Categorical predictors (ANOVA)

Maarten Speekenbrink

Experimental Psychology
University College London

Statistics lecture 6

Outline

1. Models with one categorical predictor: Oneway ANOVA
   - Post-hoc tests
2. Two or more categorical predictors: Factorial ANOVA

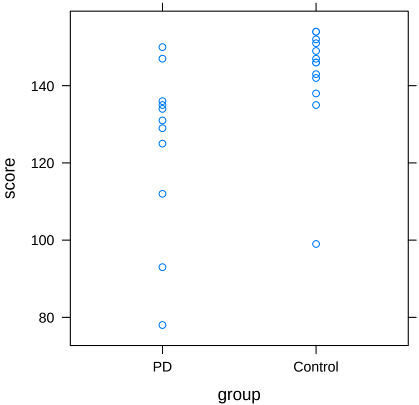
Learning and L-dopa

Parkinson's disease (PD) is a neurodegenerative disease involving a loss of dopamine producing cells. L-dopa alleviates many symptoms, but has been shown to impede learning. Speekenbrink et al. (2010) reported results of a study investigating learning by 11 PD patients (both on and off medication) and compared 13 healthy matched controls.

Participants performed the Weather Prediction Task, consisting of 200 trials in which they learned to predict the weather (Sunny or Rainy) from four cues ("tarot" cards). We'll look at performance (number of correct predictions) by PD patients on medication and controls

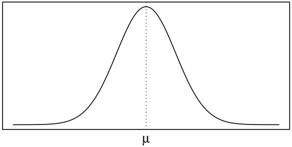
Speekenbrink et al. (2010). Models of probabilistic category learning in Parkinson's disease: Strategy use and the effects of L-dopa. *Journal of Mathematical Psychology*, *54*, 123–136.

Learning and L-dopa



Group differences (GLM)

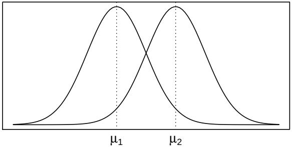
MODEL C: $Y_i = \mu + \epsilon_i$



Group differences (GLM)

MODEL C: $Y_i = \mu + \epsilon_i$

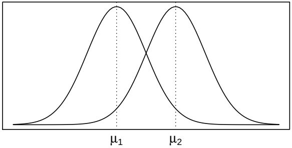
MODEL A: $Y_i = \begin{cases} \mu_1 + \epsilon_i & \text{(PD)} \\ \mu_2 + \epsilon_i & \text{(Control)} \end{cases}$



Group differences (GLM)

MODEL C: $Y_i = \mu + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \mu_1 + \epsilon_i & \text{(PD)} \\ \mu_2 + \epsilon_i & \text{(Control)} \end{cases}$



Use a linear model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

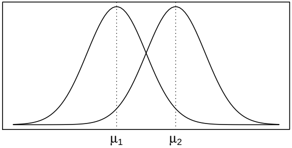
with

$$X_{1i} = \begin{cases} 0 & \text{(PD)} \\ 1 & \text{(Control)} \end{cases}$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \beta_1 X_{1i} + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 X_{1i} + \epsilon_i & \text{(Control)} \end{cases}$



Use a linear model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

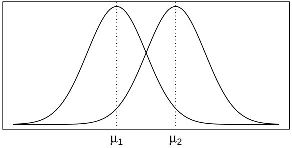
with

$$X_{1i} = \begin{cases} 0 & \text{(PD)} \\ 1 & \text{(Control)} \end{cases}$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \beta_1 \times 0 + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 \times 1 + \epsilon_i & \text{(Control)} \end{cases}$



Use a linear model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

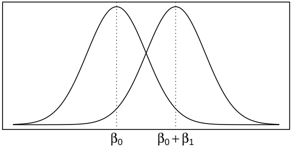
with

$$X_{1i} = \begin{cases} 0 & \text{(PD)} \\ 1 & \text{(Control)} \end{cases}$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(Control)} \end{cases}$



Use a linear model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

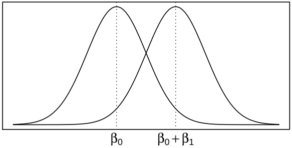
with

$$X_{1i} = \begin{cases} 0 & \text{(PD)} \\ 1 & \text{(Control)} \end{cases}$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(Control)} \end{cases}$



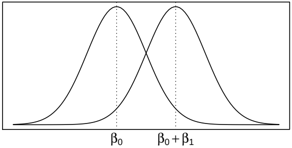
So

$$\mu_1 = \beta_0$$
$$\mu_2 = \beta_0 + \beta_1$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(Control)} \end{cases}$



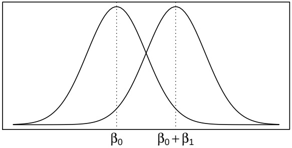
So

$$\mu_1 = \beta_0$$
$$\mu_2 = \mu_1 + \beta_1$$

Group differences (GLM)

