STAT 5200 Handout #8a

Intro. to Factorial Design (Ch. 8-9)

Handout #8 Example – Two factors of interest: $\underbrace{\text{Glass}}_{\text{2 levels}}$ & $\underbrace{\text{Phosphor}}_{\text{3 levels}}$

- Want to investigate how these together affect Output
- Consider treatment structure:

		Phosphor		
		A	B	C
Glass	1	• • •	• • •	• • •
	2	• • •	• • •	• • •

• = one exp. unit

- We could model factor effects separately:

means models

$$\left\{\begin{array}{lll} Y_{ij} & = & \mu_i^{(Glass)} + \epsilon_{ij} \qquad i=1,2 \quad j=1,\dots,9 \\ & \text{- OR -} & \\ Y_{ij} & = & \mu_i^{(Phosphor)} + \epsilon_{ij} \qquad i=1,..,3 \quad j=1,\dots,6 \end{array}\right.$$

But really, we have six treatment "combinations":

Combo:	1	2	3	4	5	6
Glass	1	1	1	2	2	2
Phosphor	A	B	C	A	B	C

This is an example of factorial treatment structure

- g treatments are comprised of all factor-level combinations of 2 (or more) factors
- Accounting for this structure can lead to interesting insight

(coming up later: interactions)

We could look at this (Handout #8) as we have before: one-way ANOVA

$$Y_{ij} = \mu_i^{(combo)} + \epsilon_{ij}; \quad i = 1, \dots, 6; \quad j = 1, 2, 3$$

- It looks like $\underbrace{\text{(Combo 1,2,3)}}_{\text{Glass}=1} > \underbrace{\text{(Combo 4,5,6)}}_{\text{Glass}=2}$ ⇒ suggests a signif. Glass effect
- But is there a Phosphor effect, or not? → need to account for treatment structure

Two-Way ANOVA (two-factor factorial design)

Consider two factors A (with levels $1, \ldots, a$) and B (with levels $1, \ldots, b$)

→ Glass → Phosphor

1. Means model

$$Y_{ijk} = \underbrace{\mu_{ij}}_{\text{true pop. mean for treatment comb. } ij} + \underbrace{\epsilon_{ijk}}_{\text{random error}} \qquad \epsilon_{ijk} \text{ iid } N(0, \sigma^2)$$

Y_{ijk} = response value for replicate (or exp. units) k in factor A level i and factor B level j

$k = 1, \ldots, n_{ij}$

$i = 1, \ldots a$

$j = 1, \ldots, b$

HW#8 ex.

$n_{ij} = 3$ for all ij

$a = 2$

$b = 3$

$N = 18$

- $\hat{\mu}_{ij} = \bar{Y}_{ij\cdot}$; $n_{ij} = n$ for balanced data, so total sample size $N =$ $a \cdot b \cdot n$

- # parameters = $a \cdot b$ = # of factor level combinations
  $\mu_{11}, \mu_{12}, \ldots$

- But this obscures the effects of A and B. To clearly separate the effects of the two factors, we need an effects parameterization (with same # parameters).

2. Effects model

sometimes → $Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ijk}$

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

HW#8 ex

$Y_{ijk} = \mu + G_i + P_j + GP_{ij} + \varepsilon_{ijk}$

Parameter	Meaning	Constraint	# parameters
μ	population mean	(none)	1
α_i	mean effect of A level i	$\sum_{i=1}^{a} \alpha_i = 0$ (or $\alpha_a = 0$)	$a-1$
β_j	mean effect of B level j	$\sum_j \beta_j = 0$ (or $\beta_b = 0$)	$b-1$
$\alpha\beta_{ij}$	interaction	$\sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0$	$(a-1)(b-1)$

With only μ, α_i, and β_j, total # parameters is only: $1 + a-1 + b-1 = a+b-1$

- We're missing some:
  $ab - (a+b-1) = (a-1)(b-1)$

Not: "A depends on B"

- What we're missing is the possibility that the effect of A on Y could depend on B.
- This is called an interaction of factors A and B
  ↳ parameters $\alpha\beta_{ij}$ $(\neq \alpha_i \cdot \beta_j)$

Visually checking interactions: Interaction plot

- Plot $\bar{Y}_{ij\cdot}$ vs. i and vs. j, with lines connected for other factor
- Key: Does the effect of one factor depend on the level of the other factor?
- Check:

Are lines [roughly] parallel?
If not (& where), then this would suggest an interaction exists

Estimates from effects model

$$\begin{aligned} \text{obs.} \quad Y_{ijk} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \\ \text{pred.} \quad \hat{Y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{\alpha\beta}_{ij} \end{aligned}$$

obs. - pred. = resid.

