

```
## Correlação do grupo 1
round(cor(subset(dados, grupo == 1)[,13:18]),2)
```

```
##      t1  t2  t3  t4  t5  t6
## t1 1.00 0.74 0.67 0.70 0.60 0.73
## t2 0.74 1.00 0.80 0.70 0.67 0.74
## t3 0.67 0.80 1.00 0.82 0.78 0.74
## t4 0.70 0.70 0.82 1.00 0.86 0.82
## t5 0.60 0.67 0.78 0.86 1.00 0.84
## t6 0.73 0.74 0.74 0.82 0.84 1.00
```

Correlação no Grupo 2

```
## Correlação do grupo
round(cor(subset(dados, grupo == 2)[,13:18]),2)
```

```
##      t1  t2  t3  t4  t5  t6
## t1 1.00 0.80 0.58 0.63 0.73 0.72
## t2 0.80 1.00 0.67 0.74 0.63 0.66
## t3 0.58 0.67 1.00 0.73 0.49 0.57
## t4 0.63 0.74 0.73 1.00 0.59 0.62
## t5 0.73 0.63 0.49 0.59 1.00 0.87
## t6 0.72 0.66 0.57 0.62 0.87 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

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	5	4425.335
1	M	64	20.2	70	2	1	0	1	0	6	1.2	6	4592.328

Gráficos

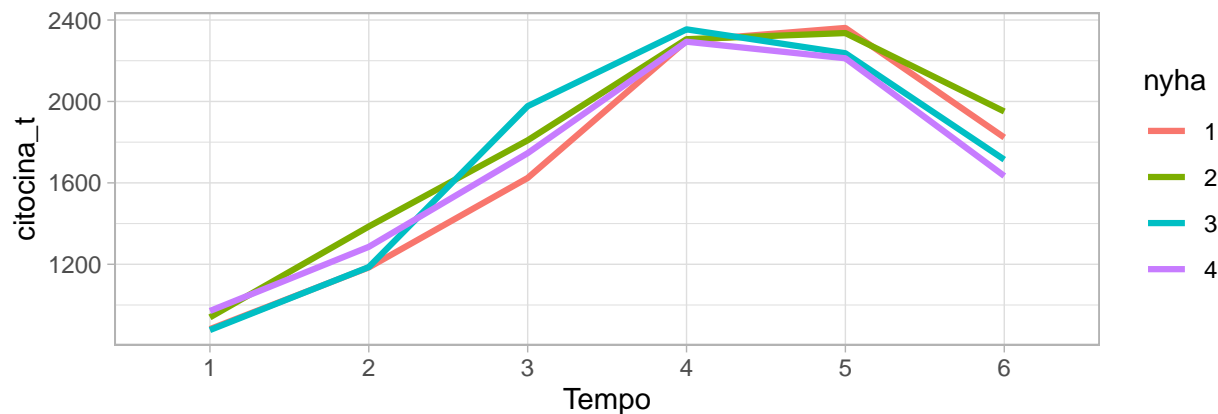
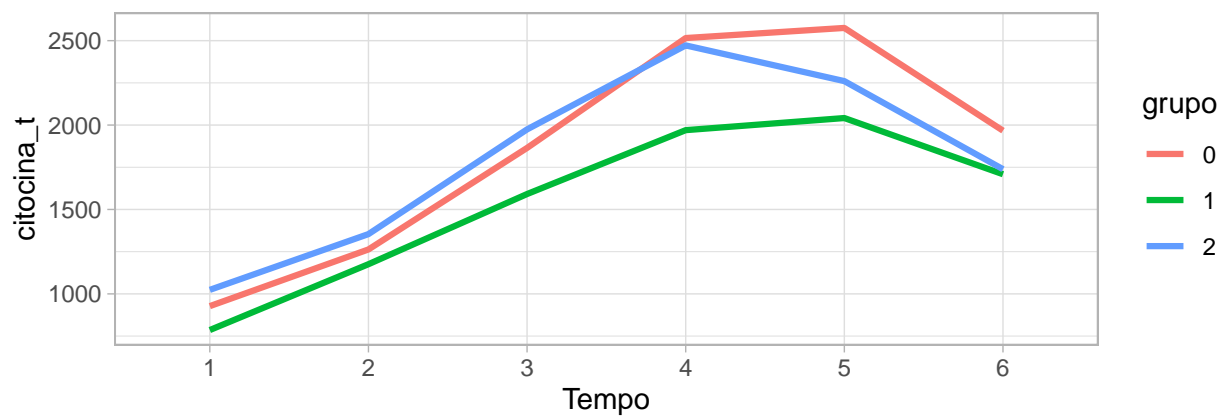
```
# Gráficos
```

