Designs to reduce the error

Session 7

MATH 80667A: Experimental Design and Statistical Methods
for Quantitative Research in Management
HEC Montréal

Outline

Blocking

Analysis of covariance

Blocking

Terminology for *nuisance*

Block

Source of variation, but
of no interest
known and controllable

Example

timing
lab technician
machine

Covariates

Explanatory measured **before** the experiment

Cannot be acted upon

Example

socioeconomic
variables
environmental
conditions

Noise factor

Under which setting is response least affected?

Example

temperature
processing

Why blocking?

Design experiment to reduce the effect of uncontrolled variations

In general, increases the power of the F test for treatment effects.

Group units in sets as alike as possible.

(Often) compare only treatments, so interactions are not included.

Assignment to treatment

Divide subjects within each block

Randomly allocate to treatment within block

(stratified sampling)

Block-treatment design

Without interaction,

$$\underset{\text{response}}{Y_{ij}} = \underset{\text{global mean}}{\mu} + \underset{\text{treatment}}{\alpha_i} + \underset{\text{blocking}}{\beta_j} + \underset{\text{error}}{\varepsilon_{ij}}$$

Compromise between

- reduced variability for residuals,
- loss of degrees of freedom due to estimation of β's.

Example: Resting metabolic rate

From Dean, Voss and Draguljić (2017), Example 10.4.1 (p. 311)

> experiment that was run to compare the effects of inpatient and outpatient protocols on the in-laboratory measurement of resting metabolic rate (RMR) in humans. A previous study had indicated measurements of RMR on elderly individuals to be 8% higher using an outpatient protocol than with an inpatient protocol. If the measurements depend on the protocol, then comparison of the results of studies conducted by different laboratories using different protocols would be difficult. The experimenters hoped to conclude that the effect on RMR of different protocols was negligible.

Fitting the complete block design

Load | Fit | Plot

```
url <- "https://edsm.rbind.io/files/data/resting_metabolic_rate.txt"
# transform integers to factors (categorical)
resting <- read.table(url, header = TRUE) |>
  dplyr::mutate(
    subject = factor(subject), #blocking factor
    protocol = factor(protocol), #experimental factor
    rate = rate/1000)
```

This is *de facto* a repeated measure design.

Fitting the complete block design

Load | **Fit** | Plot

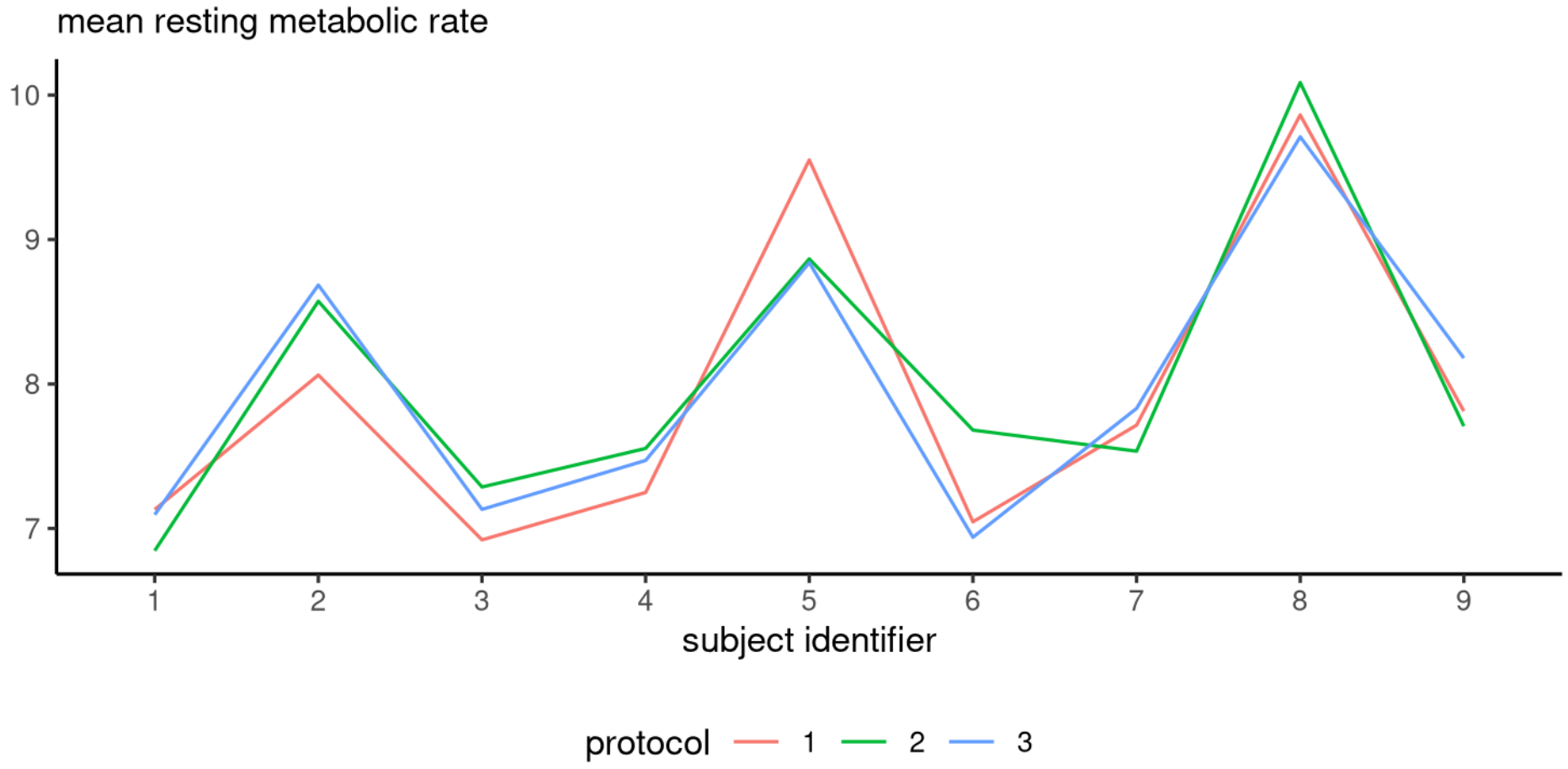
```r
# Force sum-to-zero parametrization for unordered factors
options(contrasts = c("contr.sum", "contr.poly"))
# Fit model with blocking
model_block <- aov(rate ~ subject + protocol, data = resting)
# One-way ANOVA (no blocking)
model_raw <- aov(rate ~ protocol, data = resting)
# anova(model_block)
# anova(model_raw)
```

