

We're going to analyze next the Cognitive data set (See A.5 for details). This data set is about a school lunch intervention program in Kenya. There are 2555 observations on 534 individuals. For some reason, this is different than what the book claims, so the results are somewhat different.

Each individual was given a cognitive exam (called Raven's) five times (though some are missing). After the first time, the treatment was started; each individual was randomly assigned to one of four treatments; three groups were fed a school lunch that was supplemented with either meat, milk, or oil to create a lunch with a given caloric value. The fourth group (considered the control) did not receive a lunch; instead their families received a goat at the end of the study.

We have the following variables:

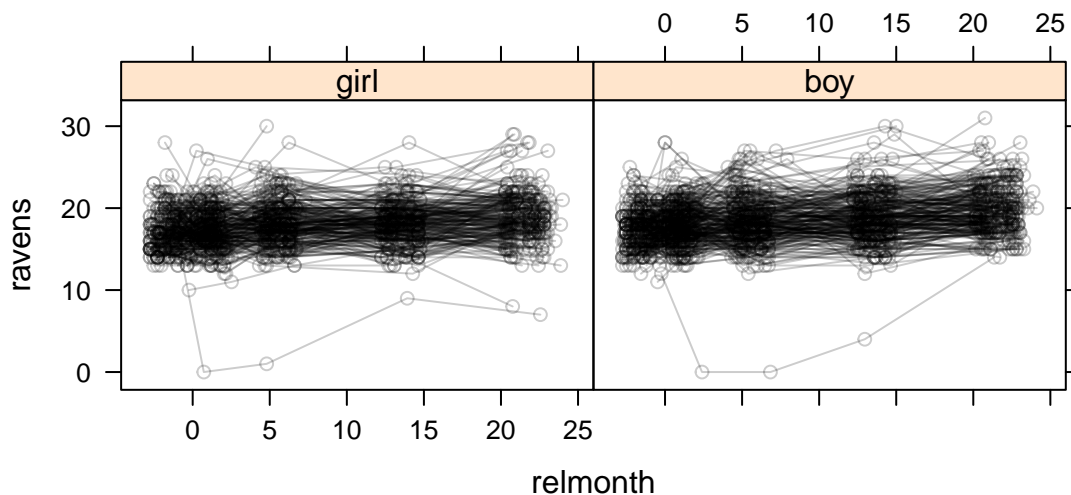
- id: identifier for each individual
- rn: the number of the cognitive exam (1 to 5)
- relmonth: the month, relative to when the treatment started
- sex: boy/girl
- ses1: a measure of the family's socioeconomic status
- ravens: the result from the Raven's cognitive measure
- treatment: the treatment the individual was assigned to

Here's some example data:

```
> subset(r1, id == 1)
```

| id | rn | relmonth | sex | ses1 | ravens | treatment |
|----|----|----------|------|------|--------|-----------|
| 1 | 1 | -1.80 | girl | 89 | 15 | meat |
| 2 | 1 | 1.68 | girl | 89 | 19 | meat |
| 3 | 1 | 5.52 | girl | 89 | 21 | meat |
| 4 | 1 | 14.28 | girl | 89 | 18 | meat |
| 5 | 1 | 20.88 | girl | 89 | 21 | meat |

And here are profile plots for the boys and girls separately. We'll leave in all the data, though there some potential outliers.



Treatment Effect (ignoring Time), different by Sex?

See page 200, section 7.2.3.5.

We'll start by modeling constant cognitive scores for each individual over time, but allowing for a change in score because of a treatment.

We'll also investigate if either these scores or the change in score due to treatment differ by sex.

The treatment effect is time-varying, just like in the pediatric pain data, so we'll make a new variable describing the treatment in effect at each observation.

```
> r1$trt <- r1$treatment
> r1$trt[r1$rn == 1] <- "control"
> subset(r1, id == 1)
```

| | id | rn | relmonth | sex | ses1 | ravens | treatment | trt |
|---|----|----|----------|------|------|--------|-----------|---------|
| 1 | 1 | 1 | -1.80 | girl | 89 | 15 | meat | control |
| 2 | 1 | 2 | 1.68 | girl | 89 | 19 | meat | meat |
| 3 | 1 | 3 | 5.52 | girl | 89 | 21 | meat | meat |
| 4 | 1 | 4 | 14.28 | girl | 89 | 18 | meat | meat |
| 5 | 1 | 5 | 20.88 | girl | 89 | 21 | meat | meat |