```
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)
```



```

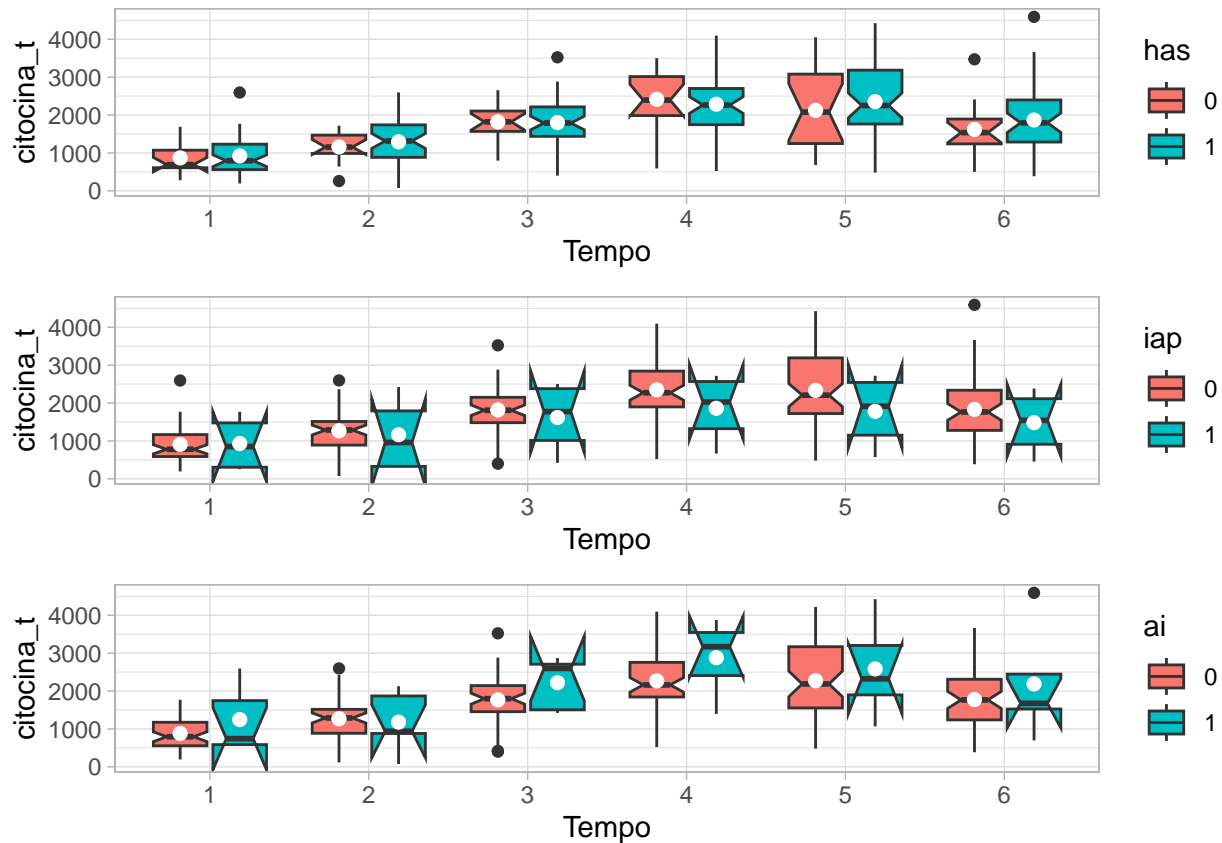
grafico3 <-
  ggplot(dados_longos, aes(x = tempo, y = citocina_t, fill = has)) +
  geom_boxplot(notch = TRUE) + theme(legend.position = "top") +
  stat_summary(
    fun = "mean",
    geom = "point",
    size = 2,
    color = "white",
    position = position_dodge(width = 0.75),
    show.legend = FALSE
  ) +
  labs(x = "Tempo") + theme_light() +
  scale_x_discrete(breaks = seq_along(rotulos), labels = rotulos)+theme(plot.title = element_text(hjust