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \begin{cases} \beta_0 + \epsilon_i & \text{(PD)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(Control)} \end{cases}$



So

$$\mu_1 = \beta_0$$
$$\mu_2 = \mu_1 + \beta_1$$

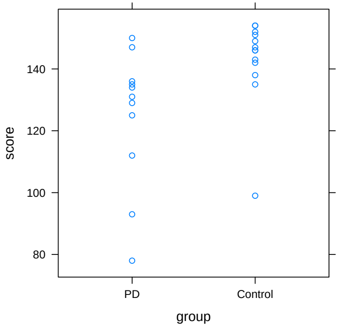
So $\beta_1 = \mu_2 - \mu_1$

Dummy coding

Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = \beta_0 + \epsilon_i$

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$

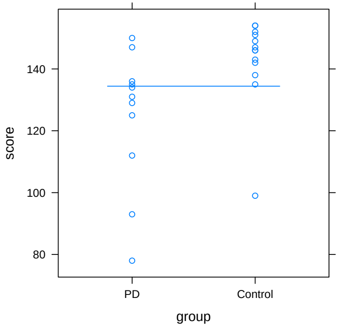


Dummy coding

Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$

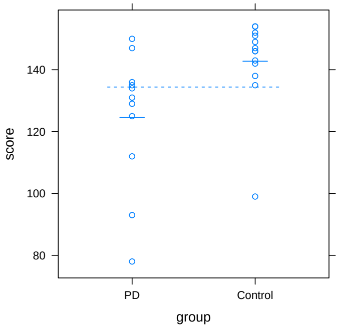


Dummy coding

Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = 124.55 + 18.22 \times X_{1i} + e_i$



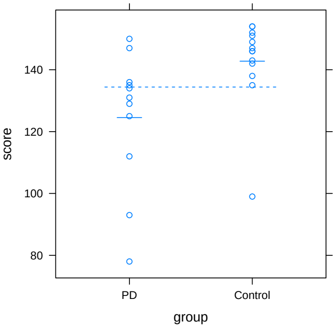
Dummy coding

Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = 124.55 + 18.22 \times X_{1i} + e_i$

$$Y_i = \begin{cases} 124.55 + 18.22 \times X_{1i} + e_i & \text{(PD)} \\ 124.55 + 18.22 \times X_{1i} + e_i & \text{(Control)} \end{cases}$$



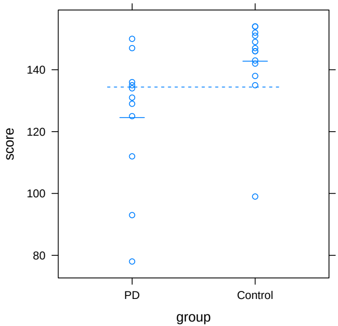
Dummy coding

Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = 124.55 + 18.22 \times X_{1i} + e_i$

$$Y_i = \begin{cases} 124.55 + 18.22 \times 0 + e_i & \text{(PD)} \\ 124.55 + 18.22 \times 1 + e_i & \text{(Control)} \end{cases}$$



Dummy coding

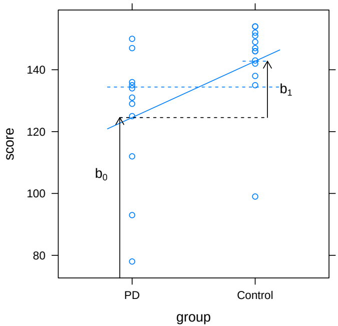
Y	group	X_1
150	PD	0
93	PD	0
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = 124.55 + 18.22 \times X_{1\,i} + e_i$

$$Y_i = \begin{cases} 124.55 + 18.22 \times 0 + e_i & \text{(PD)} \\ 124.55 + 18.22 \times 1 + e_i & \text{(Control)} \end{cases}$$

$b_0 = \overline{Y}_{\text{PD}} \qquad b_1 = \overline{Y}_{\text{Control}} - \overline{Y}_{\text{PD}}$

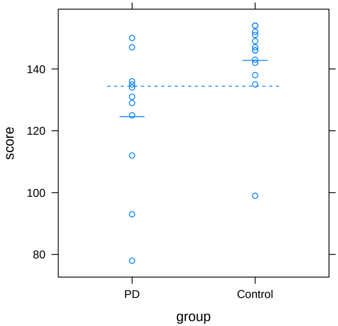


Effect coding

Y	group	X_1
150	PD	-1
93	PD	-1
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$