We'll fit three models: no treatment effect, an additive effect, and an interaction effect.

```
> m2a <- gls(ravens ~ sex, correlation = corAR1(form = ~1 | id),
+   r1, method = "ML")
> m2b <- gls(ravens ~ sex + trt, correlation = corAR1(form = ~1 |
+   id), r1, method = "ML")
> m2c <- gls(ravens ~ sex * trt, correlation = corAR1(form = ~1 |
+   id), r1, method = "ML")
```

How do we interpret these three models?

```
> anova(m2a, m2b, m2c)
```

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|-----|-------|----|----------|----------|-----------|--------|-----------|---------|
| m2a | 1 | 4 | 12459.38 | 12482.76 | -6225.689 | | | |
| m2b | 2 | 7 | 12437.08 | 12478.00 | -6211.541 | 1 vs 2 | 28.295483 | <.0001 |
| m2c | 3 | 10 | 12440.86 | 12499.31 | -6210.428 | 2 vs 3 | 2.226269 | 0.5268 |

Which do you prefer? Why?

```
> m2 <- update(m2b, method = "REML")
> summary(m2)
```

```
...
Correlation Structure: AR(1)
Formula: ~1 | id
Parameter estimate(s):
  Phi
0.3798315
```

```
Coefficients:
      Value Std.Error  t-value p-value
(Intercept) 17.660867 0.1324080 133.38214 0.0000
sexboy       0.600918 0.1544913   3.88966 0.0001
trtmilk      0.291416 0.1804535   1.61491 0.1065
trtcalorie   0.408083 0.1815113   2.24825 0.0246
trtmeat      0.973403 0.1871550   5.20105 0.0000
...
```

How do we interpret these coefficients?

```
> K
```

```
      [,1] [,2] [,3] [,4] [,5]
girls control  1  0  0  0  0
girls milk     1  0  1  0  0
girls calorie  1  0  0  1  0
girls meat     1  0  0  0  1
boys control   1  1  0  0  0
boys milk      1  1  1  0  0
boys calorie   1  1  0  1  0
boys meat      1  1  0  0  1
```

```
> t2 <- glht(m2, linfct = K)
> summary(t2, test = adjusted(type = "none"))
```

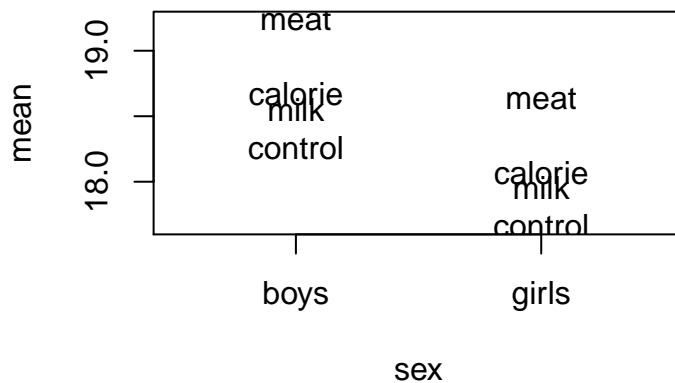
Simultaneous Tests for General Linear Hypotheses

```
Fit: gls(model = ravens ~ sex + trt, data = r1, correlation = corAR1(form = ~1 |
id), method = "REML")
```

Linear Hypotheses:

| | Estimate | Std. Error | z value | Pr(> z) |
|--------------------|----------|------------|---------|----------|
| girls control == 0 | 17.6609 | 0.1324 | 133.38 | <2e-16 |
| girls milk == 0 | 17.9523 | 0.1800 | 99.73 | <2e-16 |
| girls calorie == 0 | 18.0690 | 0.1806 | 100.04 | <2e-16 |
| girls meat == 0 | 18.6343 | 0.1843 | 101.10 | <2e-16 |
| boys control == 0 | 18.2618 | 0.1286 | 142.01 | <2e-16 |
| boys milk == 0 | 18.5532 | 0.1749 | 106.08 | <2e-16 |
| boys calorie == 0 | 18.6699 | 0.1761 | 106.04 | <2e-16 |
| boys meat == 0 | 19.2352 | 0.1839 | 104.57 | <2e-16 |

(Adjusted p values reported -- none method)