Fitting the complete block design

Load Fit Plot

```
ggplot(data = resting,
       aes(x = subject,
           y = rate,
           group = protocol,
           color = protocol)) +
    geom_line() +
    labs(y = "mean resting\n metabolic rate") +
    theme_classic() +
    theme(legend.position = "bottom")
```

Interaction plot



Impact of blocking

ANOVA table (with blocking) | ANOVA table (without blocking)

Analysis of variance table - with blocking

	Degrees of freedom	Sum of squares	Mean square	F statistic	p-value
subject	8	23.12	2.89	37.42	0.000
protocol	2	0.04	0.02	0.23	0.795
Residuals	16	1.24	0.08		

Impact of blocking

ANOVA table (with blocking) | ANOVA table (without blocking)

Analysis of variance table - without blocking

	Degrees of freedom	Sum of squares	Mean square	F statistic	p-value
protocol	2	0.04	0.02	0.02	0.982
Residuals	24	24.35	1.01		

Analysis of covariance

IJLR: It's Just a Linear Regression...

All ANOVA models covered so far are linear regression model.

The latter says that

$$\underbrace{\mathsf{E}(Y_i)}_{\text{average response}} = \underbrace{\beta_0 + \beta_1 \mathrm{X}_{1i} + \cdots + \beta_p \mathrm{X}_{pi}}_{\text{linear (i.e., additive) combination of explanatories}}$$

In an ANOVA, the model matrix $\mathbf{x}$ simply includes columns with -1, 0 and 1 for group indicators that enforce sum-to-zero constraints.

What's in a model?

In experimental designs, the explanatories are

- experimental factors (categorical)
- continuous (dose-response)

Random assignment implies
no systematic difference between groups.

ANCOVA = Analysis of covariance

- Analysis of variance with added continuous covariate(s) to reduce experimental error (similar to blocking).
- These continuous covariates are typically concomitant variables (measured alongside response).
- Including them in the mean response (as slopes) can help reduce the experimental error (residual error).

Control to gain power!

Identify external sources of variations

- enhance balance of design (randomization)
- reduce mean squared error of residuals to increase power

These steps should in principle increase power **if** the variables used as control are correlated with the response.

Example

Abstract of van Stekelenburg et al. (2021)

> In three experiments with more than 1,500 U.S. adults who held false beliefs, participants first learned the value of scientific consensus and how to identify it. Subsequently, they read a news article with information about a scientific consensus opposing their beliefs. We found strong evidence that in the domain of genetically engineered food, this two-step communication strategy was more successful in correcting misperceptions than merely communicating scientific consensus.