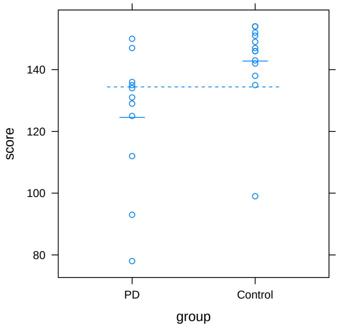
Effect coding

Y	group	X_1
150	PD	-1
93	PD	-1
⋮	⋮	⋮
154	control	1
147	control	1

MODEL C: $Y_i = 134.42 + e_i$

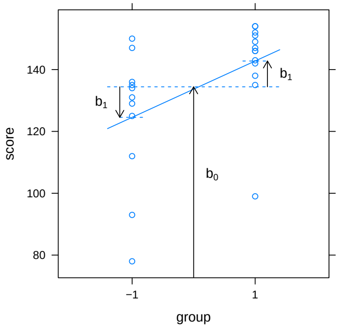
MODEL A: $Y_i = 133.66 + 9.11 \times X_{1i}$

$$Y_i = \begin{cases} 133.66 + 9.11 \times X_{1i} + e_i & \text{(PD)} \\ 133.66 + 9.11 \times X_{1i} + e_i & \text{(Control)} \end{cases}$$



Effect coding

Y	group	X_1
150	PD	-1
93	PD	-1
⋮	⋮	⋮
154	control	1
147	control	1



MODEL C: $Y_i = 134.42 + e_i$

MODEL A: $Y_i = 133.66 + 9.11 \times X_{1i}$

$$Y_i = \begin{cases} 133.66 - 9.11 + e_i & \text{(PD)} \\ 133.66 + 9.11 + e_i & \text{(Control)} \end{cases}$$

$$b_0 = \overline{Y}. = \frac{\overline{Y}_{\text{PD}} + \overline{Y}_{\text{Control}}}{2} \qquad b_1 = \overline{Y}_{\text{Control}} - \overline{Y}. \quad (= \overline{Y}. - \overline{Y}_{\text{PD}})$$

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

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

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

$$\begin{aligned} \text{SSE}(C) &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} (Y_i - b_0)^2 \\ &= \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \\ &= 9323.83 \end{aligned}$$

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

$$\begin{aligned} \text{SSE}(A) &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_{1i}))^2 \\ &= \sum_{i=1}^{n} (Y_i - \overline{Y}_{\text{group}})^2 \\ &= 7345.04 \end{aligned}$$

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

$$\begin{aligned}\text{SSE}(\text{C}) &= \textstyle\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^{n} (Y_i - b_0)^2 \\ &= \textstyle\sum_{i=1}^{n} (Y_i - \overline{Y})^2 \\ &= 9323.83\end{aligned}$$

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

$$\begin{aligned}\text{SSE}(\text{A}) &= \textstyle\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_{1i}))^2 \\ &= \textstyle\sum_{i=1}^{n} (Y_i - \overline{Y}_{\text{group}})^2 \\ &= 7345.04\end{aligned}$$

$$F = \frac{(\text{SSE}(\text{C}) - \text{SSE}(\text{A}))/(\text{PA} - \text{PC})}{\text{SSE}(\text{A})/(n - \text{PA})}$$

$$\text{PRE} = \frac{\text{SSE}(\text{C}) - \text{SSE}(\text{A})}{\text{SSE}(\text{C})}$$

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

$$\begin{aligned}\text{SSE(C)} &= \textstyle\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^{n}(Y_i - b_0)^2 \\ &= \textstyle\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \\ &= 9323.83\end{aligned}$$

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

$$\begin{aligned}\text{SSE(A)} &= \textstyle\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^{n}(Y_i - (b_0 + b_1 X_{1i}))^2 \\ &= \textstyle\sum_{i=1}^{n}(Y_i - \overline{Y}_{\text{group}})^2 \\ &= 7345.04\end{aligned}$$

$$F = \frac{(9323.83 - 7345.03)/(2-1)}{7345.03/(24-2)} = 5.93$$

$$\text{PRE} = \frac{9323.83 - 7345.03}{9323.83} = 0.21$$

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

$$\begin{aligned} \text{SSE(C)} &= \textstyle\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^n (Y_i - b_0)^2 \\ &= \textstyle\sum_{i=1}^n (Y_i - \overline{Y})^2 \\ &= 9323.83 \end{aligned}$$

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

$$\begin{aligned} \text{SSE(A)} &= \textstyle\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \textstyle\sum_{i=1}^n (Y_i - (b_0 + b_1 X_{1i}))^2 \\ &= \textstyle\sum_{i=1}^n (Y_i - \overline{Y}_{\text{group}})^2 \\ &= 7345.04 \end{aligned}$$

$$F = \frac{(9323.83 - 7345.03)/(2-1)}{7345.03/(24-2)} = 5.93$$

Critical value $F_{1,22;.05} = 4.30$

$$\text{PRE} = \frac{9323.83 - 7345.03}{9323.83} = 0.21$$

Testing difference between PD and controls

MODEL C: $Y_i = b_0 + e_i$

$$\begin{aligned}
\text{SSE(C)} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{n} (Y_i - b_0)^2 \\
&= \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \\
&= 9323.83
\end{aligned}$$

