Avaliação Pontual 3 - Cap. 6 (pg. 163-164)

Modelos para Dados Correlacionados

Nicolas Mathias Hahn

Conteúdo

6.1.1 On a single graph, construct a time plot that displays the mean weight versus time (in weeks) for the three groups. Describe the general characteristics of the time trends for the three groups. 4

6.1.2 Read the data from the external file and put the data in a "univariate" or "long" format, with five "records" per subject. 5

6.1.3 Assume that the rate of increase in each group is approximately constant throughout the duration of the study. Assuming an unstructured covariance matrix, construct a test of whether the rate of increase differs in the groups. 9

6.1.4 On a single graph, construct a time plot that displays the estimated mean weight versus time (in weeks) for the three treatment groups from the results generated from Problem 6.1.3. 10

6.1.5 Based on the results from Problem 6.1.3, what is the estimated rate of increase in mean weight in the control group (group 1)? What is the estimated rate of increase in mean weight in the thiouracil group (group 2)? What is the estimated rate of increase in mean weight in the thyroxin group (group 3)? 11

6.1.6 The study investigators conjectured that there would be an increase in weight, but that the rate of increase would level-off toward the end of the study. They also conjectured that this pattern of change may differ in the three treatment groups. Assuming an unstructured covariance matrix, construct a test of this hypothesis. 12

6.1.7 Compare and contrast the results from Problems 6.1.3 and 6.1.6. Does a model with only a linear trend in time adequately account for the pattern of change in the three treatments groups? Provide results that support your conclusion. 14

6.1.8 Given the results of all the previous analyses, what conclusions can be drawn about the effect of the additives on the patterns of change in weight? 16

Código necessário para execução dos scripts mais adiante:

```
library(haven)
df = read_dta(here::here("rat.dta"))
df$id = as.factor(df$id)
df$group = factor(df$group, labels = c("control","thiouracial","thyroxin"))

library(tidyr)
df_long = gather(data = df, key = "week", value = "weight", -id, -group)
df_long$week = factor(df_long$week, labels = c("0","1","2","3","4"))
df_long$week = as.numeric(as.character(df_long$week))
df_long$time = as.numeric(factor(df_long$week))
df_long = df_long[order(df_long$id, decreasing = F),]


library(plyr)
df_long_resumo = ddply(df_long, ~ group + week, summarize, weight.mean = mean(weight))
```

Problems

6.1 In a study of weight gain (Box, 1950) investigators randomly assigned 30 rats to three treatment groups: treatment 1 was a control (no additive); treatments 2 and 3 consisted of two different additives (thiouracil and thyroxin respectively) to the rats drinking water. Weight, in grams, was measured at baseline (week 0) and at weeks 1, 2, 3, and 4. Due to an accident at the beginning of the study, data on 3 rats from the thyroxin group are unavailable.

The raw data are stored in an external file: `rat.dat`

Each row of the data set contains the following seven variables:

