Further topics in ANOVA and ANCOVA

Maarten Speekenbrink

Experimental Psychology
University College London

Statistics lecture 7

Outline

1. Further topics in ANOVA
   - Assumptions
   - Unequal sample size
   - Testing schemes and Sums of Squares types
2. Continuous and categorical predictors: ANCOVA
   - Homogeneity of slopes

ANOVA in the GLM

"Naively coding":

	Control	Family	Cog-Beh
λ_1	1	2	3

This implies that the mean of Family is exactly halfway between the mean of Control and and Cognitive-Behavioural.

ANOVA in the GLM

"Naively coding":

	Control	Family	Cog-Beh
λ_1	1	2	3

This implies that the mean of Family is exactly halfway between the mean of Control and and Cognitive-Behavioural.

A single predictor has to make very precise predictions about the differences between group means.

ANOVA in the GLM

Contrast coding:

	Control	Family	Cog-Beh
λ_1	-2	1	1
λ_2	0	-1	1

Using more than one contrast variable, "wrong" predictions of one variable can be corrected by the other variables.

E.g., λ_1 predicts the same mean for Family and Cognitive-Behavioural, but λ_2 allows those means to differ.

ANOVA in the GLM

Contrast coding:

	Control	Family	Cog-Beh
λ_1	-2	1	1
λ_2	0	-1	1

To account for any pattern of means for m groups, we need $m-1$ contrast variables.

The contrast coded predictors (X) derived from the contrast variables (λ) are treated like any other predictor.

Assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)

Predictions

Model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \\ &= \hat{Y}_i + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \end{aligned}$$

Predictions:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i}$$

Predictions

Model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \\ &= \hat{Y}_i + \epsilon_i & \epsilon_i &\sim N(0, \sigma) \end{aligned}$$

Predictions:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i}$$

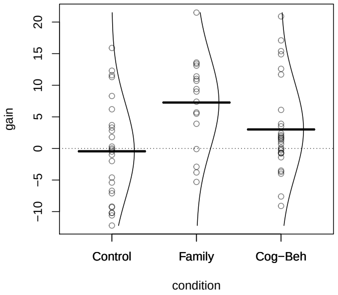
In ANOVA with m groups, there are m unique predictions. As predictors X_j have the identical values within a group, the predictions are also identical.

This implies that the assumptions must hold within each group.

Assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



In an ANOVA model, the dependent variable is Normally distributed within each group with identical variance. Can use Levene test for homogeneity of variance.

Unequal group sizes

"Classic" ANOVA designed for groups of equal size. But unequal sample sizes cannot always be avoided.

Consequences:

- Using GLM, no problem in estimation (unlike classic ANOVA)
- Contrast coded predictors (X) based on contrast codes (λ) no longer orthogonal
- Tests of predictors affected

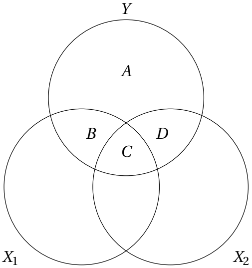
Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

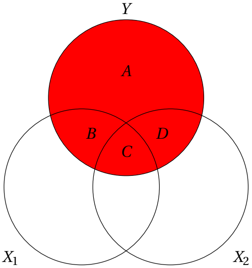
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Partitioning SSE

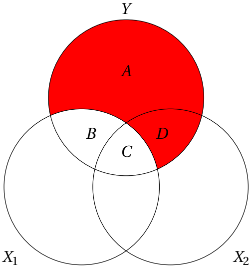
MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

$A + B + C + D = \text{SSE(C)}$



Partitioning SSE

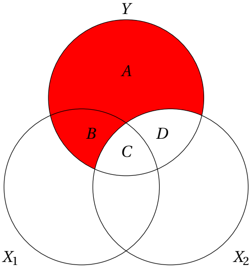
MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D =$ SSE(C)
$A + D =$ SSE(A1)

Partitioning SSE

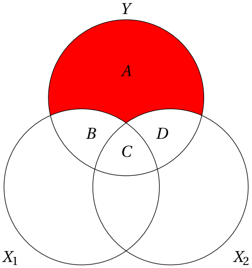
MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

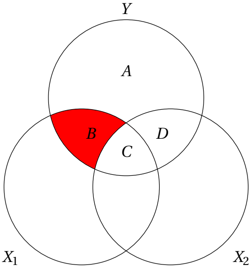
Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D =$ SSE(C)