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(hjust

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(hjust

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

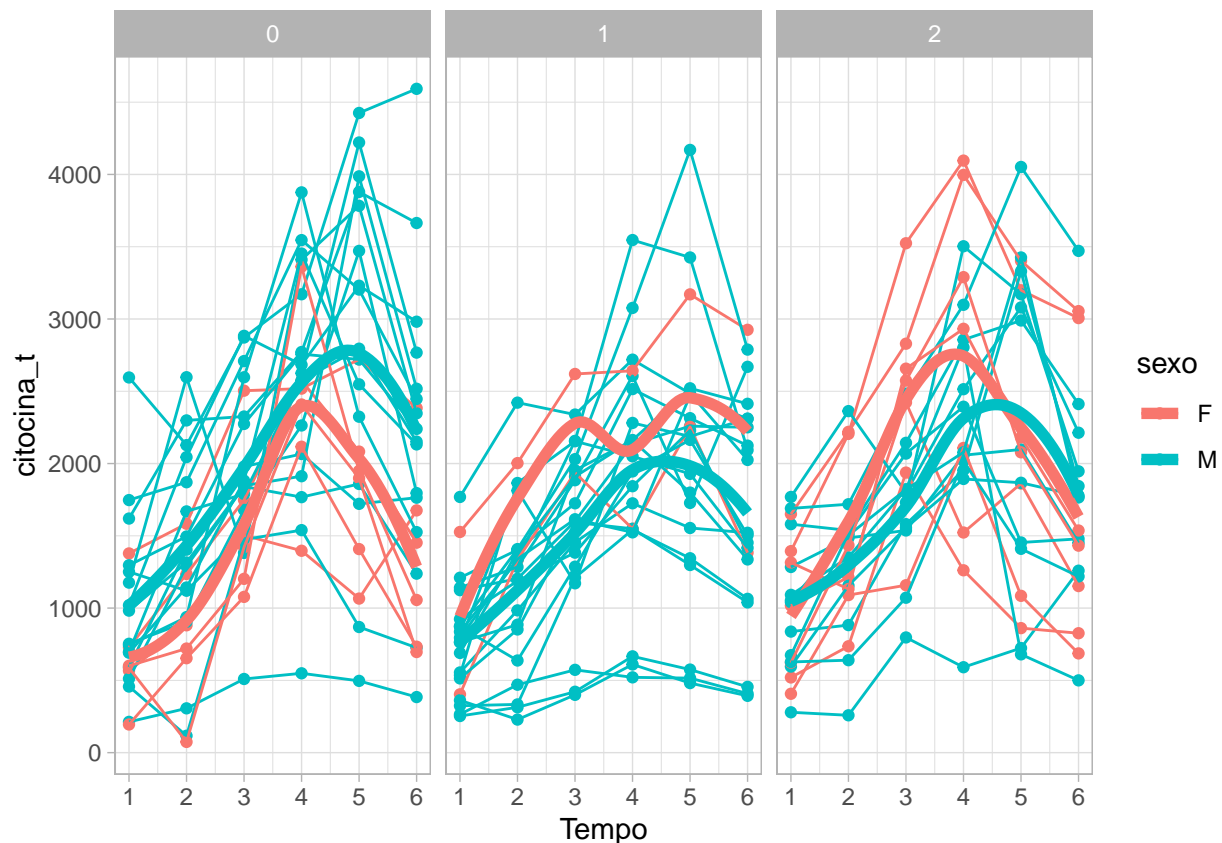
```



```
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. As variáveis **fc** (Frequência Cardíaca) e **creat** (Creatina) não foram consideradas nas análises.

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)), 4)
```

	Estimate	Std.err	Wald	Pr(> W)
## (Intercept)	-402.1399	893.6694	0.2025	0.6527
## tempo2	352.4436	50.9087	47.9287	0.0000
## tempo3	896.7094	60.0779	222.7792	0.0000
## tempo4	1408.4075	85.0438	274.2658	0.0000
## tempo5	1389.1071	98.0206	200.8340	0.0000
## tempo6	900.5282	71.8244	157.1991	0.0000
## ai1	406.6638	370.7630	1.2030	0.2727

```
## sexoM      126.9997 192.8014  0.4339  0.5101
## nyha2      212.3069 216.7565  0.9594  0.3273
## nyha3       49.9421 240.3601  0.0432  0.8354
## nyha4      122.5104 266.2635  0.2117  0.6454
## grupo1     -291.0723 206.2247  1.9921  0.1581
## grupo2      459.9408 329.6150  1.9471  0.1629
## idade       16.9134   8.0877  4.3734  0.0365
## has1        294.0793 231.3930  1.6152  0.2038
## euroes     -39.9661  68.3612  0.3418  0.5588
## imc         -1.5956  27.7531  0.0033  0.9542
## iap1       -105.3127 467.4939  0.0507  0.8218
```

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)),4)
```

```
##           Estimate Std.err    Wald Pr(>|W|)
## (Intercept) -402.1399 893.6694   0.2025  0.6527
## tempo2       352.4436  50.9087  47.9287  0.0000
## tempo3       896.7094  60.0779 222.7792  0.0000
## tempo4      1408.4075  85.0438 274.2658  0.0000
## tempo5      1389.1071  98.0206 200.8340  0.0000
## tempo6       900.5282  71.8244 157.1991  0.0000
## ai1          406.6638 370.7630   1.2030  0.2727
## sexoM        126.9997 192.8014   0.4339  0.5101
## nyha2        212.3069 216.7565   0.9594  0.3273
## nyha3         49.9421 240.3601   0.0432  0.8354
## nyha4        122.5104 266.2635   0.2117  0.6454
## grupo1       -291.0723 206.2247   1.9921  0.1581
## grupo2        459.9408 329.6150   1.9471  0.1629
## idade         16.9134   8.0877   4.3734  0.0365
## has1          294.0793 231.3930   1.6152  0.2038
## euroes       -39.9661  68.3612   0.3418  0.5588
## imc           -1.5956  27.7531   0.0033  0.9542
## iap1        -105.3127 467.4939   0.0507  0.8218
```