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

$$\begin{aligned}
\text{SSE(A)} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_{1i}))^2 \\
&= \sum_{i=1}^{n} (Y_i - \overline{Y}_{\text{group}})^2 \\
&= 7345.04
\end{aligned}$$

$$F = \frac{(9323.83 - 7345.03)/(2-1)}{7345.03/(24-2)} = 5.93$$

Critical value $F_{1,22;.05} = 4.30$

$$\text{PRE} = \frac{9323.83 - 7345.03}{9323.83} = 0.21$$

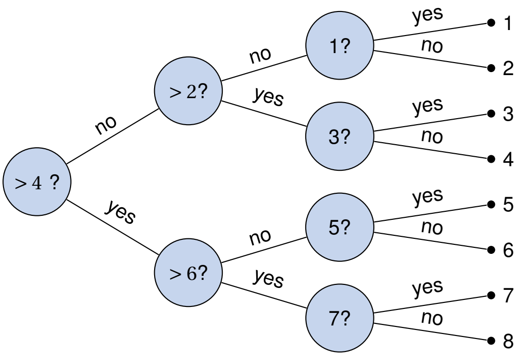
This model comparison is equivalent to an independent samples t-test ($t = \sqrt{F}$)

Coding for more than two groups

How many questions do you need to determine a number between 1 to 8?

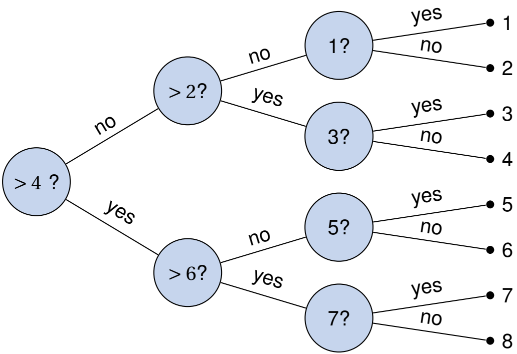
Coding for more than two groups

How many questions do you need to determine a number between 1 to 8?



Coding for more than two groups

How many questions do you need to determine a number between 1 to 8?



Need a total of 7 questions...

In general, need $m-1$ questions (variables) to choose between (code for) m levels of a variable

Coding variables

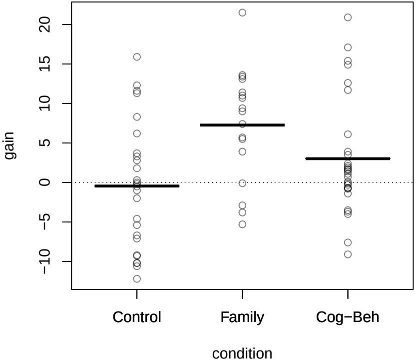
- Always: for a categorical variable (factor) with m levels (e.g., groups), we need $m-1$ contrast codes λ_j, $j = 1, \ldots, m-1$.
- Use the $m-1$ contrast codes (λ_j) to define $m-1$ contrast-coded predictors (X_j).

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

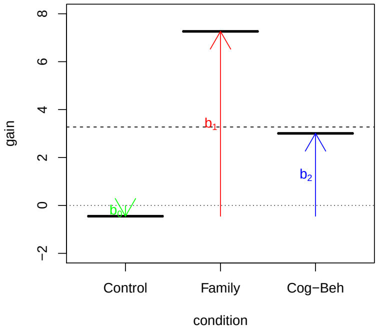
  - e.g., if i in group "Family", then $X_{1i} = 1$ and $X_{2i} = 0$.
- Can test effect of each X_j separately, but also collectively (omnibus test).

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 gain was measured after therapy.



Dummy coding



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

$$\begin{aligned} \texttt{gain}_i &= b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i \\ &= -0.45 + 7.72 X_{1i} + 3.46 X_{2i} + e_i \end{aligned}$$

effect	b	SS	df	F	p
intercept	-0.45	5.27	1	0.09	.76
condition		614.64	2	5.42	.006
X_1	7.72	611.78	1	10.79	.002
X_2	3.46	163.82	1	2.89	.09
error		3910.7	69		

Orthogonal contrast coding

- Ideally, predictors in a GLM are independent
- To ensure a set of contrast codes is orthogonal (independent), the following conditions should hold:
  1. $\sum_{k=1}^{m} \lambda_{j,k} = 0$
  2. $\sum_{k=1}^{m} \lambda_{j,k}\lambda_{l,k} = 0$, for all $j \neq l$

Orthogonal contrast coding

- Ideally, predictors in a GLM are independent
- To ensure a set of contrast codes is orthogonal (independent), the following conditions should hold:
  1. $\sum_{k=1}^{m} \lambda_{j,k} = 0$
  2. $\sum_{k=1}^{m} \lambda_{j,k}\lambda_{l,k} = 0$, for all $j \neq l$

E.g., dummy coding does *not* provide a set of orthogonal contrast codes

	Control	Family	Cog-Beh	sum
λ_1	0	1	0	1
λ_2	0	0	1	1
$\lambda_1 \times \lambda_2$	0	0	0	0

