announcements

- Problem Set 2 is due right about now
- Problem Set 2's answer key will be available tomorrow(ish)
- Problem Set 3 will be assigned on Wednesday and due next Monday

1

categorical predictors (part 3: ANOVA)

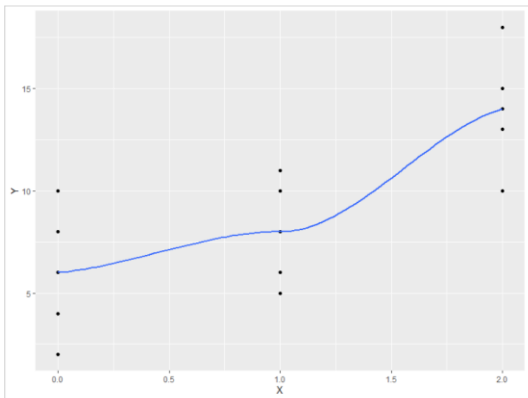
February 5, 2024

2

getting ready for >2 groups

- the first lesson to learn: creating one *X* ***won't*** suffice
- let's try
- I have a data set with 3 groups, and I assigned values of X = 1, 2, and 3 to them, respectively

3



4

using one X for >2 groups will usually induce nonlinearity

- we will ***need*** $m - 1$ new variables to numerically code our m groups
- the numbers we choose to indicate group membership will depend on what we want our slopes to tell us (among other constraints)

5

we need more Xs

- if we have m groups, we ***need*** $m - 1$ predictors (Xs), no more, no less
- the predictors could (should?) be contrast codes

$$\sum \lambda_k = 0$$

- we want contrasts to be *orthogonal* (independent)

6

orthogonality

- defined mathematically (for contrast codes; will not work for dummy codes)

$$\sum \lambda_{1k}\lambda_{2k} = 0$$

- what?!
- let's look at some *X*s for a three-group design and check for orthogonality

7

three (*m* = 3) groups → two *X*s, with the value of λ assigned to each

group	λ_1	λ_2	$\lambda_1\lambda_2$
A	1	0	
B	0	1	
C	-1	-1	

8

three (*m* = 3) groups → two *X*s, with the value of λ assigned to each

group	λ_1	λ_2	$\lambda_1\lambda_2$
A	2	0	
B	-1	1	
C	-1	-1	

9

how to choose contrast codes?

- initially, let predictions dictate what's of interest
- then, let mathematical constraints fill in the rest, as needed

10

a concrete example

- in a (hypothetical) study designed to test different memory strategies, participants were randomly assigned to learn a list of words using one of three strategies: form a mental image; find a rhyme; or just to study the list; after study & a delay, they're given a recall test
- the data are in today's script; the main results are

```
  group      M
1 control    6
2 image     12
3 rhyme     10
```

11

what predictions might we want to test?

1. is using a strategy of any kind better than not using one?
2. which strategy works better?
3. is imagery better than nothing?
4. is rhyming better than nothing?

   - each prediction corresponds to a contrast we could do
   - we can only include two (no more, no less), and we'd like them to be orthogonal

12

let's test predictions 1 & 2

assign λs using the "method of subsets"

- count the groups involved (this will be the denominator of the λs)
- count the number of groups on each "team" (these will be numerators of weights on the other team)
- assign + and – to each team, with the + going to the team expected to score higher
- for any "team" with more than one group, repeat

13

group	λ_1	λ_2	$\lambda_1\lambda_2$
image			
rhyme			
control			

14

what do we get?

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.3333     0.7045  13.248 2.49e-13
X1            5.0000     1.4944   3.346  0.00242
X2            2.0000     1.7256   1.159  0.25660
```

15

what are the predicted scores?

$$\widehat{memory} = b_0 + b_1 X_1 + b_2 X_2$$

$$\widehat{memory}_{\text{image}} = 9.33 + 5\left(\frac{1}{3}\right) + 2\left(\frac{1}{2}\right)$$

$$\widehat{memory}_{\text{image}} = 12$$

16

what is Model C?

- Model A

$$\widehat{memory} = b_0 + b_1 X_1 + b_2 X_2$$

- Model C for X_1

$$\widehat{memory} = b_0 + 0 X_1 + b_2 X_2$$

- Model C for X_2

$$\widehat{memory} = b_0 + b_1 X_1 + 0 X_2$$

17

let's focus on b_1

$$H_0\colon \beta_1 = 0$$

- this could be rewritten as

$$H_0\colon \frac{1}{3}\mu_{image} + \frac{1}{3}\mu_{rhyme} + \left(-\frac{2}{3}\right)\mu_{control} = 0$$

$$H_0\colon \frac{1}{3}\mu_{image} + \frac{1}{3}\mu_{rhyme} = \frac{2}{3}\mu_{control}$$

$$H_0\colon \mu_{image} + \mu_{rhyme} = 2\mu_{control}$$

$$H_0\colon \frac{\mu_{image} + \mu_{rhyme}}{2} = \mu_{control}$$

18

a different Model C

- Model A

$$\widehat{memory} = b_0 + b_1X_1 + b_2X_2$$

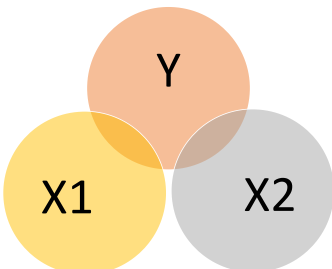
- Model C

$$\widehat{memory} = b_0 + 0X_1 + 0X_2$$

- this is Model C if one does a typical ANOVA
- crucially, $PA - PC > 1$; this is undesirable for drawing specific conclusions

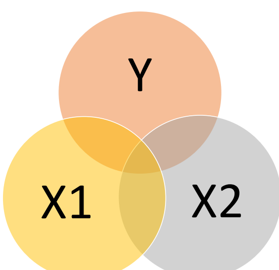
19

why is orthogonality important? it forces tolerance = 1



20

nonorthogonality → tolerance < 1



21

what happens if contrasts are not orthogonal?

group	λ_1	λ_2
image	+1/3	+½
rhyme	+1/3	0
control	-2/3	-½

22

some parameters are not what they're expected to be ...

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.000      1.349   5.930 2.54e-06
badX1          6.000      1.726   3.477  0.00173
badX2          4.000      3.451   1.159  0.25660
```

23

what happens if we use dummy codes?

group	λ_1	λ_2
image	1	0
rhyme	0	1
control	0	0

24

what is the intercept?
what are the slopes?

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.000	1.220	4.917	3.8e-05
D1	6.000	1.726	3.477	0.00173
D2	4.000	1.726	2.318	0.02827

Note that even though dummy codes are not orthogonal (for linear algebra reasons that I don't fully understand), the slopes are what we want them to be

25