- using ordinary least squares – choose $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\widehat{\alpha\beta}_{ij}$ to minimize $\sum_{ijk} \left(Y_{ijk} - \hat{Y}_{ijk}\right)^2$
- here, sum-to-zero constraints
- Recall that $\cdot$ in the subscript represents taking the sum over that subscript, and $^-$ over something represents converting the sum to the mean
- Parameter estimates:

$$\begin{aligned} \hat{\mu} &= \bar{Y}_{\cdots} \\ \hat{\alpha}_i &= \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots} \\ \hat{\beta}_j &= \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots} \\ \widehat{\alpha\beta}_{ij} &= \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots} \end{aligned}$$

- In both means model and effects model, $\hat{Y}_{ijk} =$ $\bar{Y}_{ij\cdot}$

ANOVA Table with two factors (plus interaction)

- Same information as before (Handout # 5a), plus extra partitioning of model SS:

$SS_{Trt} = SS_A + SS_B + SS_{AB}$

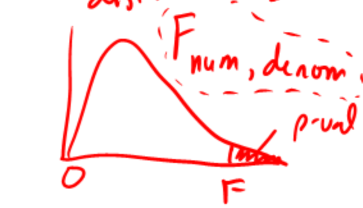
- SS are calculated as sum of squared [estimated] effects:

$$\begin{aligned} SS_A &= \sum_{ijk} \hat{\alpha}_i^2 = nb \sum_i \hat{\alpha}_i^2 \\ SS_B &= \sum_{ijk} \hat{\beta}_j^2 = na \sum_j \hat{\beta}_j^2 \\ SS_{AB} &= \sum_{ijk} \widehat{\alpha\beta}_{ij}^2 = n \sum_{ij} \widehat{\alpha\beta}_{ij}^2 \\ SS_E &= \sum_{ijk} e_{ijk}^2 = \sum_{ijk} \left(Y_{ijk} - \hat{Y}_{ijk}\right)^2 = \sum_{ijk} \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2 \end{aligned}$$

Type III SS

Source	DF	SS	MS	F	P-value
A	$a-1$	SS_A	$MS_A = SS_A/(a-1)$	$F_A = MS_A/MS_E$	
B	$b-1$	SS_B	$MS_B = SS_B/(b-1)$	$F_B = MS_B/MS_E$	
AB	$(a-1)(b-1)$	SS_{AB}	$MS_{AB} = SS_{AB}/(a-1)(b-1)$	$F_{AB} = MS_{AB}/MS_E$	
Error	$N-ab$	SS_E	$MS_E = SS_E/(N-ab)$		
Corrected Total	$N-1$	SS_{Total}			

sampling distn $F_{num, denom}$



What are nulls for F_A, F_B, and F_{AB}?

- $H_{0,A} : \alpha_1 = \alpha_2 = \ldots = \alpha_a$, or $A_1 = A_2 = \ldots = A_a$ no A main effect
- $H_{0,B} : \beta_1 = \beta_2 = \ldots = \beta_b$, or $B_1 = B_2 = \ldots = B_b$ no B main effect
- $H_{0,AB} : \alpha\beta_{11} = \ldots = \alpha\beta_{ab}$, or $AB_{11} = \ldots = AB_{ab}$ no interaction effect

We can also test these using "contrasts" (coming up in Handouts # 9 and # 10)

Factorial terminology and visual tool

- When g treatments are comprised of all factor-level combinations of 2 (or more) factors, we say those factors are "crossed". (Factors A and B are crossed.) we observe all their combinations
- Example: The ANOVA table above summarizes an $a \times b$ factorial design with n replicates at each factor level combination. ex: 2×3 factorial design w/ 3 reps
- This crossing can be visualized in a Hasse diagram (we'll build on this later in course)

← main

← interaction / crossing

← error