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.
- PROC GLM 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
```

SAS code

```
options linesize=75;

data drugs;
  infile "ancova.dat";
  input drug $ before after;

proc means;
  class drug;

proc glm;
  class drug;
  model after = drug before drug*before;
```

- Get means of before and after scores for each treatment.
- Make sure `drug` treated as categorical (“class”)
- Before score treated as numeric by default
- Interaction means “effect of before score on after score is different for each treatment”. Fit this first.

The means

The MEANS Procedure

drug	N Obs	Variable	N	Mean	Std Dev
a	10	before	10	13.1000000	6.0452001
		after	10	29.2000000	6.6131183
b	10	before	10	18.0000000	7.1492035
		after	10	28.1000000	5.4660569

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

Testing for interaction

The GLM Procedure

Dependent Variable: after

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	558.5668744	186.1889581	27.09	<.0001
Error	16	109.9831256	6.8739453		
Corrected Total	19	668.5500000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
drug	1	6.0500000	6.0500000	0.88	0.3621
before	1	540.1797947	540.1797947	78.58	<.0001
before*drug	1	12.3370798	12.3370798	1.79	0.1991

Source	DF	Type III SS	Mean Square	F Value	Pr > F
drug	1	1.2105592	1.2105592	0.18	0.6803
before	1	552.3578682	552.3578682	80.36	<.0001
before*drug	1	12.3370798	12.3370798	1.79	0.1991

Taking out interaction

- Take out non-significant interaction.
- Assuming linear dependence of after score on before score has *same slope* for both treatments (though possibly different intercept).
- Get predicted means for “after” score depending on `drug` and `before`.
- Also get means for treatments “adjusted” for before score.
- Code:

```
proc glm;
  class drug;
  model after = drug before;
  output out=z predict=pred;
  lsmeans drug;

proc print data=z;
```

Results

Dependent Variable: after

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	546.2297947	273.1148973	37.96	<.0001
Error	17	122.3202053	7.1953062		
Corrected Total	19	668.5500000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
drug	1	6.0500000	6.0500000	0.84	0.3720
before	1	540.1797947	540.1797947	75.07	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
drug	1	115.3059567	115.3059567	16.03	0.0009
before	1	540.1797947	540.1797947	75.07	<.0001

Interpreting the output

- Requires care!
- Model as a whole is significant.
- Type I SS says “is each variable significant when added *in order*: that is:
  - `drug` added to a model containing nothing (not sig)
  - `before` added to model containing only `drug` (sig)
- Not really what we want to know.
- Type III SS: “can I take this variable out of a model containing everything?” Answer in both cases no. Interpretation: once you allow for before score, there is a significant difference between treatments. (But if you don’t allow for before score, there isn’t.)

LS-means

Sample means for each treatment close:

drug	N Obs	Variable	N	Mean	Std Dev
a	10	after	10	29.2000000	6.6131183
b	10	after	10	28.1000000	5.4660569

"Least squares means": mean score for each treatment, after allowing for difference in before scores:

The GLM Procedure
Least Squares Means

drug	after LSMEAN
a	31.2273292
b	26.0726708

Treatment A noticeably (significantly) better than B, *once you allow for before score*.

Looking at the predictions

Some of them, arranged in before score order:

Obs	drug	before	pred
4	a	9	25.8073
3	a	12	28.2898
7	a	14	29.9447
8	a	18	33.2547
6	a	21	35.7371
20	b	9	20.6527
12	b	12	23.1351
19	b	14	24.7901
15	b	18	28.1000
18	b	21	30.5824

- Prediction for treatment A about 5 units higher than for treatment B at the same before score — same difference as between LSMEANS.
- Consistent because no interaction.
- If interaction had been included, A might be higher for some before scores and B higher for others: clouds interpretation.