Experiment 2: Genetically Engineered Food

We focus on a single experiment; preregistered exclusion criteria led to $n = 442$ total sample size (unbalanced design).

Three experimental conditions:

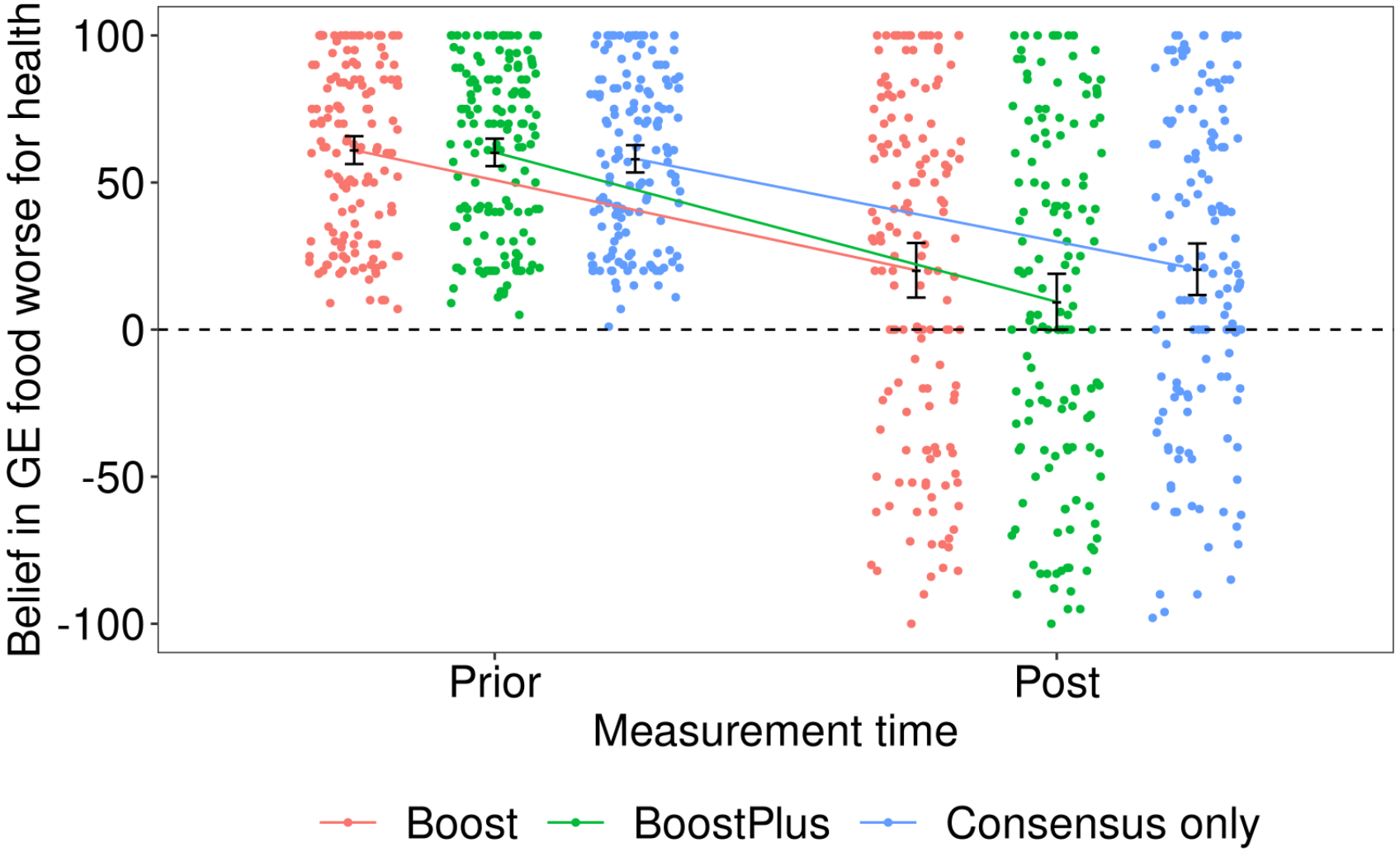
`Boost` | `Boost Plus` | Consensus only (`consensus`)

Model formulation

Use `post` as response variable and `prior` beliefs as a control variable in the analysis of covariance.

> their response was measured on a visual analogue scale ranging from –100 (I am 100% certain this is false) to 100 (I am 100% certain this is true) with 0 (I don't know) in the middle.