$A + D =$ SSE(A1)

$A + B =$ SSE(A2)

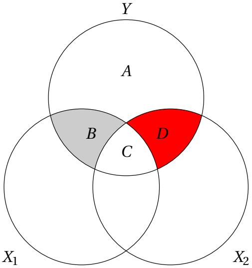
$A =$ SSE(A3)

Comparing A3 to A2:

$B = \text{SSR}(X_1)$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

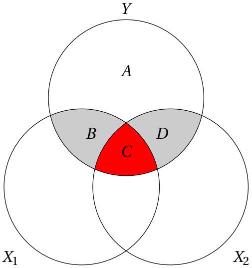


$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$
Comparing A3 to A1:
$D = \text{SSR}(X_2)$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$
Comparing A3 to A1:
$D = \text{SSR}(X_2)$
$C = ?$

Sums of Squares Types

- When predictors are redundant, different methods of model comparison give different results
- Determining the effect of a predictor by comparing a full model to one with the predictor removed is called Type III SS, or unique SS:
  - SSR terms do not add up to total SS (can be more or less!)
- Another commonly used strategy is called Type I SS, or sequential SS:
  - Start with an intercept only model (C)
  - Add a predictor (MODEL A1) and compute SSR
  - Add a second predictor (MODEL A3) and compute SSR
  - etc
- Other methods: Type II SS, Type IV SS, Type V SS, . . . (not covered here)

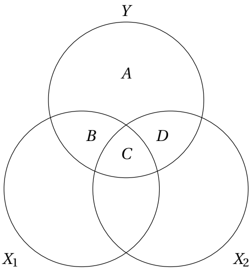
Sequential (Type I) SS

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

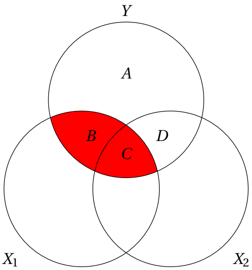
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Sequential (Type I) SS

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Comparing A1 to C:
$B + C = \mathrm{SSR}(X_1)$

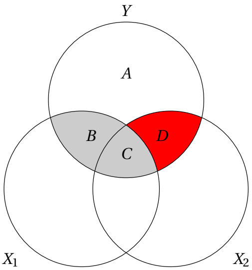
Sequential (Type I) SS

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Comparing A1 to C:

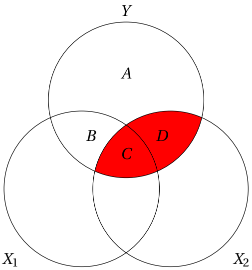
$B + C = \text{SSR}(X_1)$

Comparing A3 to A1:

$D = \text{SSR}(X_2)$

Sequential (Type I) SS

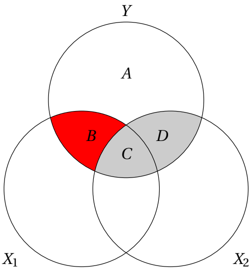
MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Comparing A2 to C:
$C + D = \mathrm{SSR}(X_2)$

Sequential (Type I) SS

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



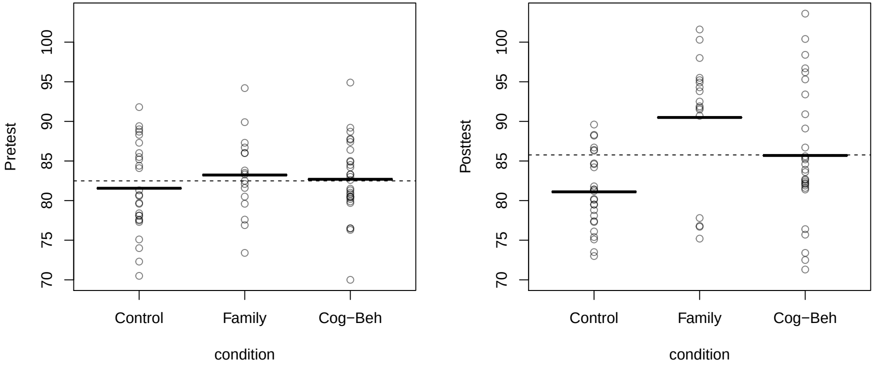
Comparing A2 to C:
$C + D = \text{SSR}(X_2)$
Comparing A3 to A2:
$B = \text{SSR}(X_1)$

