categorical predictors
(part 4: ANOVA)

February 7, 2024

1

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 main results are

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

2

what happens if we use dummy codes?

group	D_1	D_2
image		
rhyme		
control		

3

what is the intercept?
what are the slopes?

```
> summary(modelDummy)
            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
```

4

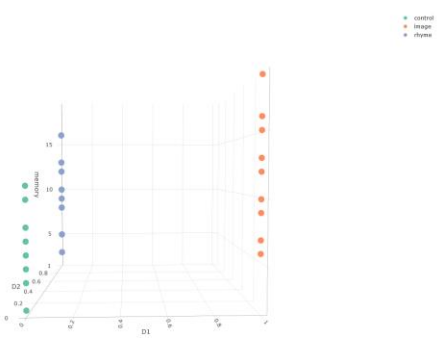
why do dummy code slopes "ignore" a group?

group	D_1	D_2
image	1	0
rhyme	0	1
control	0	0

slope is M_{rhyme} - $M_{control}$

slope is M_{image} - $M_{control}$

5



6

R defaults to dummy codes

- the summary of `lm(memory ~ group)` is

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

summary(modelDummy)
            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
```

7

"model F" is the same regardless of coding

	D1	D2
I	1	0
R	0	1
C	0	0

F(2, 27) = 6.3

	strange1	strange2
I	2	3
R	7	11
C	1	4

F(2, 27) = 6.3

Why?
Whatever values are assigned to groups, model *F* is based on

Model A: $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$
Model C: $\hat{Y} = b_0 + 0\ X_1 + 0\ X_2$

8

pairwise comparisons

- *very* often the contrasts of interest in a one-factor study are simply comparisons between all possible pairs of groups
- this is clunky to execute using orthogonal contrasts
- it requires redoing analyses multiple times and ignoring some results
- the `pairwise.t.test` function is handy for executing only pairwise comparisons
- it comes with an argument that allows one to control Type I errors ...

9

controlling Type I error rates

- if each hypothesis test one does comes with a .05 error rate ...
- ... doing many hypothesis tests leads to a *familywise error rate* of > .05
- *FWER* = the probability of at least one Type I error in a *family* of contrasts
- important digression: what is a family?
  - is it all the hypothesis tests you do in your career?
  - is it all the hypothesis tests you do in one manuscript?
  - is it all the hypothesis tests you do for one model?

10

controlling Type I error rates

- use the Bonferroni (or Dunn-Bonferroni) procedure if your contrasts are *planned*
- if c = the number of contrasts you'll perform
- use an alpha level of $.05/c$ to decide significance
- e.g., if you're doing 5 contrasts

$$\alpha = {}^{.05}/_{5} = .01$$

- alternatively, take each p and multiply it by c, and then compare to α (probably .05)

11

controlling the "false discovery rate"

- the Bonferroni procedure is designed to minimize the probability of at least one Type I error occurring
- other procedures are designed to minimize the proportion of Type I errors that occur (the "false discovery rate")
- a simple one is the Benjamini-Hochberg procedure

12

BH procedure

- for any family of contrasts
  - find *p*-values for contrasts
  - rank the *p*-values from p_1 to p_K (small to large)
    - if $p_K < FWER$, all are significant
    - if not, check if $p_{K-1} < FWER / 2$; all remaining significant
    - if not, check if $p_{K-2} < FWER / 3$; etc.

13

controlling Type I error rates

- for unplanned (post-hoc, data-snooping) contrasts, use Scheffe's procedure
- it's the method of last resort

14

writing about results

Three pairwise comparisons were executed by orthogonal contrasts. To control the Type I error rate, a Bonferroni-corrected $\alpha = .05/3 = .017$ was used. The imagery group ($M = 12$) had significantly better memory than the control group ($M = 6$), $t(27) = 3.48$, $p = .001$. The rhyme group ($M = 10$) had non-significantly better memory than the control group ($M = 6$), $t(27) = 2.32$, $p = .03$. The imagery and rhyme groups also did not differ significantly, $t(27) = 1.16$, $p = .26$.

15

or …

Three pairwise comparisons were executed by orthogonal contrasts. To control the Type I error rate, Bonferroni-corrected *p*-values were used with $\alpha = .05$. The imagery group ($M = 12$) had significantly better memory than the control group ($M = 6$), $t(27) = 3.48$, $p = .005$. The rhyme group ($M = 10$) had non-significantly better memory than the control group ($M = 6$), $t(27) = 2.32$, $p = .085$. The imagery and rhyme groups also did not differ significantly, $t(27) = 1.16$, $p = .77$.

16