Plot of post vs prior response



Model formulation

Average for the $_r$th replication of the $_i$th experimental group is

$$\begin{aligned} \mathsf{E}(\texttt{post}_{ir}) &= \mu + \alpha_i \texttt{condition}_i + \beta \texttt{prior}_{ir}. \\ \mathsf{Va}(\texttt{post}_{ir}) &= \sigma^2 \end{aligned}$$

We assume that there is no interaction between `condition` and `prior`

- the slopes for `prior` are the same for each `condition` group.
- the effect of prior is linear

Contrasts of interest

1. Difference between average boosts (`Boost` and `BoostPlus`) and control (`consensus`)
2. Difference between `Boost` and `BoostPlus` (pairwise)

Inclusion of the `prior` score leads to increased precision for the mean (reduces variability).

Contrasts with ANCOVA

- The estimated marginal means will be based on detrended values $\neq$ group averages
- In the `emmeans` package, the average of the covariate is used as value.
- the difference between levels of `condition` are the same for any value of `prior` (parallel lines), but the uncertainty changes.

Multiple testing adjustments:

- Methods of Bonferroni (prespecified number of tests) and Scheffé (arbitrary contrasts) still apply
- Can't use Tukey anymore (adjusted means are not independent anymore).

Data analysis

Loading data | Scatterplot | Model | ANOVA

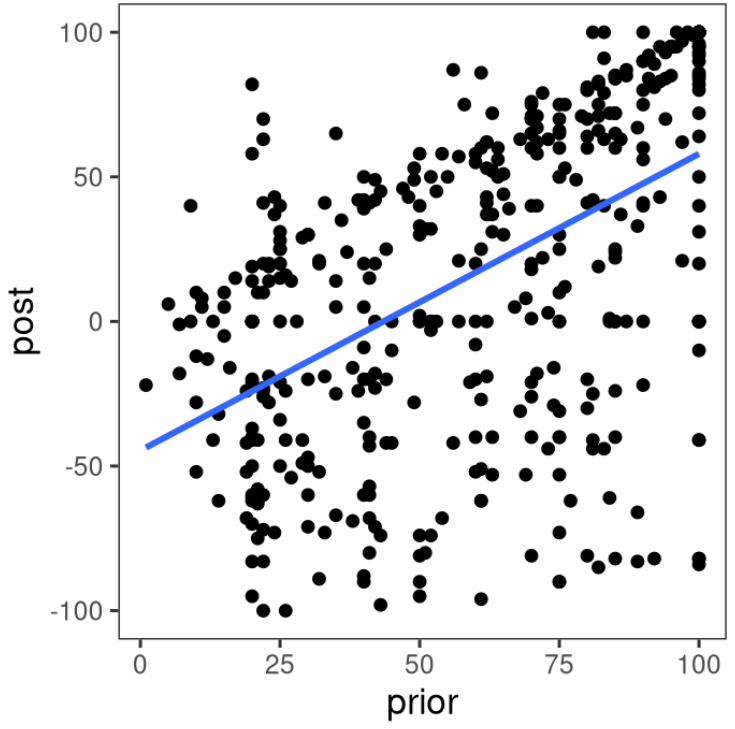
```
library(emmeans)
options(contrasts = c("contr.sum", "contr.poly"))
data(SSVB21_S2, package = "hecedsm")
# Check balance
with(SSVB21_S2, table(condition))
```

Data analysis

Loading data | **Scatterplot** | Model | ANOVA

```
library(ggplot2)
ggplot(data = SSVB21_S2,
       aes(x = prior, y = post)) +
       geom_point() +
       geom_smooth(method = "lm",
                   se = FALSE)
```

Strong correlation; note responses that achieve max of scale.



Data analysis

Loading data | Scatterplot | **Model** | ANOVA

```
# Check that the data are well randomized
car::Anova(lm(prior ~ condition, data = SSVB21_S2), type = 3)
# Fit linear model with continuous covariate
model1 <- lm(post ~ condition + prior, data = SSVB21_S2)
# Fit model without for comparison
model2 <- lm(post ~ condition, data = SSVB21_S2)
# Global test for differences - of NO INTEREST
car::Anova(model1, type = 3)
car::Anova(model2, type = 3)
```

Data analysis

Loading data | Scatterplot | Model | ANOVA

term	sum of squares	df	statistic	p-value
(Intercept)	166341	1	71.7	0.00
condition	**14107**	**2**	**3.0**	**0.05**
prior	385385	1	166.1	0.00
Residuals	1016461	438		

term	sum of squares	df	statistic	p-value
(Intercept)	123377	1	38.64	0.000
condition	**11680**	**2**	**1.83**	**0.162**
Residuals	1401846	439		