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)), 4)
```

```
##           Estimate Std.terr      Wald Pr(>|W|)
## (Intercept) -583.9262 699.7547   0.6963 0.4040
## tempo2      352.4436  50.9087  47.9287 0.0000
## tempo3      896.7094  60.0779 222.7792 0.0000
## tempo4     1408.4075  85.0438 274.2658 0.0000
## tempo5     1389.1071  98.0206 200.8340 0.0000
## tempo6      900.5282  71.8244 157.1991 0.0000
## ai1         619.8332 302.8043   4.1901 0.0407
## sexoM       -277.1634 133.4510   4.3135 0.0378
## nyha2        105.9218 158.3058   0.4477 0.5034
## nyha3       -179.6487 198.0836   0.8225 0.3644
## nyha4        -46.9516 226.0813   0.0431 0.8355
## grupo1     -141.1616 136.6009   1.0679 0.3014
## grupo2      907.3259 258.2570  12.3430 0.0004
## idade        16.4307  7.4037   4.9251 0.0265
## has1         220.7575 220.2736   1.0044 0.3162
## euros       -82.6469  50.0757   2.7240 0.0989
## imc          23.4382 19.1225   1.5023 0.2203
## iap1         164.8051 360.2800   0.2092 0.6474
```

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 **Não Estruturada**. Essa escolha se baseia na análise das correlações das medidas da citocina TNF-receptor nos diferentes tempos realizada anteriormente, onde não apresentam um padrão claro.

Seleção de Variáveis

Com base no resultado da função `summary()` aplicada ao modelo GEE com estrutura de correlação Não Estruturada (exibido novamente abaixo), observamos que diversas variáveis aparentam não serem significativas a um nível de significância de 5%.

No entanto, as variáveis idade, imc, nyha e euroscore 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_unstructured)),4)
```

```
##           Estimate Std.err      Wald Pr(>|W|)
## (Intercept) -583.9262 699.7547    0.6963  0.4040
## tempo2      352.4436  50.9087   47.9287  0.0000
## tempo3      896.7094  60.0779  222.7792  0.0000
## tempo4     1408.4075  85.0438  274.2658  0.0000
## tempo5     1389.1071  98.0206  200.8340  0.0000
## tempo6      900.5282  71.8244  157.1991  0.0000
## ai1         619.8332 302.8043   4.1901  0.0407
## sexoM      -277.1634 133.4510   4.3135  0.0378
## nyha2       105.9218 158.3058   0.4477  0.5034
## nyha3      -179.6487 198.0836   0.8225  0.3644
## nyha4       -46.9516 226.0813   0.0431  0.8355
## grupo1     -141.1616 136.6009   1.0679  0.3014
## grupo2      907.3259 258.2570  12.3430  0.0004
## idade       16.4307   7.4037   4.9251  0.0265
## has1        220.7575 220.2736   1.0044  0.3162
## euroes     -82.6469  50.0757   2.7240  0.0989
## imc         23.4382  19.1225   1.5023  0.2203
## iap1        164.8051 360.2800   0.2092  0.6474
```

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 um resumo do modelo através da função `summary()`.

```
ajuste_gee_grupotempo <- geeglm(
  citocina_t ~ tempo * grupo - grupo + sexo + imc + nyha + idade + euroes,
  data = dados_longos,
  corstr = 'unstructured',
  id = id,
  family = 'gaussian'
)

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

```
##           Estimate Std.err      Wald Pr(>|W|)
## (Intercept)   77.6670 712.5013    0.0119  0.9132
## 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
## sexoM       -224.3006 146.2259   2.3529  0.1250
## imc          14.4154  17.6039   0.6706  0.4129
## nyha2        144.2081 177.8573   0.6574  0.4175
## nyha3       -108.4636 187.4656   0.3348  0.5629
## nyha4        -86.9468 235.8547   0.1359  0.7124
## idade        11.5559   8.0141   2.0792  0.1493
## euroes       -30.1218  56.5946   0.2833  0.5946
## tempo1:grupo1 -130.1605 149.4906   0.7581  0.3839
```

```
## tempo2:grupo1 -75.9019 188.7629 0.1617 0.6876
## tempo3:grupo1 -261.7246 183.7565 2.0286 0.1544
## tempo4:grupo1 -534.2073 233.6841 5.2259 0.0223
## tempo5:grupo1 -522.3568 302.0290 2.9911 0.0837
## tempo6:grupo1 -246.7639 269.7284 0.8370 0.3603
## tempo1:grupo2 369.8312 234.6421 2.4843 0.1150
## tempo2:grupo2 364.4698 234.7681 2.4102 0.1206
## tempo3:grupo2 383.2397 253.2686 2.2897 0.1302
## tempo4:grupo2 229.9628 295.7943 0.6044 0.4369
## tempo5:grupo2 -41.3873 361.5219 0.0131 0.9089
## tempo6:grupo2 45.1529 334.7461 0.0182 0.8927
```

Com base nos resultados do modelo, podemos observar que, embora a interação entre o **Tempo 4** e o **Grupo 1** 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.

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

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

