

Analysis of Covariance

Analysis of covariance

- ANOVA: explanatory variables categorical (divide data into groups)
- traditionally, analysis of covariance has categorical x 's plus one numerical x ("covariate") to be adjusted for.
- `lm` handles this too.
- Simple example: two treatments (drugs) (a and b), with before and after scores.
- Does knowing before score and/or treatment help to predict after score?
- Is after score different by treatment/before score?

Data: treatment, before, after

```
a 5 20  
a 10 23  
a 12 30  
a 9 25  
a 23 34  
a 21 40  
a 14 27  
a 18 38  
a 6 24  
a 13 31  
b 7 19  
b 12 26  
b 27 33  
b 24 35  
b 18 30  
b 22 31  
b 26 34  
b 21 28  
b 14 23  
b 9 22
```

Packages

```
library(tidyverse)  
library(broom)  
library(marginaleffects)
```

the last of these for predictions.

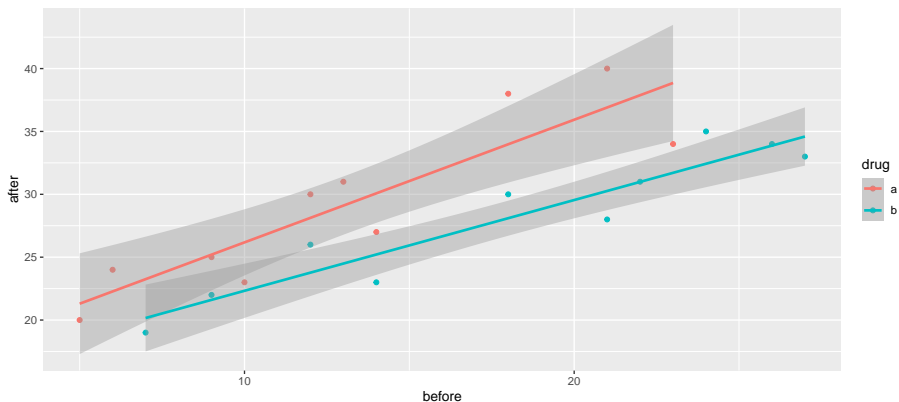
Read in data

```
url <- "http://ritsokiguess.site/datafiles/ancova.txt"
prepost <- read_delim(url, " ")
prepost
```

```
# A tibble: 20 x 3
  drug before after
<chr> <dbl> <dbl>
1 a      5     20
2 a     10     23
3 a     12     30
4 a      9     25
5 a     23     34
6 a     21     40
7 a     14     27
8 a     18     38
9 a      6     24
10 a     13     31
```

Making a plot

```
ggplot(prepost, aes(x = before, y = after, colour = drug)) +  
  geom_point() + geom_smooth(method = "lm")
```



Comments

- As before score goes up, after score goes up.
- Red points (drug A) generally above blue points (drug B), for comparable before score.
- Suggests before score effect *and* drug effect.

The means

```
prepost %>%  
  group_by(drug) %>%  
  summarize(  
    before_mean = mean(before),  
    after_mean = mean(after)  
  )
```

```
# A tibble: 2 x 3  
  drug before_mean after_mean  
  <chr>      <dbl>      <dbl>  
1 a          13.1         29.2  
2 b           18         28.1
```

- Mean “after” score slightly higher for treatment A.
- Mean “before” score much higher for treatment B.
- Greater *improvement* on treatment A.

Testing for interaction

```
prepost.1 <- lm(after ~ before * drug, data = prepost)
drop1(prepost.1, test = "F")
```

Single term deletions

Model:

after ~ before * drug

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			109.98	42.092		
before:drug	1	12.337	122.32	42.218	1.7948	0.1991

- Interaction not significant. Will remove later.

Predictions

Set up values to predict for, median and quartiles for before, the two drugs:

```
new <- datagrid(before = c(9.75, 14, 21.25),  
                drug = c("a", "b"), model = prepost.1)  
new
```

	before	drug	rowid
1	9.75	a	1
2	9.75	b	2
3	14.00	a	3
4	14.00	b	4
5	21.25	a	5
6	21.25	b	6