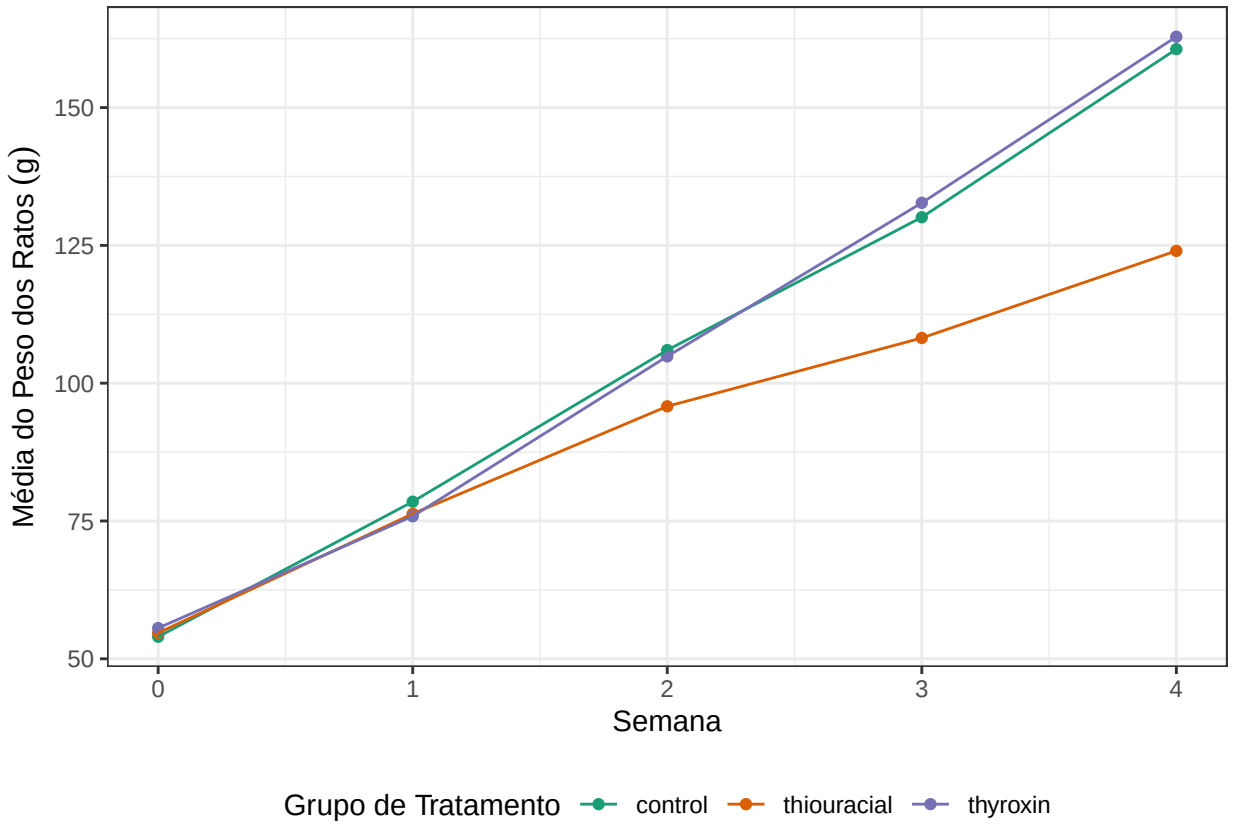
ID Group Y_1 Y_2 Y_3 Y_4 Y_5

Note: The variable Group is coded 1 = control, 2 = thiouracil, and 3 = thyroxin.

6.1.1 On a single graph, construct a time plot that displays the mean weight versus time (in weeks) for the three groups. Describe the general characteristics of the time trends for the three groups.

```
library(ggplot2)

p_obs = ggplot(data = df_long_resumo,
               mapping = aes(x = week, y = weight.mean, group = group, colour = group)) +
  geom_point() +
  geom_line() +
  scale_color_brewer(palette = "Dark2") +
  labs(x = "Semana",
       y = expression("Média do Peso dos Ratos"~(g)),
       colour = "Grupo de Tratamento") +
  theme_bw() +
  theme(legend.position = "bottom")
p_obs
```



Com base no gráfico acima, podemos observar que houve um aumento de peso médio dos ratos ao longo da semana. No entanto, é bastante notável que o peso médio dos grupos *control* e *thyroxin* ficaram bastante próximos na semana 4 (acima de 150g), mas o peso médio dos ratos do grupo *thiouracial* ficou menor do que os dos outros grupos (cerca de 125g) - o que pode indicar uma redução no ganho de massa dos ratos expostos ao *thiouracial*.

6.1.2 Read the data from the external file and put the data in a "univariate" or "long" format, with five "records" per subject.

```
knitr::kable(df)
```

id	group	y1	y2	y3	y4	y5
1	control	57	86	114	139	172
2	control	60	93	123	146	177
3	control	52	77	111	144	185
4	control	49	67	100	129	164
5	control	56	81	104	121	151
6	control	46	70	102	131	153
7	control	51	71	94	110	141
8	control	63	91	112	130	154
9	control	49	67	90	112	140
10	control	57	82	110	139	169
11	thiouracial	61	86	109	120	129
12	thiouracial	59	80	101	111	122
13	thiouracial	53	79	100	106	133
14	thiouracial	59	88	100	111	122
15	thiouracial	51	75	101	123	140
16	thiouracial	51	75	92	100	119
17	thiouracial	56	78	95	103	108
18	thiouracial	58	69	93	114	138
19	thiouracial	46	61	78	90	107
20	thiouracial	53	72	89	104	122
21	thyroxin	59	85	121	146	181
22	thyroxin	54	71	90	110	138
23	thyroxin	56	75	108	151	189
24	thyroxin	59	85	116	148	177
25	thyroxin	57	72	97	120	144
26	thyroxin	52	73	97	116	140
27	thyroxin	52	70	105	138	171

```
knitr::kable(df_long)
```