```
##           Estimate Std.err      Wald Pr(>|W|)
## (Intercept) -316.2064 683.8620   0.2138  0.6438
## tempo2      352.4436  50.9087  47.9287  0.0000
## tempo3      896.7094  60.0779 222.7792  0.0000
## tempo4     1408.4075  85.0438 274.2658  0.0000
## tempo5     1389.1071  98.0206 200.8340  0.0000
## tempo6      900.5282  71.8244 157.1991  0.0000
## sexoM       255.2614 214.1788   1.4204  0.2333
## imc         15.0799  18.7477   0.6470  0.4212
## nyha2       185.7508 176.3024   1.1101  0.2921
## nyha3        16.4539 174.2651   0.0089  0.9248
## nyha4       -29.4728 246.7537   0.0143  0.9049
## idade       10.4601   7.5537   1.9175  0.1661
## euroes       -8.8747  55.1243   0.0259  0.8721
## grupo1:sexoF 291.0020 211.1219   1.8999  0.1681
## grupo2:sexoF 802.3328 227.8074  12.4043  0.0004
## grupo1:sexoM -330.7921 164.3475   4.0512  0.0441
## grupo2:sexoM 137.1992 227.3267   0.3643  0.5462
```

Diferentemente do ajuste anterior, a interação entre **Sexo** e **Grupo** possui interações significativas. Isso é um indicativo que existe um efeito conjunto plausível da combinação dessas duas variáveis.

Com base nos ajustes marginais apresentados acima, podemos observar efeitos mais relevantes na reposta da citocina Tnf-receptor estão associados as medições (**Tempo**), Sexo Masculino, Idade e tanto as interações entre **Sexo** e **Grupo**, bem como o indivíduo pertencer ao Grupo 2 apenas.

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)	-402.1399	991.5921	305	-0.4055	0.6854
##	tempo2	352.4436	83.0711	305	4.2427	0.0000
##	tempo3	896.7094	83.0711	305	10.7945	0.0000
##	tempo4	1408.4075	83.0711	305	16.9542	0.0000
##	tempo5	1389.1071	83.0711	305	16.7219	0.0000
##	tempo6	900.5282	83.0711	305	10.8405	0.0000
##	ai1	406.6638	383.7033	49	1.0598	0.2944
##	sexoM	126.9997	243.5907	49	0.5214	0.6045
##	nyha2	212.3069	229.1455	49	0.9265	0.3587
##	nyha3	49.9421	299.3697	49	0.1668	0.8682
##	nyha4	122.5104	401.1651	49	0.3054	0.7614
##	grupo1	-291.0723	227.1548	49	-1.2814	0.2061
##	grupo2	459.9408	377.6934	49	1.2178	0.2291
##	idade	16.9134	10.4692	49	1.6155	0.1126
##	has1	294.0793	331.3537	49	0.8875	0.3791
##	euroes	-39.9661	74.9949	49	-0.5329	0.5965
##	imc	-1.5956	25.7046	49	-0.0621	0.9508
##	iap1	-105.3127	397.4076	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)	282.2829	873.9357	305	0.3230	0.7469

```
## tempo2      352.4436   83.0711 305  4.2427  0.0000
## tempo3      896.7094   83.0711 305 10.7945  0.0000
## tempo4     1408.4075   83.0711 305 16.9542  0.0000
## tempo5     1389.1071   83.0711 305 16.7219  0.0000
## tempo6       900.5282   83.0711 305 10.8405  0.0000
## grupo1     -362.3681  209.9330  52 -1.7261  0.0903
## grupo2      127.4424  287.7006  52  0.4430  0.6596
## sexoM        83.4764  231.0666  52  0.3613  0.7194
## imc         -10.7425   24.0561  52 -0.4466  0.6570
## nyha2       215.9451  225.7204  52  0.9567  0.3431
## nyha3        42.0372  289.0116  52  0.1455  0.8849
## nyha4       45.8499  394.5498  52  0.1162  0.9079
## idade       15.2619    9.1665  52  1.6650  0.1019
## euroes     -23.5416   64.4443  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 **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) -192.9952   900.0374 305 -0.2144  0.8304
## tempo2       352.4436    83.0711 305  4.2427  0.0000
## tempo3       896.7094    83.0711 305 10.7945  0.0000
## tempo4     1408.4075    83.0711 305 16.9542  0.0000
## tempo5     1389.1071    83.0711 305 16.7219  0.0000
## tempo6       900.5282    83.0711 305 10.8405  0.0000
## sexoM        645.7480   361.5724  50  1.7859  0.0802
## imc          -7.8506    23.6442  50 -0.3320  0.7413
## nyha2       244.9438    222.8106  50  1.0993  0.2769
## nyha3       168.3328    290.9609  50  0.5785  0.5655
## nyha4        47.7874    400.2388  50  0.1194  0.9054
## idade       13.5420     9.0182  50  1.5016  0.1395
## euroes      -3.6227    64.1059  50 -0.0565  0.9552
## grupo1:sexoF  499.3925   552.0860  50  0.9046  0.3700
## grupo2:sexoF  600.0604   386.3874  50  1.5530  0.1267
## grupo1:sexoM -530.5100   221.7678  50 -2.3922  0.0205
## grupo2:sexoM -173.9608   340.4501  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 + euroes,
                     data = dados_longos, random = ~1|grupo)
```

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