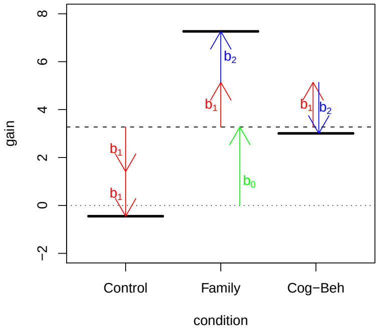
Orthogonal contrast coding

- Ideally, predictors in a GLM are independent
- To ensure a set of contrast codes is orthogonal (independent), the following conditions should hold:
  1. $\sum_{k=1}^{m} \lambda_{j,k} = 0$
  2. $\sum_{k=1}^{m} \lambda_{j,k}\lambda_{l,k} = 0$, for all $j \neq l$

but Helmert coding does:

	Control	Family	Cog-Beh	sum
λ_1	-2	1	1	0
λ_2	0	-1	1	0
$\lambda_1 \times \lambda_2$	0	-1	1	0

Helmert contrast



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

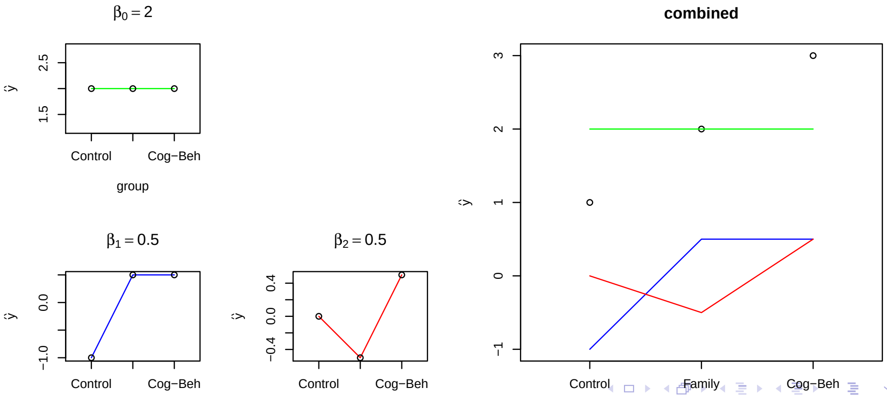
$$\begin{aligned}\texttt{gain}_i &= b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i \\ &= 3.27 + 1.86 X_{1i} - 2.13 X_{2i} + e_i\end{aligned}$$

effect	b	SS	df	F	p
intercept	3.27	732.07	1	12.92	$< .001$
condition		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		3910.7	69		

Contrast codes are flexible. . .

Helmert contrast

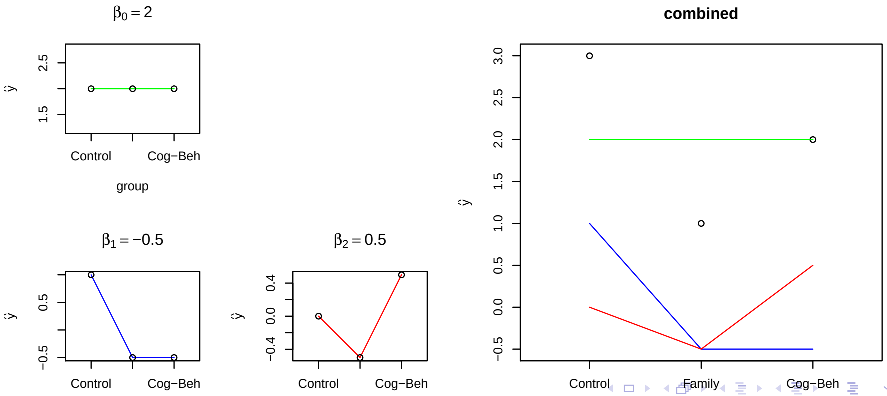
	Control	Family	Cog-Beh
λ_1	-2	1	1
λ_2	0	-1	1



Contrast codes are flexible. . .

Helmert contrast

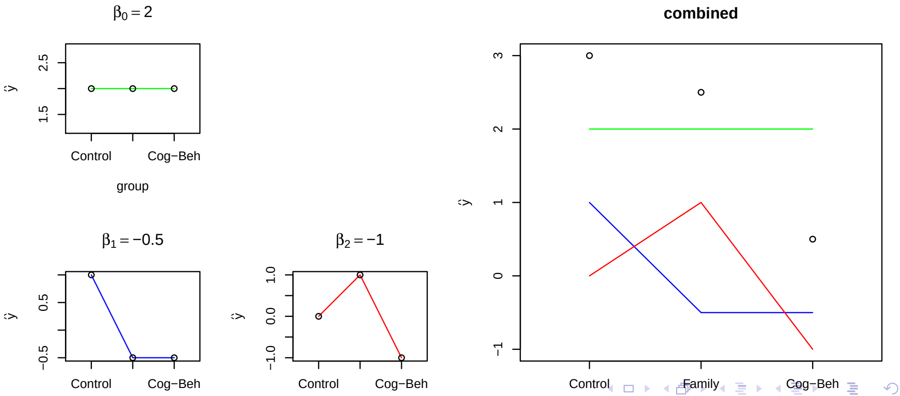
	Control	Family	Cog-Beh
λ_1	-2	1	1
λ_2	0	-1	1



Contrast codes are flexible. . .