id	group	week	weight	time
1	control	0	57	1
1	control	1	86	2
1	control	2	114	3
1	control	3	139	4
1	control	4	172	5
2	control	0	60	1
2	control	1	93	2
2	control	2	123	3
2	control	3	146	4
2	control	4	177	5
3	control	0	52	1
3	control	1	77	2
3	control	2	111	3

id	group	week	weight	time
3	control	3	144	4
3	control	4	185	5
4	control	0	49	1
4	control	1	67	2
4	control	2	100	3
4	control	3	129	4
4	control	4	164	5
5	control	0	56	1
5	control	1	81	2
5	control	2	104	3
5	control	3	121	4
5	control	4	151	5
6	control	0	46	1
6	control	1	70	2
6	control	2	102	3
6	control	3	131	4
6	control	4	153	5
7	control	0	51	1
7	control	1	71	2
7	control	2	94	3
7	control	3	110	4
7	control	4	141	5
8	control	0	63	1
8	control	1	91	2
8	control	2	112	3
8	control	3	130	4
8	control	4	154	5
9	control	0	49	1
9	control	1	67	2
9	control	2	90	3
9	control	3	112	4
9	control	4	140	5
10	control	0	57	1
10	control	1	82	2
10	control	2	110	3
10	control	3	139	4
10	control	4	169	5
11	thiouracial	0	61	1
11	thiouracial	1	86	2
11	thiouracial	2	109	3
11	thiouracial	3	120	4
11	thiouracial	4	129	5
12	thiouracial	0	59	1
12	thiouracial	1	80	2
12	thiouracial	2	101	3
12	thiouracial	3	111	4
12	thiouracial	4	122	5
13	thiouracial	0	53	1
13	thiouracial	1	79	2
13	thiouracial	2	100	3
13	thiouracial	3	106	4
13	thiouracial	4	133	5

id	group	week	weight	time
14	thiouracial	0	59	1
14	thiouracial	1	88	2
14	thiouracial	2	100	3
14	thiouracial	3	111	4
14	thiouracial	4	122	5
15	thiouracial	0	51	1
15	thiouracial	1	75	2
15	thiouracial	2	101	3
15	thiouracial	3	123	4
15	thiouracial	4	140	5
16	thiouracial	0	51	1
16	thiouracial	1	75	2
16	thiouracial	2	92	3
16	thiouracial	3	100	4
16	thiouracial	4	119	5
17	thiouracial	0	56	1
17	thiouracial	1	78	2
17	thiouracial	2	95	3
17	thiouracial	3	103	4
17	thiouracial	4	108	5
18	thiouracial	0	58	1
18	thiouracial	1	69	2
18	thiouracial	2	93	3
18	thiouracial	3	114	4
18	thiouracial	4	138	5
19	thiouracial	0	46	1
19	thiouracial	1	61	2
19	thiouracial	2	78	3
19	thiouracial	3	90	4
19	thiouracial	4	107	5
20	thiouracial	0	53	1
20	thiouracial	1	72	2
20	thiouracial	2	89	3
20	thiouracial	3	104	4
20	thiouracial	4	122	5
21	thyroxin	0	59	1
21	thyroxin	1	85	2
21	thyroxin	2	121	3
21	thyroxin	3	146	4
21	thyroxin	4	181	5
22	thyroxin	0	54	1
22	thyroxin	1	71	2
22	thyroxin	2	90	3
22	thyroxin	3	110	4
22	thyroxin	4	138	5
23	thyroxin	0	56	1
23	thyroxin	1	75	2
23	thyroxin	2	108	3
23	thyroxin	3	151	4
23	thyroxin	4	189	5
24	thyroxin	0	59	1
24	thyroxin	1	85	2