```
##              Value Std.Error   DF t-value p-value
## (Intercept) -192.9952  460.7324  353 -0.4189  0.6756
## tempo2      352.4436  129.9313  353  2.7125  0.0070
## tempo3      896.7094  129.9313  353  6.9014  0.0000
## tempo4     1408.4075  129.9313  353 10.8396  0.0000
## tempo5     1389.1071  129.9313  353 10.6911  0.0000
## tempo6      900.5282  129.9313  353  6.9308  0.0000
## sexoM       645.7480  168.4643  353  3.8331  0.0001
## imc         -7.8506   11.0163  353 -0.7126  0.4765
## nyha2       244.9438  103.8122  353  2.3595  0.0188
## nyha3       168.3328  135.5649  353  1.2417  0.2152
## nyha4        47.7874  186.4798  353  0.2563  0.7979
## idade       13.5420    4.2018  353  3.2229  0.0014
## euroes      -3.6227   29.8683  353 -0.1213  0.9035
## grupo1:sexoF 499.3925  355.2351  353  1.4058  0.1607
## grupo2:sexoF 600.0604  304.0309  353  1.9737  0.0492
## grupo1:sexoM -530.5100  265.8979  353 -1.9952  0.0468
## grupo2:sexoM -173.9608  291.8675  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))
))
```

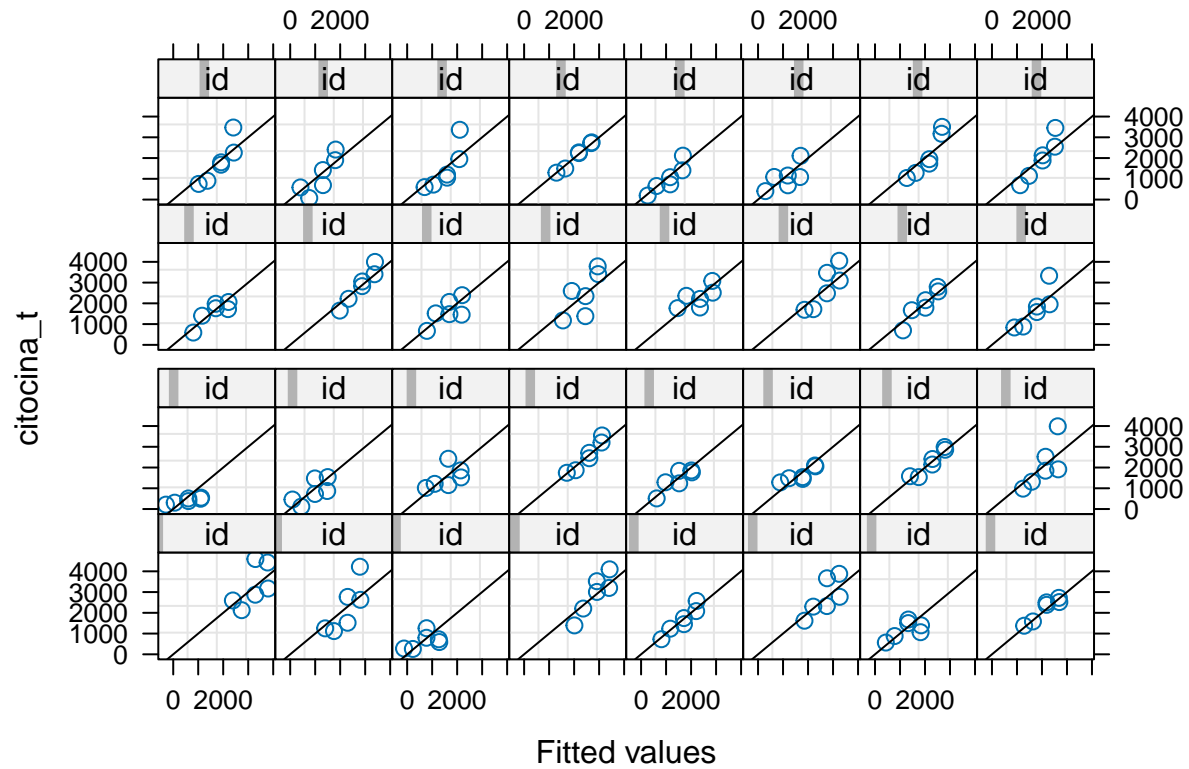
Modelos	AIC	BIC
Ajuste 1	5593.911	5671.297
Ajuste 2	5631.217	5697.138
Ajuste 3	5602.445	5676.015
Ajuste 4	5798.929	5872.500

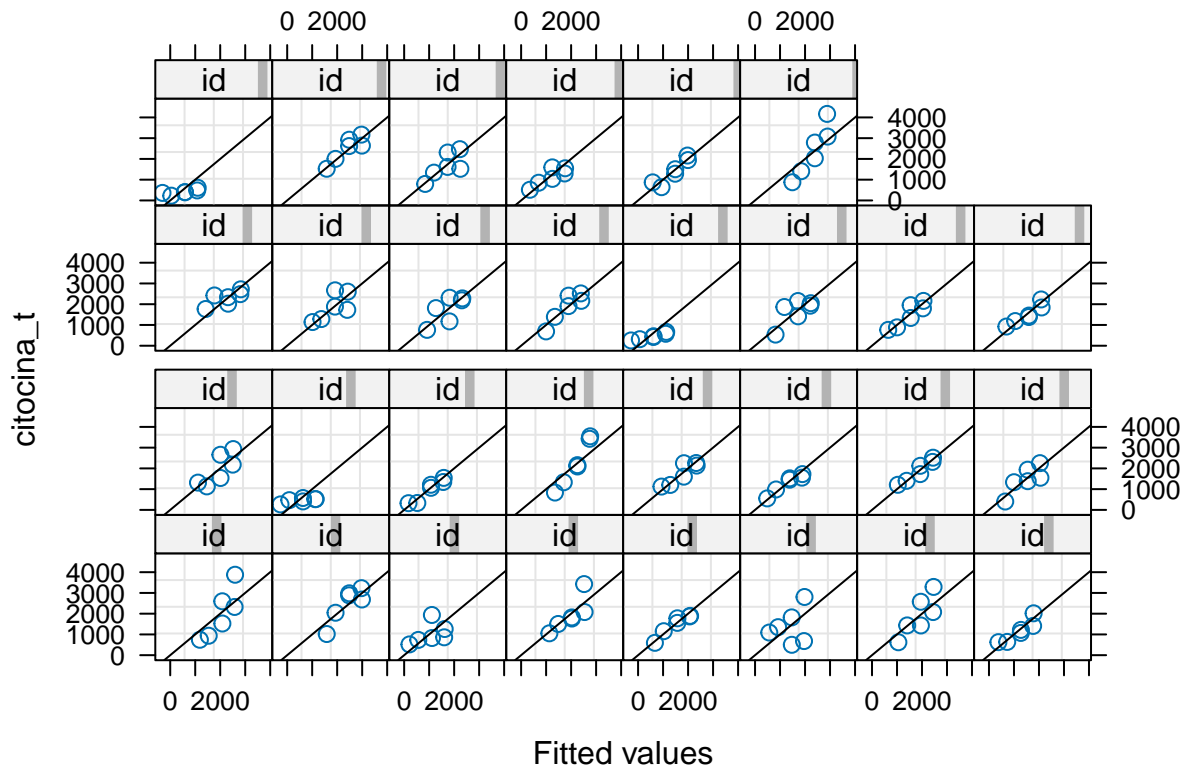
Podemos observar que o modelo que produz o menor AIC e BIC, parece ser o modelo que inclui todas as variáveis e um **intercepto aleatório** (Ajuste 1).

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))
```