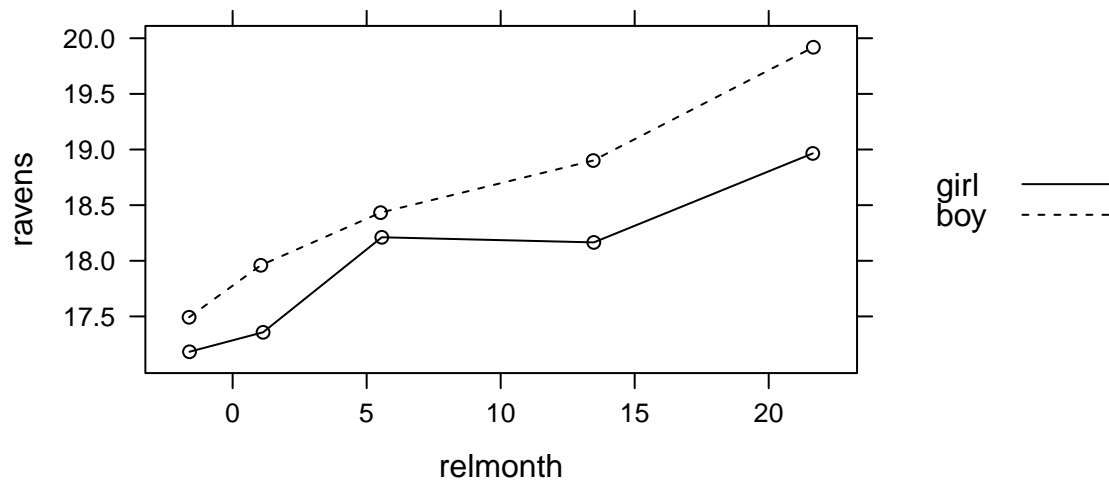


What is this model missing? How might we add it?

Time as linear effect, different by Sex?

See page 208, section 7.3.1.

Here are the means for each testing time for each gender.



It looks roughly linear, so we'll fit some models assuming linearity and compare them.

What terms are included in each of these models? How would you interpret them? Do they all make sense?

```
> m1 <- gls(ravens ~ relmonth, correlation = corAR1(form = ~1 |
+   id), data = r1, method = "ML")
> m2 <- gls(ravens ~ relmonth + sex, correlation = corAR1(form = ~1 |
+   id), data = r1, method = "ML")
> m3 <- gls(ravens ~ relmonth + relmonth:sex, correlation = corAR1(form = ~1 |
+   id), data = r1, method = "ML")
> m4 <- gls(ravens ~ relmonth * sex, correlation = corAR1(form = ~1 |
+   id), data = r1, method = "ML")
```

```
> anova(m1, m2, m3, m4, test = FALSE)
```

| | Model | df | AIC | BIC | logLik |
|----|-------|----|----------|----------|-----------|
| m1 | 1 | 4 | 12340.92 | 12364.30 | -6166.459 |
| m2 | 2 | 5 | 12327.70 | 12356.93 | -6158.848 |
| m3 | 3 | 5 | 12326.65 | 12355.88 | -6158.325 |
| m4 | 4 | 6 | 12325.39 | 12360.47 | -6156.697 |

```
> anova(m2, m4)
```

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|----|-------|----|----------|----------|-----------|--------|----------|---------|
| m2 | 1 | 5 | 12327.70 | 12356.93 | -6158.848 | | | |
| m4 | 2 | 6 | 12325.39 | 12360.47 | -6156.697 | 1 vs 2 | 4.302586 | 0.0381 |

Which do you prefer? Why?

```
> summary(m4)
```

```
...  
Correlation Structure: AR(1)  
Formula: ~1 | id  
Parameter estimate(s):  
  Phi  
0.3674618
```

```
Coefficients:  
                Value Std.Error  t-value p-value  
(Intercept)    17.417994 0.13709804 127.04773 0.0000  
relmonth         0.068478 0.01045732   6.54837 0.0000  
sexboy           0.341878 0.18944732   1.80461 0.0713  
relmonth:sexboy  0.029807 0.01437492   2.07353 0.0382  
...
```

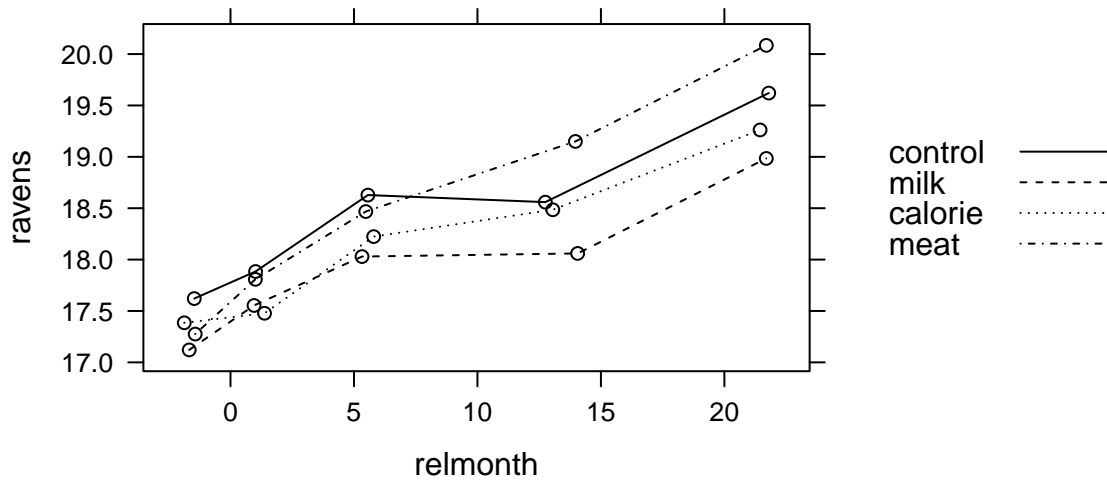
How do you interpret these coefficients?