id	group	week	weight	time
24	thyroxin	2	116	3
24	thyroxin	3	148	4
24	thyroxin	4	177	5
25	thyroxin	0	57	1
25	thyroxin	1	72	2
25	thyroxin	2	97	3
25	thyroxin	3	120	4
25	thyroxin	4	144	5
26	thyroxin	0	52	1
26	thyroxin	1	73	2
26	thyroxin	2	97	3
26	thyroxin	3	116	4
26	thyroxin	4	140	5
27	thyroxin	0	52	1
27	thyroxin	1	70	2
27	thyroxin	2	105	3
27	thyroxin	3	138	4
27	thyroxin	4	171	5

```
knitr::kable(df_long_resumo)
```

group	week	weight.mean
control	0	54.00000
control	1	78.50000
control	2	106.00000
control	3	130.10000
control	4	160.60000
thiouracial	0	54.70000
thiouracial	1	76.30000
thiouracial	2	95.80000
thiouracial	3	108.20000
thiouracial	4	124.00000
thyroxin	0	55.57143
thyroxin	1	75.85714
thyroxin	2	104.85714
thyroxin	3	132.71429
thyroxin	4	162.85714

6.1.3 Assume that the rate of increase in each group is approximately constant throughout the duration of the study. Assuming an unstructured covariance matrix, construct a test of whether the rate of increase differs in the groups.

```
library(nlme)
# modelo de tendências lineares
mod_lin = gls(weight ~ week +
                    week:group,
                  corr = corSymm(form = ~ time | id),
                  weights = varIdent(form = ~ 1 | time),
                  method = "REML",
                  data = df_long)

knitr::kable(
  summary(mod_lin)$tTable,
  digits = c(4, 4, 2),
  col.names = c("Estimativa", "EP", "Z", "p-valor"))
```

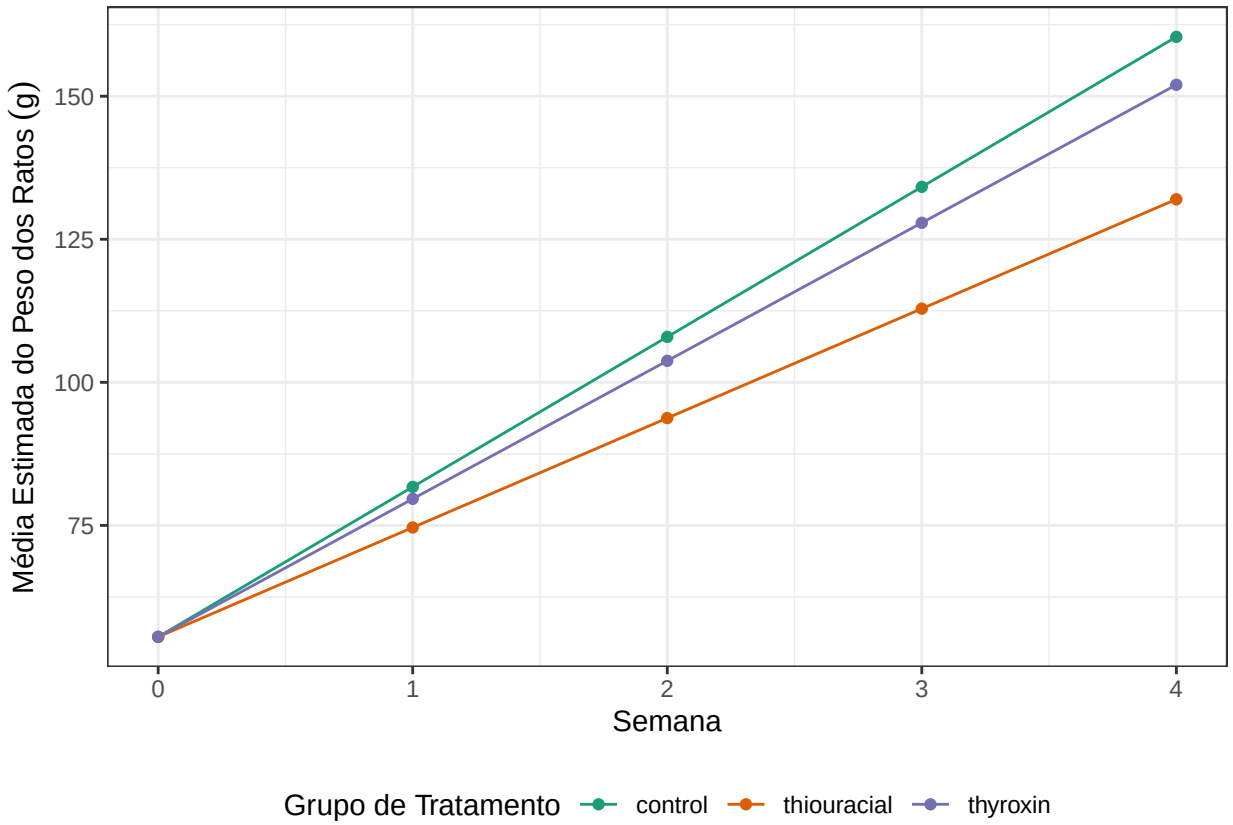
	Estimativa	EP	Z	p-valor
(Intercept)	55.5138	0.8189	67.79	0.000
week	26.2151	1.0163	25.79	0.000
week:groupthiouracial	-7.0964	1.4132	-5.02	0.000
week:groupthyroxin	-2.0944	1.5573	-1.34	0.181