ANCOVA

- When model has both categorical and continuous predictors, it's custom to speak of an ANalysis of COVAriance (ANCOVA)
- In GLM, this is nothing new...
- Reasons for including continuous predictors:
  - Statistically controlling for differences in experimental conditions
  - Because they are theoretically interesting

Treatment for anorexia

Everitt (1994) conducted a study on the effectiveness of treatments for anorexia. Patients were either given Cognitive/Behavioral therapy ($n = 29$), Family therapy ($n = 17$) or assigned to a control group ($n = 26$ on waiting list). Patients' weight was measured before (pre) and after (post) therapy.

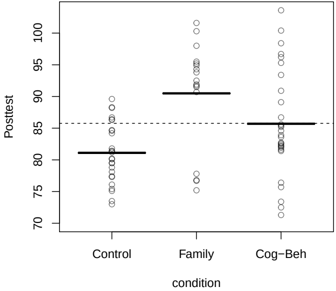


Differences after treatment

Helmert contrast:

	Control	Family	Cog/Beh
λ_1	-2	1	1
λ_2	0	-1	1

$$\texttt{post}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$



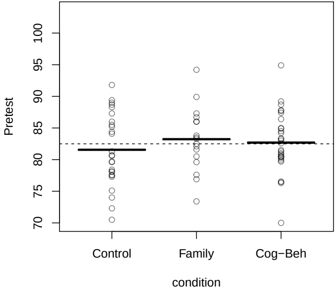
Source	b	SS	df	F	p
intercept	85.77	502417	1	9458.7	$< .001$
Treatment		918.99	2	8.65	$< .001$
X_1	2.33	790.24	1	14.88	$< .001$
X_2	-2.40	246.68	1	4.64	$.035$
error		3665.1	69		

Pre-existing differences?

Helmert contrast:

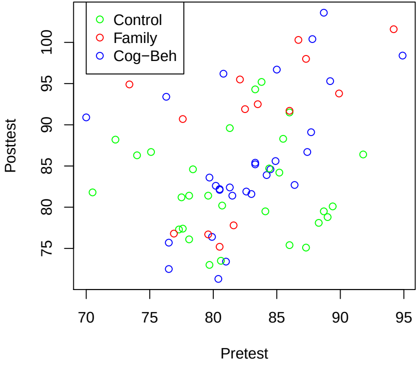
	Control	Family	Cog/Beh
λ_1	-2	1	1
λ_2	0	-1	1

$$\texttt{pre}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$



Source	b	SS	df	F	p
intercept	82.49	464793	1	17110	$< .001$
Treatment		32.57	2	0.60	0.55
X_1	0.47	31.81	1	1.17	0.28
X_2	-0.27	3.12	1	0.11	.74
error		1874.35	69		

Controlling for pretest differences (ANCOVA)



Controlling for pretest differences (ANCOVA)

$$\texttt{post}_i = \beta_0 + \beta_{\text{pre}}\texttt{pre}_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Source	b	SS	df	F	p
intercept	49.93	683.81	1	14.043	$< .001$
pretest	0.43	353.79	1	7.27	.009
Treatment		766.27	2	7.89	$< .001$
X_1	2.13	647.50	1	13.30	$< .001$
X_2	-2.28	222.78	1	4.58	.036
error		3311.3	68		

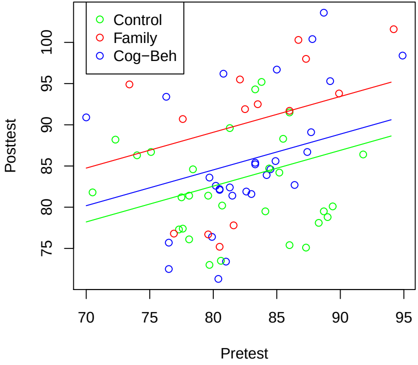
Controlling for pretest differences (ANCOVA)

$$\texttt{post}_i = \beta_0 + \beta_{\text{pre}}\texttt{pre}_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Source	b	SS	df	F	p
intercept	49.93	683.81	1	14.043	$< .001$
pretest	0.43	353.79	1	7.27	.009
Treatment		766.27	2	7.89	$< .001$
X_1	2.13	647.50	1	13.30	$< .001$
X_2	-2.28	222.78	1	4.58	.036
error		3311.3	68		

Assumes the effect of pretest is identical in each group!