Helmert contrast

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



Contrast codes and hypotheses about group means

- Many possible orthogonal contrast codes
- Can choose “standard” ones (like previous slide) but also specific ones to test interesting/sensible hypotheses
- Example: 4 weight loss regimes: none, diet, exercise, diet and and exercise. Questions:
  - Is doing something better than doing nothing?
  - Is there a difference between diet and exercise?
  - Is a combination better than either alone?

contrast	none	diet	exercise	both

Contrast codes and hypotheses about group means

- Many possible orthogonal contrast codes
- Can choose “standard” ones (like previous slide) but also specific ones to test interesting/sensible hypotheses
- Example: 4 weight loss regimes: none, diet, exercise, diet and and exercise. Questions:
  - Is doing something better than doing nothing?
  - Is there a difference between diet and exercise?
  - Is a combination better than either alone?

contrast	none	diet	exercise	both
λ_1	-3	1	1	1

Contrast codes and hypotheses about group means

- Many possible orthogonal contrast codes
- Can choose “standard” ones (like previous slide) but also specific ones to test interesting/sensible hypotheses
- Example: 4 weight loss regimes: none, diet, exercise, diet and and exercise. Questions:
  - Is doing something better than doing nothing?
  - Is there a difference between diet and exercise?
  - Is a combination better than either alone?

contrast	none	diet	exercise	both
λ_1	-3	1	1	1
λ_2	0	-1	1	0

Contrast codes and hypotheses about group means

- Many possible orthogonal contrast codes
- Can choose "standard" ones (like previous slide) but also specific ones to test interesting/sensible hypotheses
- Example: 4 weight loss regimes: none, diet, exercise, diet and and exercise. Questions:
    - Is doing something better than doing nothing?
    - Is there a difference between diet and exercise?
    - Is a combination better than either alone?

contrast	none	diet	exercise	both
λ_1	-3	1	1	1
λ_2	0	-1	1	0
λ_3	0	-1	-1	2

Interpretation of parameters

- Essentially identical to regression parameters, but can also interpret in terms of group means
- When $m-1$ orthogonal contrast variables are in the model, the estimated slopes can be written as

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

Interpretation of parameters

contrast	none	diet	exercise	both
λ_1	-3	1	1	1
λ_2	0	-1	1	0
λ_3	0	-1	-1	2

$$
\begin{aligned}
b_1 &= \frac{\sum_{k=1}^{m} \lambda_{1k} \overline{Y}_k}{\sum_{k=1}^{m} \lambda_{1k}^2} \\
&= \frac{-3 \times \overline{Y}_n + \overline{Y}_d + \overline{Y}_e + \overline{Y}_b}{(-3)^2 + 1^2 + 1^2 + 1^2} \\
&= \frac{\overline{Y}_d + \overline{Y}_e + \overline{Y}_b - 3 \times \overline{Y}_n}{12} \\
&= \frac{\frac{\overline{Y}_d + \overline{Y}_e + \overline{Y}_b}{3} - \overline{Y}_n}{4}
\end{aligned}
$$

Interpretation of parameters

contrast	none	diet	exercise	both
λ_1	-3	1	1	1
λ_2	0	-1	1	0
λ_3	0	-1	-1	2

$$
\begin{aligned}
b_2 &= \frac{\sum_{k=1}^{m} \lambda_{2k} \overline{Y}_k}{\sum_{k=1}^{m} \lambda_{2k}^2} \\
&= \frac{-1 \times \overline{Y}_d + \overline{Y}_e}{(-1)^2 + 1^2} \\
&= \frac{\overline{Y}_e - \overline{Y}_d}{2}
\end{aligned}
$$

Interpretation of parameters

contrast	none	diet	exercise	both
λ_1	-3	1	1	1
λ_2	0	-1	1	0
λ_3	0	-1	-1	2

$$
\begin{aligned}
b_3 &= \frac{\sum_{k=1}^{m} \lambda_{3k} \overline{Y}_k}{\sum_{k=1}^{m} \lambda_{3k}^2} \\
&= \frac{-1 \times \overline{Y}_d - 1 \times \overline{Y}_e + 2 \times \overline{Y}_b}{(-1)^2 + (-1)^2 + 2^2} \\
&= \frac{2\overline{Y}_b - \overline{Y}_d - \overline{Y}_e}{6} \\
&= \frac{\overline{Y}_b - \frac{\overline{Y}_d + \overline{Y}_e}{2}}{3}
\end{aligned}
$$