Com base na tabela acima, temos que a taxa de aumento de peso entre os grupos *controle* e *thyroxin* não difere significativamente a 5%. Já entre os grupos *controle* e *thiouracil* há diferença significativa a 5% de significância.

6.1.4 On a single graph, construct a time plot that displays the estimated mean weight versus time (in weeks) for the three treatment groups from the results generated from Problem 6.1.3.

```
library(ggeffects)
est_mean = ggpredict(mod_lin, terms = c("week", "group"))

p_estmean_modlin = ggplot(data = est_mean,
            mapping = aes(x = x, y = predicted, group = group, colour = group)) +
  geom_point() +
  geom_line() +
  scale_color_brewer(palette = "Dark2") +
  labs(x = "Semana",
       y = expression("Média Estimada do Peso dos Ratos"~(g)),
       colour = "Grupo de Tratamento") +
  theme_bw() +
  theme(legend.position = "bottom")
p_estmean_modlin
```



Com base no gráfico acima, podemos chegar na mesma conclusão que em *6.1.1*: que houve um aumento de peso médio dos ratos ao longo da semana. No entanto, é bastante notável que o peso médio estimado dos grupos *control* e *thyroxin* ficaram bastante próximos na semana 4 (acima de 150g), mas o peso médio estimado dos ratos do grupo *thiouracial* ficou menor do que os dos outros grupos (cerca de 125g) - o que pode indicar uma ganho de massa menor dos ratos expostos ao *thiouracial*.

6.1.5 Based on the results from Problem 6.1.3, what is the estimated rate of increase in mean weight in the control group (group 1)? What is the estimated rate of increase in mean weight in the thiouracil group (group 2)? What is the estimated rate of increase in mean weight in the thyroxin group (group 3)?

```
est = as.numeric(summary(mod_lin)$tTable[,"Value"])
rslt = data.frame(estimated_rate=double(3))
rownames(rslt) = c("control","thiouracil","thyroxin")
rslt$estimated_rate = c(est[2],est[2]+est[3],est[2]+est[4])

knitr::kable(rslt, digits = 4)
```

	estimated_rate
control	26.2151
thiouracil	19.1187
thyroxin	24.1207

Podemos observar, na tabela acima, a taxa estimada de alteração do peso para cada grupo:

- grupo *control* com taxa estimada de 26.21509g;
- grupo *thiouracil* com taxa estimada de 19.11870g;
- grupo *thyroxin* com taxa estimada de 24.12073g.