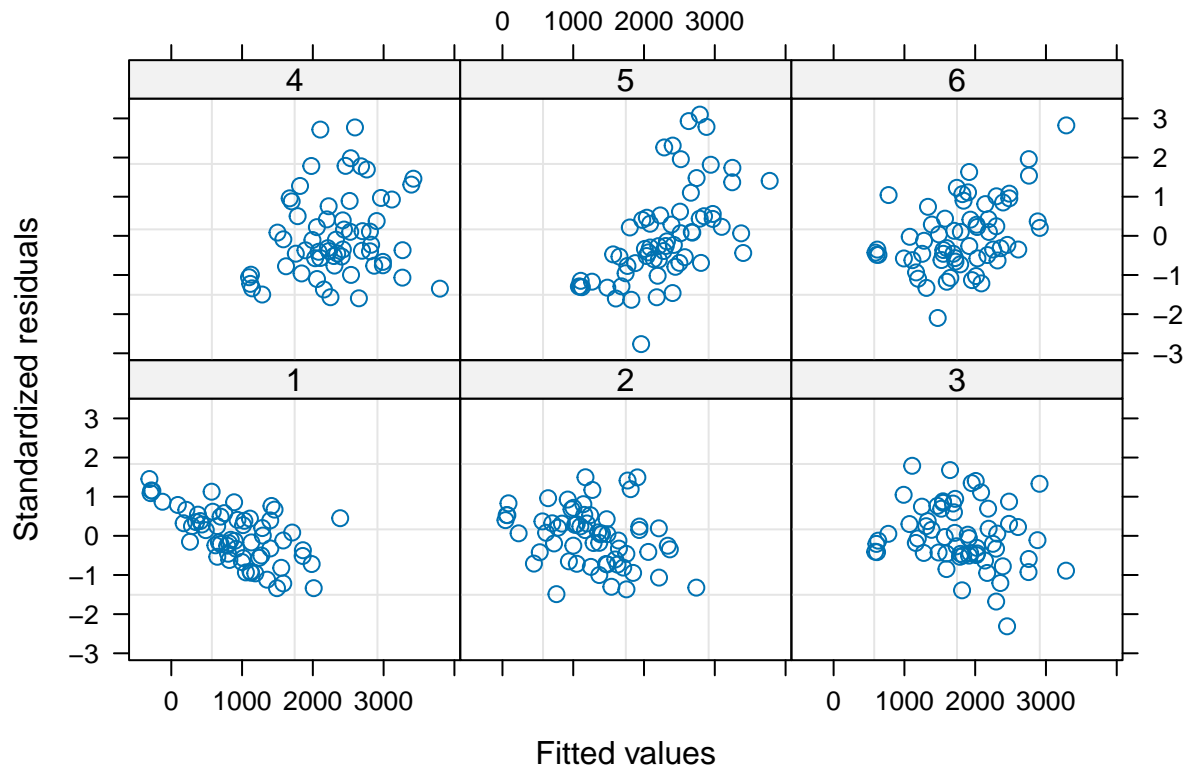




Com base nos gráficos acima, podemos observar que o modelo aparenta apresentar um bom ajuste para os indivíduos. Porém, podemos ver que para alguns pacientes, os pontos se afastam bastante da reta identidade, indicando que, para esse pacientes, o modelo não se ajusta tão bem assim.

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)
```

De maneira geral, os resíduos aparentam se distribuir em torno de zero, o que poderia ser considerado um indicativo de um bom ajuste. Porém, podemos notar que a distribuição dos resíduos não segue um padrão aleatório na maior porção dos tempos, o que indica que esse modelo pode não ser adequado aos dados.

Conclusão

Em ambos os modelos, observamos que a variável associada à observação está fortemente relacionada ao valor observado da citocina Tnf-receptor. Além disso, pacientes do sexo masculino apresentam um aumento médio nos níveis da citocina, assim como a variável idade também está positivamente associada. Por outro lado, não identificamos um efeito isolado claro do grupo (tipo de cirurgia) em nenhum dos modelos. Esse efeito apenas se manifesta consistentemente quando consideramos a interação entre grupo e sexo. Isso sugere que o tipo de cirurgia, por si só, não influencia diretamente a medição da citocina, mas sim em combinação com o sexo do paciente.

A opção entre um modelo misto e marginal, depende muito do objetivo da análise. De modo geral, O modelo misto é ideal quando o objetivo é capturar a variação intraindivíduo, enquanto o modelo marginal (GEE) foca nas diferenças entre grupos e tratamentos aplicados.

Para o contexto desta análise, optamos pelo modelo marginal (GEE) com um termo de interação entre **Sexo** e **Grupo**. A interpretação dos coeficientes no GEE é mais direta do que no modelo misto, e a inclusão da interação entre Sexo e Grupo revelou-se relevante, pois evidenciou um efeito relacionado ao tipo de cirurgia.

A seguir, apresentamos a interpretação de alguns coeficientes do modelo em relação à variável resposta:

```
knitr::kable(
  round(coef(summary(ajuste_gee_sexgrup)), 4))
```

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-316.2064	683.8620	0.2138	0.6438
tempo2	352.4436	50.9087	47.9287	0.0000

	Estimate	Std.err	Wald	Pr(> W)
tempo3	896.7094	60.0779	222.7792	0.0000
tempo4	1408.4075	85.0438	274.2658	0.0000
tempo5	1389.1071	98.0206	200.8340	0.0000
tempo6	900.5282	71.8244	157.1991	0.0000
sexoM	255.2614	214.1788	1.4204	0.2333
imc	15.0799	18.7477	0.6470	0.4212
nyha2	185.7508	176.3024	1.1101	0.2921
nyha3	16.4539	174.2651	0.0089	0.9248
nyha4	-29.4728	246.7537	0.0143	0.9049
idade	10.4601	7.5537	1.9175	0.1661
euroes	-8.8747	55.1243	0.0259	0.8721
grupo1:sexoF	291.0020	211.1219	1.8999	0.1681
grupo2:sexoF	802.3328	227.8074	12.4043	0.0004
grupo1:sexoM	-330.7921	164.3475	4.0512	0.0441
grupo2:sexoM	137.1992	227.3267	0.3643	0.5462

De maneira geral, observamos que cada intervalo de tempo está associado a um aumento na média da citocina Tnf-receptor. Os maiores aumentos médios foram registrados no tempo 4, com um incremento de 1408,4 unidades, e no tempo 5, com um incremento de 1389,1 unidades, considerando as demais covariáveis constantes.

Também identificamos que a interação entre sexo feminino e os diferentes grupos está associada a medições mais altas da citocina em comparação aos indivíduos do sexo masculino nos mesmos grupos. Destaca-se o grupo 2, que apresentou o maior aumento médio em relação aos demais.

Além disso, verificamos que um aumento na idade está associado a um incremento médio de 10,4 unidades na citocina, considerando todas as demais covariáveis fixadas.