Time as Linear Effect, different by Treatment?

See page 212, section 7.3.4.

Finally, we could include time as linear effect, and see how it differs by treatment.

Here's a plot of those means.



What kind of lines would we want to fit here? How would we want the control line to be related to the treatment line? What should change at time 0 (when the treatment starts)?

```
> m1 <- gls(ravens ~ relmonth + relmonth:trt, correlation = corAR1(form = ~1 |
+   id), data = r1)
> summary(m1)
```

```
...
Correlation Structure: AR(1)
Formula: ~1 | id
Parameter estimate(s):
  Phi
0.3701123
```

Coefficients:

| | Value | Std.Error | t-value | p-value |
|---------------------|-----------|------------|-----------|---------|
| (Intercept) | 17.602786 | 0.09534034 | 184.63105 | 0.0000 |
| relmonth | 0.092043 | 0.01209884 | 7.60758 | 0.0000 |
| relmonth:trtmilk | -0.035178 | 0.01601865 | -2.19606 | 0.0282 |
| relmonth:trtcalorie | -0.015876 | 0.01626785 | -0.97589 | 0.3292 |
| relmonth:trtmeat | 0.021695 | 0.01635388 | 1.32659 | 0.1848 |

...

Here are slopes:

```
> K1
```

| | [,1] | [,2] | [,3] | [,4] | [,5] |
|---------------|------|------|------|------|------|
| control slope | 0 | 1 | 0 | 0 | 0 |
| milk slope | 0 | 1 | 1 | 0 | 0 |
| calorie slope | 0 | 1 | 0 | 1 | 0 |
| meat slope | 0 | 1 | 0 | 0 | 1 |

```
> t1 <- glht(m1, linfct = K1)
> summary(t1, test = adjusted(type = "none"))
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: gls(model = ravens ~ relmonth + relmonth:trt, data = r1, correlation = corAR1(form = ~1 |
id))
```

Linear Hypotheses:

| | Estimate | Std. Error | z value | Pr(> z) |
|--------------------|----------|------------|---------|----------|
| control slope == 0 | 0.09204 | 0.01210 | 7.608 | 2.80e-14 |
| milk slope == 0 | 0.05686 | 0.01196 | 4.756 | 1.98e-06 |
| calorie slope == 0 | 0.07617 | 0.01234 | 6.174 | 6.64e-10 |
| meat slope == 0 | 0.11374 | 0.01240 | 9.171 | < 2e-16 |

(Adjusted p values reported -- none method)

Here are some pairwise comparisons:

```
> K2
```

```
      [,1] [,2] [,3] [,4] [,5]
meat-control    0    0    0    0    1
meat-calorie    0    0    0   -1    1
meat-milk       0    0   -1    0    1
control-calorie 0    0    0   -1    0
control-milk    0    0   -1    0    0
calorie-milk    0    0   -1    1    0
```

```
> t2 <- glht(m1, linfct = K2)
```

```
> summary(t2)
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: gls(model = ravens ~ relmonth + relmonth:trt, data = r1, correlation = corAR1(form = ~1 | id))
```

Linear Hypotheses:

| | Estimate | Std. Error | z value | Pr(> z) |
|----------------------|----------|------------|---------|----------|
| meat-control == 0 | 0.02169 | 0.01635 | 1.327 | 0.54584 |
| meat-calorie == 0 | 0.03757 | 0.01614 | 2.328 | 0.09201 |
| meat-milk == 0 | 0.05687 | 0.01591 | 3.575 | 0.00195 |
| control-calorie == 0 | 0.01588 | 0.01627 | 0.976 | 0.76318 |
| control-milk == 0 | 0.03518 | 0.01602 | 2.196 | 0.12426 |
| calorie-milk == 0 | 0.01930 | 0.01580 | 1.222 | 0.61304 |

(Adjusted p values reported -- single-step method)