6.1.6 The study investigators conjectured that there would be an increase in weight, but that the rate of increase would level-off toward the end of the study. They also conjectured that this pattern of change may differ in the three treatment groups. Assuming an unstructured covariance matrix, construct a test of this hypothesis.

```
# modelo linear por partes (spline linear com knot em semana 2)
mod_comp = gls(weight ~ week + I( (week - 2) * (week >= 2) ) +
                    week:group + I( (week - 2) * (week >= 2) ):group,
                  corr = corSymm(form = ~ time | id),
                  weights = varIdent(form = ~ 1 | time),
                  method = "ML",
                  data = df_long)

# modelo de tendências lineares
mod_red = gls(weight ~ week +
                    week:group,
                  corr = corSymm(form = ~ time | id),
                  weights = varIdent(form = ~ 1 | time),
                  method = "ML",
                  data = df_long)


# logLik(mod_comp) # modelo linear por partes
# logLik(mod_red)  # modelo linear

anova(mod_comp,mod_red)
```

```
##          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod_comp     1 22 856.3033 920.2194 -406.1517
## mod_red      2 19 867.1784 922.3786 -414.5892 1 vs 2 16.87511   7e-04
```

O resultado do teste indica que o modelo linear por partes melhora, significativamente, o ajuste global da resposta média ao longo do tempo quando comparado ao modelo linear ($p \approx 0.0007$). Sendo assim, vamos observar as taxas estimadas de aumento de peso para cada grupo no início (antes da semana 2) e no final (após semana 2) do estudo:

```
# modelo linear por partes (spline linear com knot em semana 2)
mod_linparts = gls(weight ~ week + I( (week - 2) * (week >= 2) ) +
                    week:group + I( (week - 2) * (week >= 2) ):group,
                  corr = corSymm(form = ~ time | id),
                  weights = varIdent(form = ~ 1 | time),
                  method = "REML",
                  data = df_long)

est_mean2 = ggpredict(mod_linparts, terms = c("week", "group"))

knitr::kable(
  summary(mod_linparts)$tTable,
  digits = c(4, 4, 2),
  col.names = c("Estimativa", "EP", "Z", "p-valor"))
```

	Estimativa	EP	Z	p-valor
(Intercept)	55.5587	0.8190	67.84	0.0000
week	26.5723	1.1034	24.08	0.0000
I((week - 2) * (week >= 2))	-1.0303	1.2243	-0.84	0.4016
week:groupthiouracial	-5.4309	1.5278	-3.55	0.0005
week:groupthyroxin	-2.4583	1.6836	-1.46	0.1467
I((week - 2) * (week >= 2)):groupthiouracial	-4.9436	1.7235	-2.87	0.0048
I((week - 2) * (week >= 2)):groupthyroxin	1.0803	1.8992	0.57	0.5705