Hypothesis tests

Restated in terms of population means:

$$\beta_j = \frac{\sum_{k=1}^{m} \lambda_{jk} \mu_k}{\sum_{k=1}^{m} \lambda_{jk}^2}$$

Hypothesis tests

$$\beta_1 = \frac{\frac{\mu_d+\mu_e+\mu_b}{3} - \mu_n}{4}$$

If $\beta_1 = 0$:

$$0 = \frac{\frac{\mu_d+\mu_e+\mu_b}{3} - \mu_n}{4}$$

$$0 = \frac{\frac{\mu_d+\mu_e+\mu_b}{3}}{4} - \frac{\mu_n}{4}$$

$$\frac{\mu_n}{4} = \frac{\frac{\mu_d+\mu_e+\mu_b}{3}}{4}$$

$$\mu_n = \frac{\mu_d + \mu_e + \mu_b}{3}$$

Hypothesis tests

$$\beta_2 = \frac{\mu_e - \mu_d}{2}$$

If $\beta_2 = 0$:

$$0 = \frac{\mu_e}{2} - \frac{\mu_d}{2}$$
$$\mu_d = \mu_e$$

Hypothesis tests

$$\beta_3 = \frac{\mu_b - \frac{\mu_d + \mu_e}{2}}{3}$$

If $\beta_3 = 0$:

$$\begin{aligned} 0 &= \frac{\mu_b}{3} - \frac{\frac{\mu_d + \mu_e}{2}}{3} \\ \frac{\mu_d + \mu_e}{2} &= \mu_b \end{aligned}$$

Post hoc tests

- After significant omnibus test, often want to know which groups differ
- Using orthogonal contrast codes, we can test for $m-1$ pairs of groups
- But there are $\frac{m(m-1)}{2}$ possible paired comparisons
- Can form more (nonorthogonal) contrast codes to test for other pairs
  - Inflation of Type I error
  - Bonferroni correction, or Scheffé adjusted critical value

$$(m-1)F_{m-1,n-\text{PA};\alpha}$$

  where m is the total number of groups.
- Use specific post-hoc test procedure to test for all possible pairs
  - Tukey HSD test (or Tukey-Kramer for unequal cell size)

Two categorical predictors

- Weight loss regimes example:
  - Two dichotomous variables: diet (yes, no) and exercise (yes, no)
  - All 4 combinations possible
  - Can analyse effects separately
- Need to code both categorical variables

Two categorical predictors

Dummy coding:

- Diet: yes ($\lambda_1 = 1$), no ($\lambda_1 = 0$)
- Exercise: yes ($\lambda_2 = 1$), no ($\lambda_2 = 0$)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$$= \begin{cases} \beta_0 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(diet)} \\ \beta_0 + \beta_2 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	0	1	0	1
λ_2	0	0	1	1

Two categorical predictors

Dummy coding:

- Diet: yes ($\lambda_1 = 1$), no ($\lambda_1 = 0$)
- Exercise: yes ($\lambda_2 = 1$), no ($\lambda_2 = 0$)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$$= \begin{cases} \beta_0 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(diet)} \\ \beta_0 + \beta_2 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	0	1	0	1
λ_2	0	0	1	1

Assumes combined effect of diet and exercise is sum of its parts

Two categorical predictors

Dummy coding:

- Diet: yes ($\lambda_1 = 1$), no ($\lambda_1 = 0$)
- Exercise: yes ($\lambda_2 = 1$), no ($\lambda_2 = 0$)
- Interaction: both ($\lambda_3 = 1$) other ($\lambda_3 = 0$)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

$$= \begin{cases} \beta_0 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(diet)} \\ \beta_0 + \beta_2 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	0	1	0	1
λ_2	0	0	1	1
$\lambda_3 (= \lambda_1 \lambda_2)$	0	0	0	1

Combined effect of diet and exercise not necessarily sum of its parts

Two categorical predictors

Dummy coding:

- Diet: yes ($\lambda_1 = 1$), no ($\lambda_1 = 0$)
- Exercise: yes ($\lambda_2 = 1$), no ($\lambda_2 = 0$)
- Interaction: both ($\lambda_3 = 1$) other ($\lambda_3 = 0$)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

$$= \begin{cases} \beta_0 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 + \epsilon_i & \text{(diet)} \\ \beta_0 + \beta_2 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	0	1	0	1
λ_2	0	0	1	1
$\lambda_3 (= \lambda_1 \lambda_2)$	0	0	0	1

$$\beta_0 = \mu_\text{n}$$
$$\beta_1 = \mu_\text{d} - \mu_\text{n}$$
$$\beta_2 = \mu_\text{e} - \mu_\text{n}$$
$$\beta_3 = \mu_\text{b} - \mu_\text{d} - \mu_\text{e} + \mu_\text{n}$$