Data analysis

Contrasts | t-tests | Assumptions

```
emm1 <- emmeans(model1, specs = "condition")
# Note order: Boost, BoostPlus, consensus
emm2 <- emmeans(model2, specs = "condition")
# Not comparable: since one is detrended and the other isn't
contrast_list <- list(
   "boost vs control" = c(0.5,  0.5, -1),
   #av. boosts vs consensus
   "Boost vs BoostPlus" = c(1, -1,  0))
contrast(emm1,
         method = contrast_list,
         p.adjust = "holm")
```

Data analysis

Contrasts | t-tests | Assumptions

contrast	estimate	se	df	t stat	p-value
boost vs control	-8.37	4.88	438	-1.72	0.09
Boost vs BoostPlus	9.95	5.60	438	1.78	0.08

Contrasts with ANCOVA with `prior` (Holm-Bonferroni adjustment with $k = 2$ tests)

contrast	estimate	se	df	t stat	p-value
boost vs control	-5.71	5.71	439	-1.00	0.32
Boost vs BoostPlus	10.74	6.57	439	1.63	0.10

Contrasts for ANOVA (Holm-Bonferroni adjustment with $k = 2$ tests)

Data analysis

Contrasts | t-tests | Assumptions

```
# Test equality of variance
levene <- car::leveneTest(
   resid(model1) ~ condition,
   data = SSVB21_S2,
   center = 'mean')
# Equality of slopes (interaction)
car::Anova(lm(post ~ condition * prior,
              data = SSVB21_S2),
              model1, type = 3)
```

term	sum of squares	df	statistic	p-value
(Intercept)	165573	1	71.3	0.0
condition	4245	2	0.9	0.4
prior	382596	1	164.9	0.0
condition:prior	**3257**	**2**	**0.7**	**0.5**
Residuals	1016461	438		

Levene's test of equality of variance: F (2, 439) = 2.05 with a p-value of 0.13.

Model with interaction `condition*prior`. Slopes don't differ between condition.

The kitchen sink approach

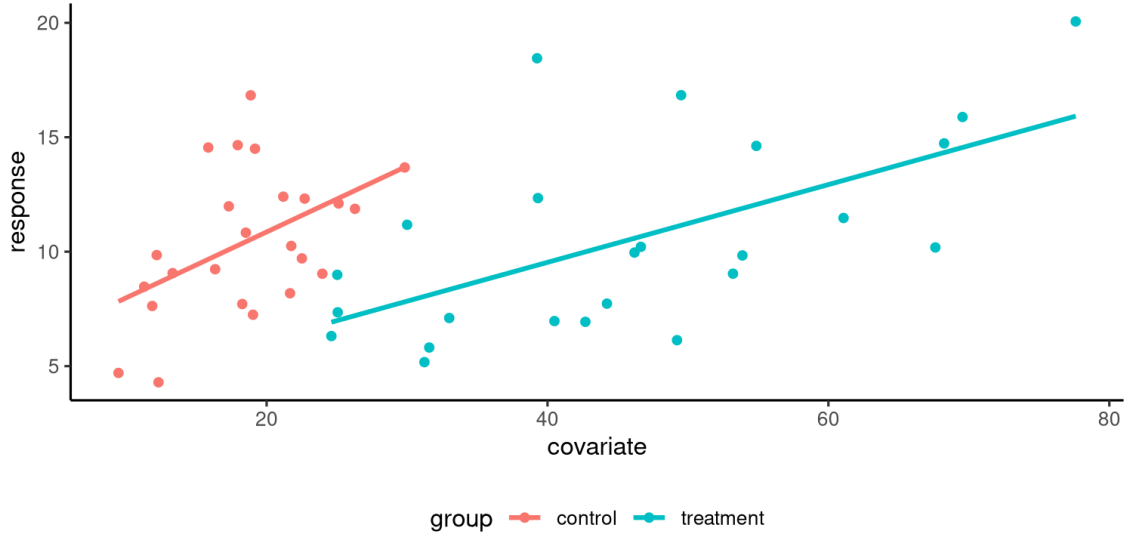
Should we control for more stuff?

NO! ANCOVA is a design to reduce error

- Randomization should ensure that there is no confounding
- No difference (on average) between group given a value of the covariate.
- If it isn't the case, adjustment matters

Equal trends

- If trends are different, meaningful comparisons (?)
- Differences between groups depend on value of the covariate



Due to lack of overlap, comparisons hazardous as they entail extrapolation one way or another.

Testing equal slope

Compare two nested models

- Null $\mathscr{H}_0$: model with covariate
- Alternative $\mathscr{H}_a$: model with interaction covariate * experimental factor

Use `anova` to compare the models in **R**.