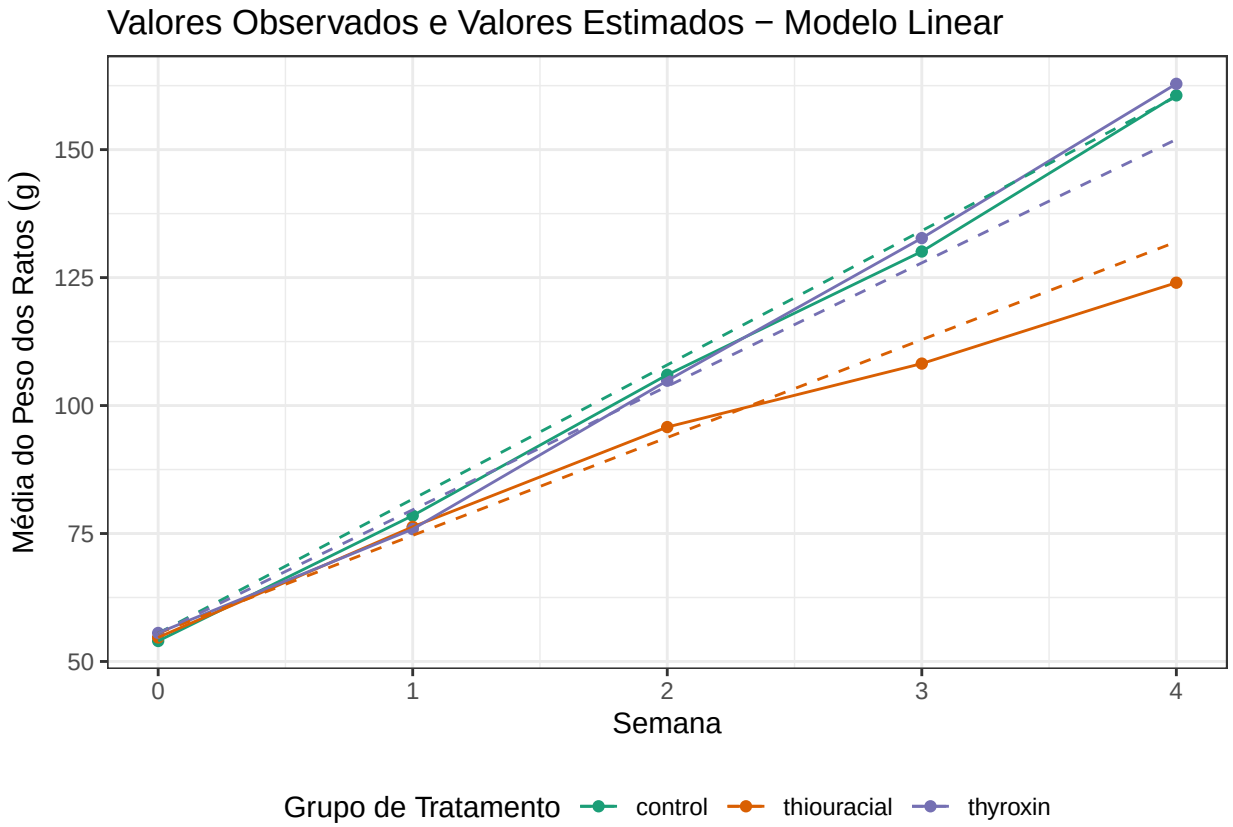
Com base na tabela acima, temos que não há diferença significativa (a 5%) na taxa de aumento de peso dos grupos *control* e *thyroxin* entre o início e o final do estudo. No entanto, há diferença significatica (a 5% de significância) no grupo *thiouracil*, evidenciando uma redução na taxa de aumento de peso. Dessa forma, podemos concluir que há sim uma queda na taxa de aumento de peso no final do estudo, mas apenas para o grupo *thiouracil*.

6.1.7 Compare and contrast the results from Problems 6.1.3 and 6.1.6. Does a model with only a linear trend in time adequately account for the pattern of change in the three treatments groups? Provide results that support your conclusion.

Um modelo com apenas uma tendência linear no tempo é capaz de se ajustar para o padrão de mudança no peso dos ratos de cada grupo. Apesar de ser realizar um ajuste satisfatório, há opções mais adequadas. Vamos visualizar tal fato nos gráficos abaixo:

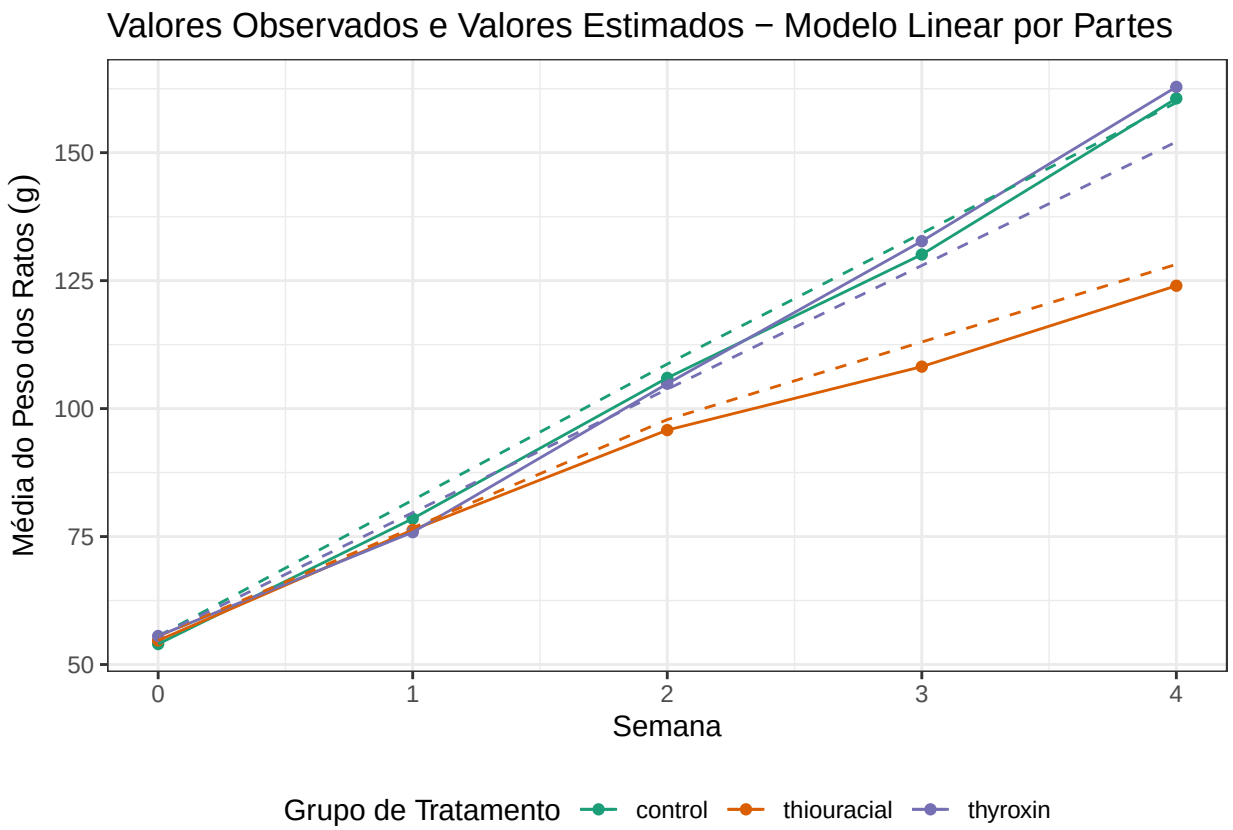
- **Obs.: linhas tracejadas são as médias estimadas do peso, enquanto que as linhas sólidas são as médias observadas do peso**.

```
p = p_obs + geom_line(data = est_mean, aes(x=x,y=predicted,group=group,colour=group),
                      linetype="dashed") +
  ggtitle("Valores Observados e Valores Estimados - Modelo Linear")
p
```



Fica claro, pelo gráfico acima, que o modelo linear tem uma boa aproximação do peso médio de cada grupo. No entanto, vale destacar a mudança mais expressiva do ganho de massa dos ratos do grupo *thiouracil* após a semana 2. Agora, vamos observar o modelo linear por partes como um modelo alternativo para os dados:

```
p = p_obs + geom_line(data = est_mean2, aes(x=x,y=predicted,group=group,colour=group),
                      linetype="dashed") +
  ggtitle("Valores Observados e Valores Estimados - Modelo Linear por Partes")
p
```



Como é possível observar, o modelo linear por partes capta de maneira mais adequada a mudança de peso do grupo *thiouracil*, que é o grupo que de fato tem uma mudança significativa na taxa de aumento de peso após a semana 2. Portanto, podemos concluir - com base nos resultados - que o modelo linear considera satisfatoriamente o padrão de mudança de peso dos grupos de tratamento estudados, mas há alternativas mais adequadas, como o modelo linear por partes.

6.1.8 Given the results of all the previous analyses, what conclusions can be drawn about the effect of the additives on the patterns of change in weight?

Com base nos resultados anteriores, podemos concluir que:

- os grupos *control* e *thyroxin* não diferem significativamente no ganho de peso ao longo do tempo;
- o grupo *thiouracil* difere significativamente no ganho de peso dos outros grupos;
- o grupo *thiouracil* é o único com diferença expressiva no ganho de peso (sendo, em média, menor o ganho de peso) após a semana 2.

Por conseguinte, temos evidências para concluir que o aditivo *thiouracil* impacta significativamente no ganho de peso dos ratos ao longo do tempo, e que o aditivo *thyroxin* não impactou, significativamente (a 5% de significância) no ganho de massa dos ratos.