Two categorical predictors

Orthogonal contrast

- Diet: $\lambda_1 = \begin{cases} 1 & \text{(yes)} \\ -1 & \text{(no)} \end{cases}$
- Exercise: $\lambda_2 = \begin{cases} 1 & \text{(yes)} \\ -1 & \text{(no)} \end{cases}$
- Interaction: $\lambda_3 = \lambda_1 \times \lambda_2$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

$$= \begin{cases} \beta_0 - \beta_1 - \beta_2 + \beta_3 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 - \beta_2 - \beta_3 + \epsilon_i & \text{(diet)} \\ \beta_0 - \beta_1 + \beta_2 - \beta_3 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	-1	1	-1	1
λ_2	-1	-1	1	1
$\lambda_3 (= \lambda_1 \lambda_2)$	1	-1	-1	1

Two categorical predictors

Orthogonal contrast

- Diet: $\lambda_1 = \begin{cases} 1 & \text{(yes)} \\ -1 & \text{(no)} \end{cases}$
- Exercise: $\lambda_2 = \begin{cases} 1 & \text{(yes)} \\ -1 & \text{(no)} \end{cases}$
- Interaction: $\lambda_3 = \lambda_1 \times \lambda_2$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

$$= \begin{cases} \beta_0 - \beta_1 - \beta_2 + \beta_3 + \epsilon_i & \text{(none)} \\ \beta_0 + \beta_1 - \beta_2 - \beta_3 + \epsilon_i & \text{(diet)} \\ \beta_0 - \beta_1 + \beta_2 - \beta_3 + \epsilon_i & \text{(exerc.)} \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i & \text{(both)} \end{cases}$$

	no exercise		exercise	
contrast	no diet	diet	no diet	diet
λ_1	-1	1	-1	1
λ_2	-1	-1	1	1
$\lambda_3 (= \lambda_1 \lambda_2)$	1	-1	-1	1

$$\beta_0 = \frac{\mu_\text{n} + \mu_\text{d} + \mu_\text{e} + \mu_\text{b}}{4} = \mu_{..}$$

$$\beta_1 = \frac{\mu_\text{d} + \mu_\text{b} - \mu_\text{n} - \mu_\text{e}}{4} = \frac{\frac{\mu_\text{d} + \mu_\text{b}}{2} - \frac{\mu_\text{n} + \mu_\text{e}}{2}}{2}$$

$$\beta_2 = \frac{\mu_\text{e} + \mu_\text{b} - \mu_\text{n} - \mu_\text{d}}{4} = \frac{\frac{\mu_\text{e} + \mu_\text{b}}{2} - \frac{\mu_\text{n} + \mu_\text{d}}{2}}{2}$$

$$\beta_3 = \frac{\mu_\text{n} + \mu_\text{b} - \mu_\text{d} - \mu_\text{e}}{4} = \frac{\frac{\mu_\text{n} + \mu_\text{b}}{2} - \frac{\mu_\text{d} + \mu_\text{e}}{2}}{2}$$

ANOVA tables

Example:

- Drug (placebo, A, B)
- Therapy (no, yes)

	no therapy			psychotherapy		
contrast	placebo	A	B	placebo	A	B
λ_1	-2	1	1	-2	1	1
λ_2	0	1	-1	0	1	-1
λ_3	-1	-1	-1	1	1	1
λ_4	2	-1	-1	-2	1	1
λ_5	0	-1	1	0	1	-1

ANOVA tables ("Classic")

Example:

- Drug (placebo, A, B)
- Therapy (no, yes)

Source	SS	df	MS	F	p
Model (between)	960	5	192.0	46.1	.0001
Drug	453	2	226.5	53.9	.0001
Therapy	450	1	450	108.0	.0001
Drug × Therapy	57	2	28.5	6.8	.01
Error (within)	50	12	4.2		
Total	1010	17			

ANOVA tables (Better)

Example:

- Drug (placebo, A, B)
- Therapy (no, yes)

Source	b	SS	df	MS	F	p	PRE
Model (between)		960	5	192.0	46.1	.0001	.95
Drug		453	2	226.5	53.9	.0001	.90
X_1	3.5	441	1	441	105.8	.0001	.90
X_2	1.0	12	1	12	2.9	.11	.19
Therapy		450	1	450	108.0	.0001	.90
X_3	5.0	450	1	450	108.0	.0001	.90
Drug × Therapy		57	2	28.5	6.8	.01	.53
X_4	0.5	9	1	9	2.2	.16	.15
X_5	2.0	48	1	48	11.5	.005	.49
Error (within)		50	12	4.2			
Total		1010	17				

More than two categorical variables

- Basically the same as for two categorical predictors
  - Contrast codes to code for levels of each variable
  - Interaction effects by multiplying contrast variables
- Now also higher order interactions...
- Testing identical to before

Further Reading

This week:

- Judd, McClelland & Ryan, chapters 8 & 9

Next week:

- Judd, McClelland & Ryan, chapter 10