Controlling for pretest differences (ANCOVA)



Slopes for contrast coded predictors in ANCOVA

Consider model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{m-1} X_{m-1,i} + b_z Z_i + e_i$$

where the X_j are contrast coded predictors and Z is a covariate. We can adjust the m group means for covariate differences by computing adjusted means

$$\overline{Y}^*_k = \overline{Y}_k - b_z(\overline{Z}_k - \overline{Z})$$

The slopes of the contrast coded predictors for the original variable Y are then

$$b_j = \frac{\sum_{k=1}^m \lambda_{j,k} \overline{Y}^*_k}{\sum_{k=1}^m \lambda_{j,k}^2}$$

alternatively, these can be computed directly as

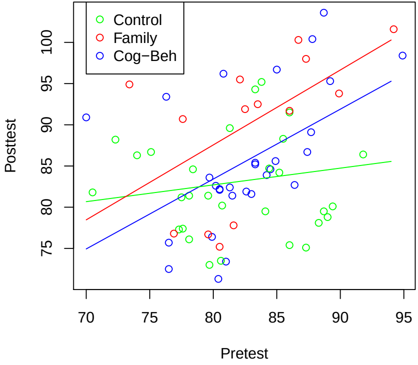
$$b_j = \frac{\sum_{k=1}^m \lambda_{j,k} \overline{Y}_k}{\sum_{k=1}^m \lambda_{j,k}^2} - b_z \frac{\sum_{k=1}^m \lambda_{j,k} \overline{Z}_k}{\sum_{k=1}^m \lambda_{j,k}^2}$$

Homogeneity of regression slopes?

$$\texttt{post}_i = \beta_0 + \beta_{\text{pre}}\texttt{pre}_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3(\texttt{pre} \times X_1)_i + \beta_4(\texttt{pre} \times X_2)_i + \epsilon_i$$

Source	b	SS	df	F	p
intercept	40.82	417.6	1	9.69	.003
pretest	0.54	503.42	1	11.68	.001
Treatment		400.54	2	4.65	.013
X_1	-25.62	390.33	1	9.06	.004
X_2	0.38	0.02	1	0.00	.098
Interaction		466.48	2	5.41	.007
pre $\times X_1$	0.34	460.13	1	10.68	.001
pre $\times X_2$	-0.03	0.94	1	0.02	.88
error		2844.78	66		

Homogeneity of regression slopes?



Homogeneity of regression slopes (2)?

(centered `pre` first)

$$\texttt{post}_i = \beta_0 + \beta_{\text{pre}}\texttt{pre0}_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3(\texttt{pre0} \times X_1)_i + \beta_4(\texttt{pre0} \times X_2)_i + \epsilon_i$$

Source	b	SS	df	F	p
intercept	85.40	488192	1	11326	$< .001$
pretest	0.54	503.42	1	11.68	$.001$
Treatment		785.11	2	9.11	$< .001$
X_1	2.20	691.99	1	16.05	$< .001$
X_2	-2.14	193.48	1	4.49	$.038$
Interaction		466.48	2	5.41	$.007$
pre $\times X_1$	0.34	460.13	1	10.68	$.001$
pre $\times X_2$	-0.03	0.94	1	0.02	$.88$
error		2844.78	66		

Analysis of difference scores

$$(\texttt{post} - \texttt{pre})_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Source	b	SS	df	F	p
intercept	3.27	732.07	1	12.92	$< .001$
Treatment		614.64	2	5.42	.006
X_1	1.86	504.97	1	8.91	.004
X_2	-2.13	194.29	1	3.43	.068
error		3910.74	69		

Analysis of difference scores

$$(\texttt{post} - \texttt{pre})_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Source	b	SS	df	F	p
intercept	3.27	732.07	1	12.92	$< .001$
Treatment		614.64	2	5.42	.006
X_1	1.86	504.97	1	8.91	.004
X_2	-2.13	194.29	1	3.43	.068
error		3910.74	69		

Usually not as powerful!

Equal to model

$$\texttt{post}_i = \beta_0 + 1 \times \texttt{pre}_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Further reading

This week:

- Judd, McClelland & Ryan, chapter 10

Next week:

- Judd, McClelland & Ryan, chapter 11