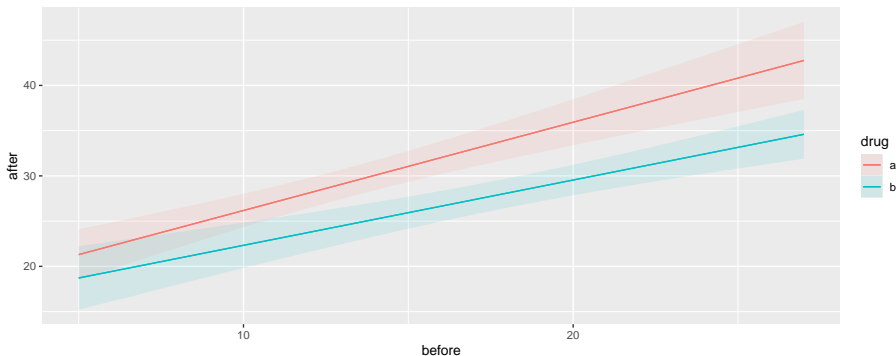
and then

```
cbind(predictions(prepost.1, newdata = new)) %>%  
  select(drug, before, estimate, conf.low, conf.high)
```

	drug	before	estimate	conf.low	conf.high
1	a	9.75	25.93250	24.05059	27.81442
2	b	9.75	22.14565	19.58681	24.70450
3	a	14.00	30.07784	28.43296	31.72271
4	b	14.00	25.21304	23.32649	27.09959
5	a	21.25	37.14929	34.32557	39.97300
6	b	21.25	30.44565	28.64373	32.24758

Predictions (with interaction included), plotted

```
plot_predictions(model = prepost.1,  
                 condition = c("before", "drug"))
```



Lines almost parallel, but not quite.

Taking out interaction

```
prepost.2 <- update(prepost.1, . ~ . - before:drug)
drop1(prepost.2, test = "F")
```

Single term deletions

Model:

after ~ before + drug

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			122.32	42.218			
before	1	540.18	662.50	74.006	75.074	1.211e-07	***
drug	1	115.31	237.63	53.499	16.025	0.0009209	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Take out non-significant interaction.
- before and drug strongly significant.
- Do predictions again and plot them.

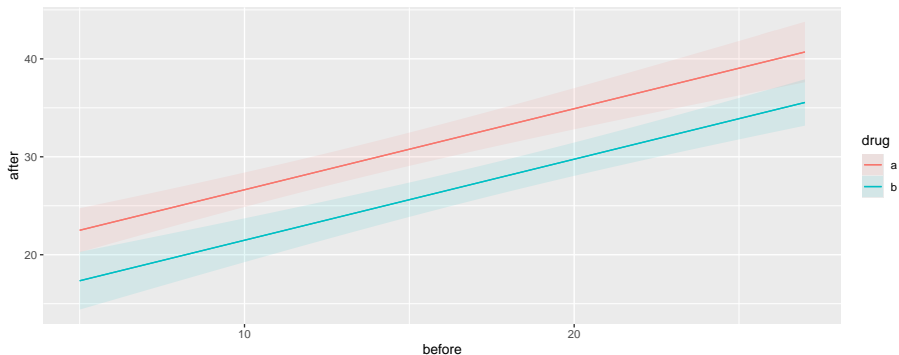
Predictions

```
cbind(predictions(prepost.2, newdata = new)) %>%  
  select(drug, before, estimate)
```

	drug	before	estimate
1	a	9.75	26.42794
2	b	9.75	21.27328
3	a	14.00	29.94473
4	b	14.00	24.79007
5	a	21.25	35.94397
6	b	21.25	30.78931

Plot of predicted values

```
plot_predictions(prepost.2, condition = c("before", "drug"))
```



This time the lines are *exactly* parallel. No-interaction model forces them to have the same slope.

Different look at model output

- `drop1(prepost.2)` tests for significant effect of before score and of drug, but doesn't help with interpretation.
- `summary(prepost.2)` views as regression with slopes:

```
summary(prepost.2)
```

Call:

```
lm(formula = after ~ before + drug, data = prepost)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6348	-2.5099	-0.2038	1.8871	4.7453

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	18.3600	1.5115	12.147	8.35e-10	***
before	0.8275	0.0955	8.665	1.21e-07	***
drugb	-5.1547	1.2876	-4.003	0.000921	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.682 on 17 degrees of freedom

Multiple R-squared: 0.817. Adjusted R-squared: 0.7955

Understanding those slopes

```
tidy(prepost.2)
```

```
# A tibble: 3 x 5
  term          estimate std.error statistic  p.value
<chr>         <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   18.4       1.51      12.1 8.35e-10
2 before        0.827     0.0955     8.66 1.21e- 7
3 drugb        -5.15     1.29     -4.00 9.21e- 4
```

- before ordinary numerical variable; drug categorical.
- lm uses first category drug_a as baseline.
- Intercept is prediction of after score for before score 0 and *drug A*.
- before slope is predicted change in after score when before score increases by 1 (usual slope)
- Slope for drug_b is *change* in predicted after score for being on drug B rather than drug A. Same for *any* before score (no interaction).

Summary

- ANCOVA model: fits different regression line for each group, predicting response from covariate.
- ANCOVA model with interaction between factor and covariate allows different slopes for each line.
- Sometimes those lines can cross over!
- If interaction not significant, take out. Lines then parallel.
- With parallel lines, groups have consistent effect regardless of value of covariate.