categorical predictors (part 5: "post-tests")

February 12, 2024

1

review

- both contrast codes and dummy codes provide ways to compare groups using *single-df contrasts*
- why single *df*? because each comparison is about a single parameter
- some (most?) software defaults to dummy-coding
- whatever codes you use, *F* for the whole model (all of the predictors combined) will be the same

2

"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: $\hat{Y} = b_0 + 0\ X_1 + 0\ X_2$

3

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 (or dummy coding)
- it requires redoing analyses multiple times (and in some cases generating irrelevant contrasts)
- 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 ...

4

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?

5

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)

6

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

7

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.

8

controlling Type I error rates

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

9

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, $t(27) = 1.16$, $p = .26$.

10

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, $t(27) = 1.16$, $p = .77$.

11

a little theory

(time permitting)

12

reminders about SSE and SSR

- in a design with three groups, Model A is

$$Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_{ij}$$

- we can express predicted scores as follows

$$\hat{Y}_{ij} = b_0 + b_1 X_1 + b_2 X_2 \text{ or } \hat{Y}_{ij} = \bar{Y}_{.j}$$

- and we can express (estimates of) residuals

$$e_{ij} = Y_{ij} - \bar{Y}_{.j}$$

- and SSE(A) is

$$SSE(A) = \sum (Y_{ij} - \bar{Y}_{.j})^2$$

13

reminders about SSE and SSR

- in a design with three groups, Model C (for the usual ANOVA) is

$$Y_{ij} = \beta_0 + 0X_1 + 0X_2 + \varepsilon_{ij}$$

- we can express predicted scores as follows

$$\hat{Y}_{ij} = b_0 \text{ or } \hat{Y}_{ij} = \bar{Y}_{..}$$

- and we can express (estimates of) residuals

$$e_{ij} = Y_{ij} - \bar{Y}_{..}$$

- and SSE(C) thus is

$$SSE(C) = \sum \left(Y_{ij} - \bar{Y}_{..}\right)^2$$

14

reminders about SSE and SSR

- if we compare Model A to Model C, we get SSR

$$SSR = SSE(C) - SSE(A)$$

- SSR is the reduction (improvement) in SSE
- it can be re-expressed as follows

$$SSR = \sum n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

- or (less formally, but more clearly, I hope)

$$SSR = \sum n_{\text{group}} (\bar{Y}_{\text{group}} - \bar{Y}_{\text{overall}})^2$$

15

these SS values have aliases in the context of ANOVA

$$SSR = SS_{\text{between}}$$
$$SSE(A) = SS_{\text{within}}$$
$$SSE(C) = SS_{\text{total}}$$

- SS_{between} is a measure of differences between groups, with sample size playing a role
- Why do group means differ?
  - real differences + noise
- SS_{within} is a measure of differences within groups
- Why do scores within groups differ?
  - noise

16

now MSs

- the *df* associated with *SS*s can be used to calculate *MS* values, as follows

$$MS_{\text{between}} = {SS_{\text{between}}}/{k-1}$$

$$MS_{\text{within}} = {SS_{\text{within}}}/{n-k}$$

17

finally the *F*-ratio

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

18

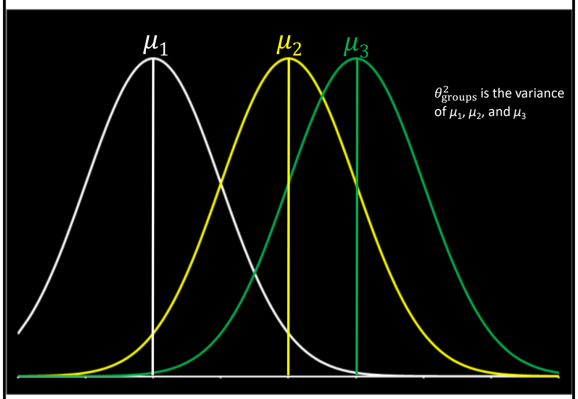
what contributes to the *F*-ratio?

formally

$$F = \frac{E(MS_{\text{between}})}{E(MS_{\text{within}})} = \frac{\sigma_{\text{e}}^2 + n\theta_{\text{groups}}^2}{\sigma_{\text{e}}^2}$$

what?!

19



20

what contributes to the *F*-ratio?

formally

$$F = \frac{E(MS_{\text{between}})}{E(MS_{\text{within}})} = \frac{\sigma_{\text{e}}^2 + n\theta_{\text{groups}}^2}{\sigma_{\text{e}}^2}$$

informally!

$$F = \frac{\text{noise} + \text{sample size} \times \text{group diffs}}{\text{noise}}$$

21

why does this matter?

$$F = \frac{\text{noise} + \text{sample size } \times \text{group diffs}}{\text{noise}}$$

- if noise is minimized, power goes up
- if sample size is increased, power goes up
- if groups are more different, power goes up

- this also is how an *F*-ratio is constructed: if there are no group diffs (it's 0), the numerator and denominator are both noise and *F* is expected to equal 1

22